

Analiza prednosti i nedostataka rudarenja podataka

Analysis of advantages and disadvantages of data mining

Mahir Zajmović

Fakultet za primenjeni menadžment, ekonomiju i finansije Beograd, Univerzitet „Privredna akademija“ u Novom Sadu, Jevrejska 24, 11000 Beograd, R.Srbija, mahir.zajmovic@mef.edu.rs

Apstrakt: *Svako preduzeće ili bilo kakve organizacije prikupljaju velike količine podataka potrebnih za svoje poslovanje. Međutim, nas zanima, postoje li još nekakve dodatne informacije u tim već prikupljenim i poznatim podacima koje mogu biti korisne za poslovanje nekog preduzeća. Možemo li iz „iskopanih“ podataka „iskopati“ još korisnih informacija? Rudarenje podataka koristi metode koje su poznate u statistici, vještačkoj inteligenciji, matematici i računovodstvu radi automatizovanog otkrivanja skrivenih modela i uzoraka podataka u velikim količinama podataka. Svrha ovog procesa jeste predvidjeti ponašanje u budućim poslovnim događajima osnovi znanja otkrivenog u podacima. Na početku rada je postavljen metodološki okvir rada, a nakon toga objašnjen sam pojam rudarenja podataka i koje metode rudarenja razlikujemo. U nastavku rada su nabrojani i opisani programi koji se koriste za rudarenje podataka, zatim prednosti i nedostaci rudarenja podataka i na kraju opisano je područje primjene i osnovni cilj rudarenja podataka.*

Ključne riječi: *informacione tehnologije, podaci, baza podataka, rudarenje, znanje*

Abstract: *Every company or any organization collects large amounts of data necessary for its operations. However, we are interested in whether there is any additional information in this already collected and known data that can be useful for the business of a company. Can we "excavate" more useful information from the "mined" data? Data mining uses methods known in statistics, artificial intelligence, mathematics, and accounting to automatically discover hidden patterns and data patterns in large amounts of data. The purpose of this process is to predict behavior in future business events based on the knowledge discovered in the data. At the beginning of the work, the methodological framework of the work was set, and after that, the concept of data mining and the different mining methods were explained. In the rest of the paper, the programs used for data mining are listed and described, followed by the advantages and disadvantages of data mining, and finally, the field of application and the basic goal of data mining are described.*

Keywords: *information technology, data, database, mining, knowledge*

Uvod

Rudarenje podataka (engl. *Data Mining*) je naziv za fazu analize velike količine podataka u procesu otkrivanja novog znanja iz tih podataka (engl. *knowledge discovery in databases*). Rudarenje podataka se bavi izvlačenjem ili otkrivanjem novih i korisnih informacija iz podataka na osnovu otkrivanja ponavljajućih uzoraka. Otkrivanje znanja zasniva se na brojnim algoritmima od kojih je većina razvijena u području vještačke inteligencije i mašinskog učenja koji tako čine njegovu tehničku osnovu. Mašinsko

učenje opisuje se kao pronalaženje/dohvaćanje strukturalnih opisa iz primjera. Ti opisi se kasnije koriste za predviđanje i što je najveća prednost ispred statističkih modela, objašnjavanje i razumijevanje fenomena koji se istražuju. Rudarenje podataka predstavlja BI (engl. *Business Intelligence*) metodologiju koja pruža uvid u „skriven“ podatke o poslovanju čime se poboljšava proces donošenja strateških poslovnih odluka zasnovan na jasnoj i razumljivoj interpretaciji postojećih rezultata. Rudarenjem podataka dolazi se do logičnosti u podacima, odnosno otkrivanja odnosa, pravilnosti, zakonitosti i ostalih struktura među podacima.

Rudarenje podataka podrazumijeva multidisciplinarni pristup vrednovanju i organizovanju baza s posebnim naglaskom na čišćenje i pretprocesuiranje podataka kako bi se pristupilo znanju u bazama podataka. Cilj metodologije rudarenja je sticanje znanja na osnovu postojećih podataka u bazama prilikom čega je, upravo zbog orijentacije podataka, primjena neograničena s obzirom na industriju ili područje poslovnog djelovanja. Sam termin mogli bismo objasniti kao proces pronalaženja korisnog znanja ili informacija, odnosno otkrivanje znanja iz velike količine podataka. Rudarenjem se također otkrivaju odnosi, logičnost, pravilnost te uopšteno bilo kakve strukture među podacima. Rudarenje podrazumijeva organizovanje baza čišćenjem podataka kako bi se pristupilo znanju i sticanju istog na osnovu postojećih podataka u bazama. Razvoj tehnologije, računara, interneta bitno doprinosi lakšem organizovanju podataka, no da bi oni postali korisni, potrebno je njihovo pretvaranje u informacije i znanje.

Termin rudarenja često se poistovjećuje sa dva različita procesa: otkrivanje i predviđanje znanja. Proces otkrivanja znanja implicira korisnikovo razumijevanje eksplicitnih informacija za koje je bitno da su u čitljivom obliku. Predviđanje se odnosi na buduće događaje i u nekim pristupima može biti čitljivo i prozirno dok u drugim neprovidno.

Rudarenje podataka

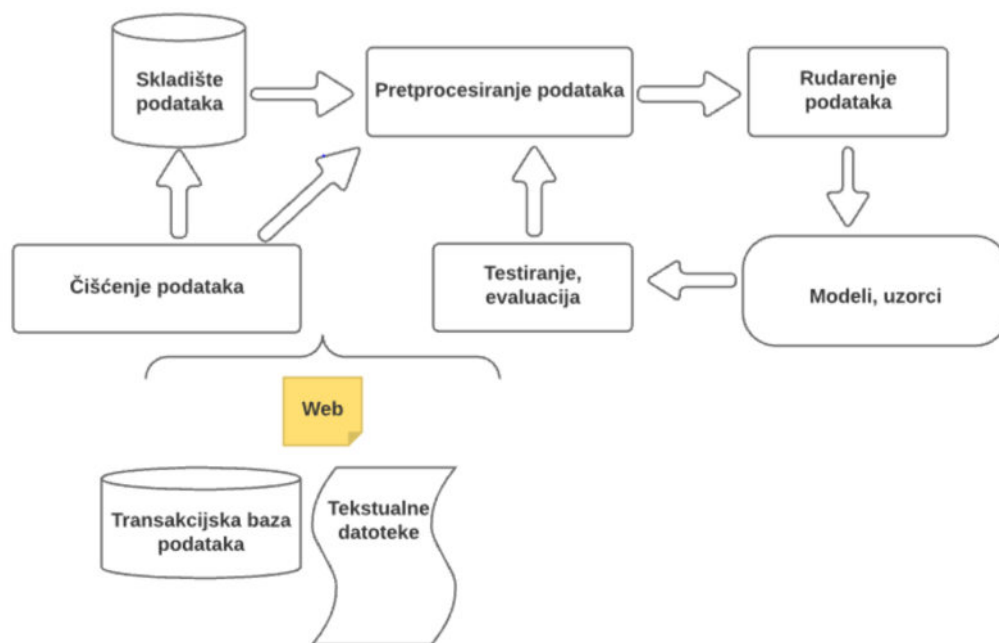
Rudarenje podatka (engl. *Data mining*) je pronalaženje zakonitosti u podacima. Podaci mogu biti organizovani u baze podataka, mogu biti tekstualni podaci, nestrukturirani podaci proizašli iz Web-a ili podaci koji su organizovani u vremenske serije. Primjenom raznih metoda za rudarenje podataka pronalazimo zakonitosti u podacima. Statistika, matematika, teorija informacija te vještačka inteligencija samo su neka od područja iz kojih potiču korijeni tih metoda (Panian & Klepac, 2003).

Tradicionalno rudarenje podataka podrazumijeva tradicionalne baze kao izvore podataka, a metode se primjenjuju nad tako formatiranim podacima. U novije vrijeme izdvajaju se pod područja s obzirom na izvore podataka kao što je to rudarenje Web-a, rudarenje teksta, te analiza vremenskih serija. Osnovni razlog izdvajanja ovih područja proizilazi iz činjenice što podaci nisu strukturirani u relacione tablice, već su nestrukturirani, ili pak strukturirani na osnovu specifičnog formata. Čišćenje podataka je jedan od osnovnih procesa u metodici otkrivanja znanja, bez obzira na izvor podataka. U tom se procesu podaci filtriraju od nečistoća, a posebna se pažnja obraća na analizu mogućih ekstremnih, tzv. stršćih vrijednosti koje ne moraju nužno biti šum u podacima već vrlo vrijedan podatak (Klepac, 2020).

U sistemima poslovne inteligencije rudarenje podataka je podržano skladištima podataka. Ukoliko dođe do nepostojanja skladišta podataka, što je u praksi čest slučaj, tada se odmah nakon procesa čišćenja podataka ulazi u proces obrade podataka. Korištenje skladišta podataka prilikom rudarenja podataka

ima prednost što skladišta podataka u vrlo kratkom vremenskom periodu mogu selektovati određene skupove podataka. Iako ovakav pristup štedi vrijeme analitičaru, u praksi se često ipak praktikuje direktan pristup podacima. Rudarenje podataka se može primjenjivati u svim područjima u kojima raspoložemo sa velikom količinom podataka iz domena tog područja i na osnovu tih podataka želimo otkriti određene pravilnosti, veze i zakonitosti. Ta područja mogu biti ekonomija, medicina, genetika, mehanika i druga područja. S obzirom na to da postoji čitav niz faktora koji mogu uticati na neki određen događaj, odnosno ishod istog, zadatak rudarenja podataka jeste otkriti najznačajnije faktore i njihova obilježja u odnosu na ciljano stanje (Klepac, 2020).

Slika 1. Otkrivanje znanja primjenom metoda rudarenja podataka



Izvor: Panian & Klepac, 2003

Rudarenje podataka je traženje novih i vrijednih informacija u velikim količinama podataka. Omogućeno je kroz saradnju čovjeka i kompjutera. Najbolji rezultati se dobivaju upravo balansom između znanja stručnjaka koji opisuje problem i mogućnostima traženja kompjutera. U praksi, primarni ciljevi rudarenja podataka jesu predviđanje i opisivanje. Predviđanje uključuje korištenje određenih varijabli ili polja u skupovima podataka kako bi se predvidjeli nepoznate ili buduće vrijednosti drugih potrebnih varijabli. Opisivanje se odnosi na pronalaženje obrazaca na način da se opisuju podaci koje čovjek može interpretirati. Aktivnosti rudarenja podataka se mogu smjestiti u jednu od dvije kategorije:

1. Prediktivno rudarenje podataka, koje proizvodi model sistema na osnovu raspoloživih skupova podataka, ili
2. Deskriptivno rudarenje podataka, koje proizvodi nove, netrivialne informacije zasnovane na dostupnim skupovima podataka (Kantardžić, 2011).

Krajnji cilj prediktivnog rudarenja podataka jeste proizvodnja modela, koji je izražen kao izvršni kod, koji se može koristiti za izvođenje klasifikacije, predikcije, procjene ili drugih sličnih zadataka. Krajnji cilj

deskriptivnog rudarenja podataka jeste da se dobije razumijevanje analiziranog sistema otkrivanjem obrazaca i veza unutar velikih skupova podataka. Ciljevi prediktivnog i deskriptivnog rudarenja podataka postignuti su korištenjem tehnika/metoda rudarenja podataka, za sljedeće primarne zadatke rudarenja podataka:

1. Klasifikacija: otkrivanje funkcija prediktivnog učenja koje klasifikuje podatak u jednu od predefinisanih klasa.
2. Regresija: otkrivanje funkcija prediktivnog učenja koje smještaju podatak u prediktivnu varijablu koja ima stvarnu vrijednost.
3. Klasteriranje: deskriptivni zadatak u kojem jedan nastoji identifikovati završni skup kategorija ili klastera kako bi se opisao podatak.
4. Sumarizacija: dodatni deskriptivni zadatak koji uključuje metode za pronalaženje kompaktnih opisa za skupove podataka.
5. Ovisno modeliranje (engl. *Dependency modeling*): traženje lokalnog modela koji opisuje značajne ovisnosti između varijabli ili između vrijednosti neke karakteristike unutar skupa podataka ili dijelovima skupa podataka.
6. Promjena i otkrivanje devijacija: otkrivanje najznačajnijih promjena unutar skupova podataka (Larose, 2004).

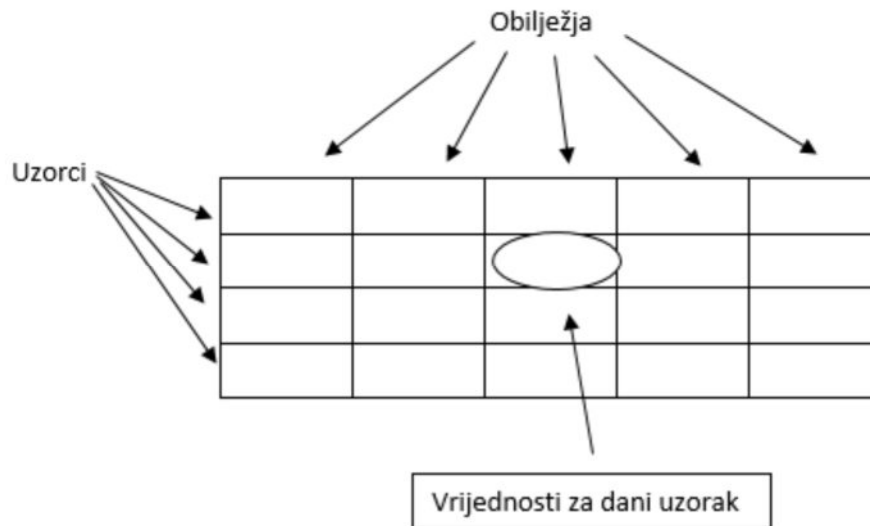
Veliki skupovi podataka

Količina podataka sa kojima raspolazemo je često prevelika da bi se obradili ručnom analizom, čak i za neke analize koje su bazirane na korištenju računara. Logično je da će neka osoba, odnosno voditelj raditi efektivnije ukoliko raspolaze sa velikom količinom podataka, nekoliko stotina ili hiljada podataka u arhivi. Poslovno društvo je svjesno današnjeg problema velike opterećenosti informacijama, te jedna analiza pokazuje sljedeće:

- a) 61% menadžera vjeruje kako je opterećenje informacija prisutno u njihovom radnom okruženju;
- b) 80% vjeruje kako će situacija postati još lošija;
- c) preko 50% menadžera ignorišu podatke u trenutnom procesu donošenja odluka zbog preopterećenja informacija;
- d) 84% menadžera pohranjuje te podatke za budućnost, ne koriste ih u aktualnim analizama;
- e) 60% vjeruje da je trošak prikupljanja informacija veći od same vrijednosti (Policki, 2019).

Različiti tipovi podataka su generisani i digitalno pohranjeni u današnjem multimedijalnom okruženju koji ima veliku infrastrukturu interneta. Kako bi se odabrale odgovarajuće metode za rudarenje podataka, moraju se analizirati osnovni tipovi i karakteristike skupova podataka. Prvi korak u toj analizi je sistematizacija podatka u odnosu na kompjuterski prikaz i upotrebu. Podaci koji su obično izvor za proces rudarenja podataka mogu biti klasifikovani u strukturisane, polu strukturisane i nestrukturisane podatke. Većina poslovnih baza podataka sadrže strukturisane podatke koji se sastoje od dobro definisanih polja sa numeričkim ili alfanumeričkim vrijednostima, dok naučne baze podataka mogu sadržavati sve tri klase podataka.

Slika 2. Tabelarni prikaz skupa podataka



Izvor: Kantardžić, 2011

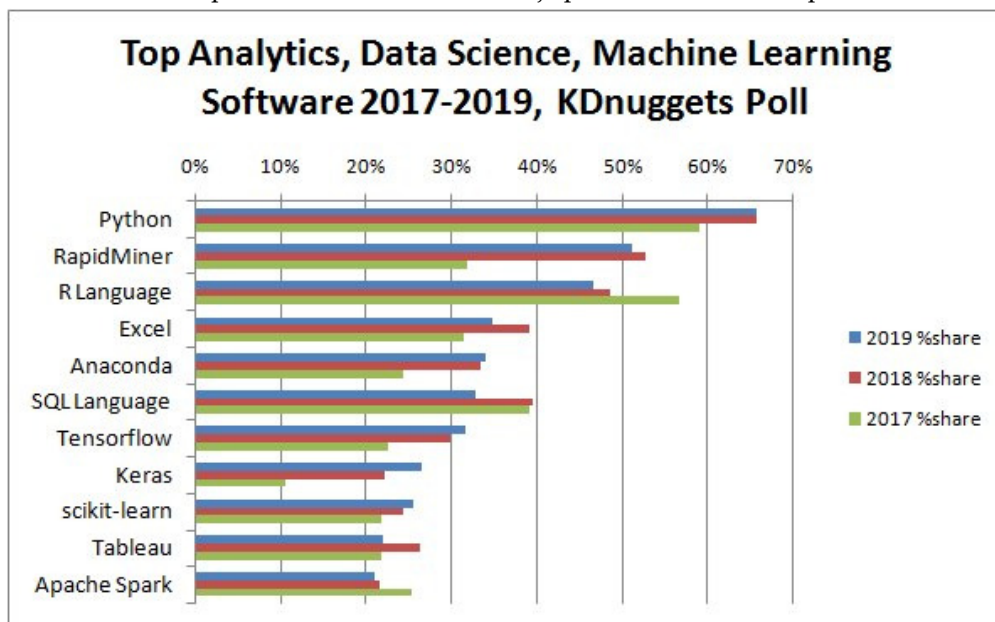
Primjeri polu strukturisanih podataka su elektronske slike poslovnih dokumenata, medicinski izvještaji, izvršni sažeci i slično. Većina Web dokumenata takođe spada u ovu kategoriju. Primjeri nestrukturisanih podataka je video snimljen nadzornom kamerom u nekom odjeljenju. U literaturi rudarenja podataka koristi se pojam uzorak (eng. *sample*) ili slučaj (eng. *case*) za redove. Mnogo različitih tipova obilježja (atributi ili varijable), što su, polja u arhivi strukturisanih podataka su zajednička u rudarenju podataka. Nisu sve metode za rudarenje podataka jednako dobro kada je riječ o različitim tipovima obilježja. Današnji računari te odgovarajući softverski alati podržavaju procesiranje grupe podataka koji imaju milione uzoraka i stotine obilježja.

Programi za rudarenje podataka

Programi za rudarenje podacima omogućavaju nam rješavanje problema rudarenja podataka. Koristimo ih za rješavanje problema klasifikacije, klasterisanja, bayesovih mreža, asocijativnih pravila, te kod ostalih metodologija rudarenja podataka. Postoje alati koji su besplatni, te komercijalni alati. Prema anketi koja je provedena 2019. godine sa stranice www.kdnuggets.com, 91% korisnika koji su učestvovali u anketi koristi komercijalne alate, a 73% besplatne alate, oko 27% koristi samo komercijalne alate, a samo 9% koristi isključivo besplatne alate. Većina od 64% koristi i komercijalne i besplatne alate za rudarenje podacima, dok je u anketi 2017. godine, tek 49% ispitanika koristilo komercijalne i besplatne.

Najpopularniji alat na cjelokupnom području rudarenja podataka i nauke podataka jest Python, a slijedi ga RapidMiner. Top deset analitičkih alata, alata za rudarenje podacima i nauka podataka korištenih u 2019. godini su Python sa 65,8%, RapidMiner sa 51,2%, R Language sa 46,6%, Excel sa 34,8%, Anaconda sa 33,9%, SQL sa 32,8%, Tensorflow sa 31,7%, Keras sa 26,6%, Scikit-learn sa 25,5%, Tableau sa 22,1%, Apache Spark sa 21,0%.

Slika 3. Top alati za analitiku, rudarenje podataka, znanost o podacima



Izvor: www.kdnuggets.com

Tabela 1. Udio alata za analitiku, rudarenje podataka, znanost o podacima

| Program (software) | Udio % (2019) | Udio % (2018) | Udio % (2017) |
|--------------------|---------------|---------------|---------------|
| Python | 65,8% | 65,6% | 59,0% |
| RapidMiner | 51,2% | 52,7% | 31,9% |
| R Language | 46,6% | 48,5% | 56,6% |
| Excel | 34,8% | 39,1% | 31,5% |
| Anaconda | 33,9% | 33,4% | 24,3% |
| SQL Language | 32,8% | 39,6% | 39,2% |
| Tensorflow | 31,7% | 29,9% | 22,7% |
| Keras | 26,6% | 22,2% | 10,7% |
| scikit-learn | 25,5% | 24,4% | 21,9% |
| Tableau | 22,1% | 26,4% | 21,8% |
| Apache Spark | 21,0% | 21,5% | 25,5% |

Izvor: www.kdnuggets.com

Iz gore navedena tabele možemo uočiti da je Python najkorišteniji program za rudarenje podataka.

Prednosti i nedostaci rudarenja podataka

Rudarenje podataka uključuje proces traženja velike količine prethodno nepoznatih podataka, a zatim ih koristi za donošenje važnih poslovnih odluka. Ključna fraza ovdje je „nepoznati podatak“, što znači da je podatak zatrpan među ogromnim količinama operativnih podataka koji, ako su analizirani, pružaju relevantne informacije organizacionim donosiocima odluka. Značajni podaci ponekad nisu pronađeni, jer

je većina podataka zarobljena i čuva se u određenom odjeljenju. Ono što može da izgleda irelevantno ili neinteresantno na nivou odeljenja, može postati uočljivo i važno na organizacionom nivou, posebno, ponašanje kupaca i modeli kupovine. Određivanje efikasnosti promocija prodaje, otkrivanje prevare, procjenjivanje rizika i procjenjivanje kvaliteta, kao i analiziranje prava osiguranja su neke od oblasti u kojima ta informacija može pomoći u donošenju odluka. Često, podaci koji su potrebni nalaze se na nekoliko različitih sistema, u različitim formatima i strukturi, pa čak mogu biti donekle i suvišni. Ovi podaci ne daju nikakvu vrijednost za poslovne korisnike, bez određene metode pristupanja njima. Ovo je mjesto gdje skladište podataka ulazi u igru kao bolji izvor konsolidovanih i očišćeni podataka, olakšavajući analizu mnogo efikasnije nego regularne datoteke ili baze podataka.

Tri koraka su potrebna za identifikaciju i korišćenje skrivenih informacija:

1. Pronalaženje skrivenih podataka i dobijanje informacija iz njih.
2. Informacije moraju biti posebno organizovane kako bi se pojednostavilo donošenje odluka.
3. Podaci moraju biti analizirani ili iskorišteni za vrijednu informaciju (Sharp, 2020).

Tehnike rudarenja podataka kakve se danas primjenjuju rezultat su decenija usavršavanja. Njihova evolucija započela je onoga trenutka kada su prvi podaci o poslovanju smješteni u memorije računara, a nastavila se usavršavanjem načina za pristupanje tim podacima. U koraku koji je uslijedio, ista su nastojanja rezultovala tehnologijama koje kupcima omogućavaju navigaciju kroz podatke smještene u skladištima u realnom vremenu. Aplikacije za rudarenje podataka ovim su procesom dovedene do nivoa na kojem su u stanju grupisati podatke prema nekim zajedničkim karakteristikama, predviđati buduće događaje te ponuditi brze i tačne odgovore na mnoga pitanja koja muče korisnika.

Prednosti rudarenja podataka

Prednosti i ograničenja ove metodologije često se iskrivljeno interpretiraju prezentujući potencijal kao rješenje svih poslovnih problema ili rješenja svih problema savremenog poslovanja. Činjenice pokazuju da se primjenom metoda i alata za rudarenje podataka mogu otkriti ponekad iznenađujući obrasci ponašanja nekih osoba ili neočekivani odnosi među pojavama, ali to nije razlog da se jasno ne izraze ograničenja odnosno preuveličavaju mogućnosti ovih tehnika. Jedna od definicija rudarenja podataka glasi: rudarenje podataka je sistematičan, interaktivan i iterativan (ponavljajući) proces izvođenja i prikazivanja korisnog, implicitnog i inovativnog znanja podataka. Data mining je izrazito multidisciplinarno područje. Ono obuhvata područja baza podataka, ekspertnih sistema, teorije informacija, statistike, matematike, logike i čitav niz pridruženih područja. Područja u kojima se rudarenje podataka može uspješno primjenjivati su raznolika, na primjer, poslovanje organizacija, ekonomija, mehanika, medicina, genetika itd. Uopće gledajući, rudarenje podataka primjenljivo je u svim onim područjima gdje se raspoloživo velikom masom podataka na osnovu kojih se žele otkriti određene pravilnosti, veze i zakonitosti.

Data mining treba razlikovati od automatske obrade podataka, sa kojom se često poistovjećuje. Dok su na prvi pogled rezultati obje tehnike grafikoni i tabele sa podacima koji predviđaju neke trendove i kvantifikuju postojeće pojave, razlika se ogleda u tome što se automatska obrada podataka odvija po prethodno definisanim šablonima i procedurama, dok data-mining aplikacije pokazuju samostalnost u analizi podataka. Dva su osnovna načina korišćenja znanja otkrivenog u procesu rudarenja podataka:

1. prezentovanje znanja i otkrivenih pravila u obliku izvještaja;
2. formalizovanje i skladištenje znanja u sistemu zasnovane na pravilima (*Rule Based Systems, RBS*) vrstu pravila koja se mogu izraziti u obliku AKO-ONDA (Panijan & Klepac, 2003).

Postoje i hibridni sistemi, koji su osim pravila sposobni prihvatiti i dinamičke modele, poput Bayesovih i neuralnih mreža. Kada se radi o sistemima zasnovanim na pravilima, mogu se izdvojiti dva osnovna tipa takvih sistema – tradicionalni ekspertni sistemi i sistemi zasnovani na neizravnoj logici. Rezultati istraživanja kao što je ovo predstavljeno u radu mogu se iskoristiti kao podloga ili dopuna nekom od ovakvih sistema.

Nedostaci rudarenja podataka

Glavni nedostatak svih metoda rudarenja podataka jest što predmetne metode spadaju u tzv. nedeterminističke metode. Nedeterminizam je fleksibilniji način računanja gdje ishod pojedine računске operacije nije jednoznačno određen. Nedeterminizam treba shvatiti kao misaoni koncept koji pomaže u teorijskim razmatranjima i omogućuje jednostavnije dokazivanje teorema. Računalni determinizam (i nedeterminizam) opisuje postupak proračuna (Sipser, 2013). Deterministički je proračun onaj koji ima funkcionalni prelaz iz stanja u stanje, što znači da se tačno zna iz kog će stanja preći u koje, dok se u slučaju nedeterminizma prelaz iz stanja u stanje definiše relacijski. Ovo je smisljeno, ali u praktičnom se smislu može postići jedino tako da se proračun masivno paralelizuje i da se svaki od tih procesa paralelno ostvaruje u memoriji.

Usljed svega navedenog nastaju sljedeći problemi i nedostaci:

- nije poznat tačan broj grana na koji se proces izgrana (ovo često ne djeluje kao problem, ali u praksi jeste zbog potrebe definisanja memorije na početku procesa);
- broj paralelnih proračuna može biti izuzetno veliki.

U skladu sa navedenom, vrlo često se ističu prednosti determinističkih metoda i proračuna, ali nedeterminističke metode odnosno rudarenje podataka može imati brojne prednosti u otkrivanju skrivenih uzoraka među podacima do kojih neki skriveni uzorci mogu rezultirati signalima ranog upozorenja odnosno tzv. EWS – *Early Warning Signals*. Predmetni signali se često ne mogu izračunati deterministički pa stoga ove metode mogu imati brojne prednosti takvog tipa. Uopšteno, alati razvijeni za rudarenje podataka vrlo su komplikovani i jaki alati. Ali, zahtijevaju vrlo stručnu osobu za pripremu podataka i razumijevanje rezultata. Budući „rudar podataka“ mora poznavati različite obrasce i odnose čiji se odnos mora valjano izvesti. Nadalje, kako se prikupljaju ogromne količine podataka u sistemima za prikupljanje podataka, neki od tih podataka koji su vrlo kritični i osjetljivi mogu biti hakovani, kao što se i dogodilo sa mnogim velikim kompanijama. Takođe, informacije generisane tehnikom rudarenja podataka mogu se iskoristiti i u pogrešne svrhe. Kod operatora elektroenergetskih sistema ti podaci i informacije mogu se iskoristiti za različite aktivnosti u sferi terorizma, napada na infrastrukturu zemlje itd (Cotilla-Sanchez & Danforth, 2013).

Uprkos brojnim primjerima uspješnog korištenja u nauci i inženjeringu te ostalim granama istraživanja, metoda rudarenja podataka pokazala je i brojne mane. Pri nepravilnom korištenju, rudarenjem podataka moguće je kao rezultat dobiti nekvalitetne izlazne informacije. Primjeri takvih informacija poznati su i u

istoriji kao na primjer korelacija proizvodnje maslaca u Bangladešu i indeksa dionica S&P 500. Takođe, rudarenjem podataka dođe se do netačnog zaključka da na primjer pobjeda ekipe iz zapadnog dijela lige u američkom fudbalu implicira porast dionica na berzi. Poznat je citat da „uz dovoljno vremena, dovoljno pokušaja i dovoljno mašte, skoro svaki skup podataka može generisati potpuno nesuvisle informacije“. Generalno, postoje četiri glavne zamke prilikom modelovanja tehnikom rudarenja podataka i to su:

1. uvijek postoji izazov za modelovanjem „čudnih“ podataka ili „čudnog“ ponašanja uzorka podataka;
2. može se pronaći dokaz za svaku teorijsku pretpostavku ako se pusti računarski proces da dovoljno dugo kopa po podacima;
3. zaključci imaju više smisla ako postoji plauzibilna teorija koja to može potkrijepiti, ali „zabavni“ zaključak sakriti slabosti u podacima;
4. što više faktora ili karakteristika se razmatra u skupu podataka, to je veća vjerovatnost da će se tehnikom rudarenja podataka pronaći određena poveznica bez obzira da li je opravdana ili ne.

Zaključak

Rudarenje podataka se primjenjuje kako bismo iz velike količine podataka došli do zanimljivih i potrebnih informacija ili pak znanja. Kako bi rudarenje podataka bilo moguće, postoji čitav niz metoda za rudarenje podacima. S obzirom na široku primjenu tog područja neke metode nisu deklarirane kao isključive metode rudarenja podataka, ali u literaturi se naravno spominju i one koje spadaju u područje rudarenja podataka. U ovom radu navedene su neke od metoda. Stabla odlučivanja, te Naivni Bayesov klasifikator spadaju među najpoznatije i najkorištenije metode za rješavanje problema klasifikacije i predikcije. Ukoliko postoji potreba za razvrstavanjem, odnosno grupisanjem uzoraka u predefinisane grupe tada koristimo metode za klasterisanje kao što su hijerarhijsko klasterisanje ili K-means klasterisanje. U ovom radu spomenuto je i memorijski temeljno razlučivanje koje je građevni element algoritma K-means metode klasterisanja, te metoda potrošačke korpe koja spada pod asocijacijska pravila te se najčešće koristi za otkrivanje povezanosti unutar transakcija koja se tiču prodaje robe u maloprodajnim centrima.

Savremeni računari imaju velike, ali još uvek i nedovoljno istraživane mogućnosti. Informacija u današnje vrijeme nije dovoljna, potrebno je znanje šta učiniti sa tim informacijama. Kaže se da će u ovom vijeku biti uspješna samo ona organizacija koja bude znala koristiti sadašnje i generisati novo znanje. To znanje je već u računarskim memorijama i bazama podataka, ali je gotovo nevidljivo. Familija proizvoda *Business Intelligence* zajedno sa data mining proizvodima je unijela nov, revolucionaran pogled na rješavanje problema upravljanja podacima u poslovnim sistemima. Pojavom data mining-a kompanije konačno mogu da u potpunosti iskoriste sakupljene podatke i da poboljšaju ne samo svoje poslovanje, donošenje poslovnih odluka, već i izuzetno važnu komunikaciju sa kupcima. Konačno su kupac i njegove potrebe stavljene u prvi plan. Menadžerima je olakšano donošenje važnih odluka na osnovu data mining analiza.

Rudarenje podataka je sve više prisutno u preduzetništvu zbog velikog gomilanja podataka koje se stvara u preduzećima. U tom velikom skupu podataka sama preduzeća mogu naći neke vrlo korisne informacije za samo poslovanje i samim tim preduzeća u tim podacima vide veliku korist. Isto tako bitno je da podaci koje preduzeće analizira budu kvalitetni, to jest korisni za preduzeće. Takođe za kvalitetno rudarenje

podataka potrebno je izabrati pravilnu metodu rudarenja podataka za određeni skup podataka. Tu u fokus dolaze stručnjaci za rudarenje podataka koji su sve više traženi. Ukoliko preduzeće pravilno rudari dostupne podatke i uspješno ih analizira i primjenjuje u poslovanju tada ono može očekivati pozitivne rezultate razvoja i poslovanja.

Na kraju možemo zaključiti da se rudarenjem podataka poboljšava proces donošenja odluka na strateško-poslovnom nivou pružajući uvid u analizu i upravljanje tržištem, korporativnu analizu i upravljanje rizicima, kao i otkrivanje prevare.

Reference

Panian, Ž., Klepac, G. (2003). Poslovna inteligencija, Masmedia, Zagreb

Zajmović, M., Brzaković, M., Karabašević, D. (2021). Improving business processes by applying the big data concept in modern business, Faculty of applied management, economics and finance Belgrade, University Business Academy in Novi Sad

Kantardžić, M. (2011). Data mining concepts, models, methods and algorithms, Kanada, John Wiley & Sons, Inc., Hoboken, New Jersey

Larose, D. (2004). Discovering Knowledge in Data: An Introduction to Data Mining, Wiley-Interscience

Policki, M. (2019). Proces data mininga nad podacima o prodaji tekstila, Završni rad, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Zagreb

Vinčević, V., Zajmović, M. (2021). Big Data koncept i primjena u savremenom kriznom poslovanju, SKEI - međunarodni interdisciplinarni časopis, Vol. 2 No. 2, Sveučilište/Univerzitet "Vitez", Vitez

Foss, B., Stone, M. (2001). Successful Customer Relationship Marketing, Kogan Page Limited, London

Sharp, D. E. (2002). Customer Relationship Management Systems Handbook. Boca Raton, FL 33487, USA: Auerbach Publications

Sipser, M. (2013). Introduction to the theory of computation, Cengage Learning, Massachusetts, SAD. Boston

Zajmović, M., Obhodaš, I., Dželilović, E. (2022). Primjena Big Data koncepta kod upravljanja kvantitativnim podacima u modernom poslovanju, Međunarodni centar za slavjanska prosveta, Sveti Nikole

Cotilla-Sanchez, E., Hines, P. & Danforth, C. (2013). Predicting critical transitions from time series synchrophasor data, IEEE Power and Energy Society General Meeting, Canada, Vancouver

<http://www.goranklepac.com>

www.kdnuggets.com