Analiza i rudarenje posjeta web mjesta

Pelesk, Paolo

Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
«Dr. Mijo Mirković»

PAOLO PELESK

ANALIZA I RUDARENJE POSJETA WEB MJESTA

Završni rad

Pula, 2016.

Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
«Dr. Mijo Mirković»

PAOLO PELESK

ANALIZA I RUDARENJE POSJETA WEB MJESTA

Završni rad

JMBAG: 0242011514, izvanredan student

Studijski smjer: Informatika

Predmet: Sustavi temeljeni na znanju

Mentor: dr. sc. Vanja Bevanda

Pula, 2016.

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Paolo Pelesk, kandidat za prvostupnika informatike 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 necitiranog rada, te da ikoji 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:

U Puli, 01. travnja 2016. _____

IZJAVA

o korištenju autorskog djela

Ja, Paolo Pelesk dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom „Analiza i rudarenje posjeta web mjesta“ 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, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, 01. travnja 2016.

Potpis

_____

Sadržaj

1. Uvod

Rudarenje podataka kao pomoć pri donošenju poslovnih odluka sve češće se primjenjuje u svim granama poslovanja i života. Rudarenje posjeta Web mjesta koristi sve metode i algoritme rudarenja podataka uz dodatak posebnih procesa i metoda specifične za Web analitiku. Cilj Web analitike je identificirati korisnike, njihove potrebe, navike i načine korištenja Web mjesta.

Sustavi preporuka i personalizirani sadržaj koriste sve Web trgovine, a sve više se koristi i na drugim Web mjestima. Tako će dnevni portali prikazati relevantne članke ovisno o analiziranim interesima, prethodno pregledanom Web mjestu i slično.

Cilj ovog rada je objasniti procese rudarenja podataka, rudarenja Weba i Web analitike i koje znanje je moguće dobiti iz podataka. Nakon teoretskog djela u kojem će se objasnit osnovne metode, procesi i definicije rudarenja podataka, bit će na primjeru prikazano rudarenje podataka nad proizvoljnim podacima o posjeti Web mjesta.

Ovaj rad sastoji se od sedam poglavlja. Nakon uvoda, u drugom poglavlju obrađuje se Web analitika, objašnjavaju se osnovni pojmovi i proces. Treće poglavlje objašnjava metode, tehnike i proces rudarenja podataka.

Četvrto poglavlje objašnjava proces i metode korištene za rudarenje Weba. Peto poglavlje objašnjava različite Web metrike i koje informacije o Web mjestu možemo dobiti od njih. Šesto poglavlje je rudarenje i analiza podataka na primjeru podataka o posjeti proizvoljnog Web mjesta. Sedmo poglavlje daje zaključak i opisan proces od posjetitelja do znanja.

2. Web analitika

Korisnici korištenjem informacijskih sustava svakim danom generiraju ogromne količine korisničkih i transakcijskih podataka. Web analitika je proces koji proučava načine ponašanja korisnika na Web mjestu. Cilj Web analitike je prikupiti, obraditi i analizirati podatke, a svrha svega je optimizirati Web mjesto, poboljšati prodaju, kvantitativno mjeriti učinkovitost marketinške kampanje, ponuditi personalizirani sadržaj putem sustava preporuka, otkriti sumnjivo ponašanje i kriminalne aktivnosti. Istraživanjem ponašanja posjetitelja na Web mjestu mogu se dobiti važni pokazatelji o korisnicima ili kupcima, njihovim interesima, demografskim podacima i slično. Znanje dobiveno iz procesa Web analitike pomažu poduzećima donijeti poslovne odluke o proizvodima i uslugama. Poduzeća uz dobivene informacije o trenutnim i potencijalnim korisnicima mogu dobiti i informacije o budućoj potražnji analizirajući trendove, društvene mreže i *feedback*[1] korisnika.

Proces rudarenja posjeta web mjesta može se podijeliti u dvije međusobno ovisne faze. Prva faza je prikupljanje i obrada podataka, a druga je otkrivanje i analiza uzoraka. *Clickstream*[2] podaci se čiste i odvajaju u setove korisničkih transakcija predstavljajući aktivnosti svakog korisnika prilikom različitih posjeta mjestu. U fazi otkrivanja uzoraka, koriste se metode iz statistike, baza podataka i strojnog učenja kako bi se pronašli “skriveni” uzorci ponašanja korisnika, kao i sažetak statistika web resursa, sesija i korisnika. Posljednja faza procesa je prethodno otkriveni uzorci i statistički podaci se dalje obrađuju i filtriraju kako bi na kraju rezultirali agregiranom korisničkom modelu koji se može koristiti u drugim aplikacijama, sustavima za preporuke, alatima za vizualizaciju i alatima za Web analitiku i izradu izvještaja.

2.1. Prikupljanje podataka o prometu

Važan zadatak u procesu za rudarenja podataka je izrada adekvatnog seta ciljanih podataka nad kojim se mogu izvršiti algoritmi. Čišćenje podataka je posebno važno za analizu posjeta web mjesta jer *clickstream* podaci i ostali povezani podaci dolaze iz različitih izvora.

Na čišćenje i obradu podataka potroši se najviše vremena i ponekad zahtjeva

[1] engl. *feedback. P*ovratna informacija.
[2] engl. *clickstream.* Podaci prikupljeni o aktivnostima korisnika na Web mjestu.

korištenje posebnih algoritama i metoda. Ovaj proces je ključan za uspješno otkivanje korisnih uzoraka iz podataka. Postupak može uključivati obradu originalnih podataka, integraciju iz više izvora i transformaciju integriranih podataka u oblik koji je prigodan za specifične metode rudarenja podataka. Priprema podataka donijela je zadatke kao što su: spajanje i čišćenje podataka, identifikacija korisnika i sesija, identifikacija *pageviewa*[3]. U analizi podataka za e-trgovinu, ove tehnike su dalje proširene kako bi omogućile otkrivanje važnih metrika o korisnicima i stranici.

2.2. Izvori i vrsta podataka

Primarni izvori podataka u rudarenju weba su poslužiteljski zapisi koji uključuju zapise Web i aplikacijskih poslužitelja. Dodatni izvori podataka su također bitni za pripremu podataka i otkrivanje uzoraka kao što su: meta podaci i baze podataka. U nekim slučajevima i za neke korisnike dodatni podaci su dostupni od strane davatelja internetskih usluga (ISP).

2.2.1 Poslužiteljski zapisi

Usage data[4] se automatski prikupljaju na Web i aplikacijskim poslužiteljima i predstavljaju detaljan pregled ponašanja korisnika na stranici ili aplikaciji. Zapisi poslužitelja su primarni izvor podataka. Svaki pregled stranice stvara HTTP zahtjev koji se bilježi na poslužitelju. Zapis se sastoji od nekoliko polja: vrijeme i datum zahtjeva, IP adresu klijenta (korisnika), zatraženi resurs, status zahtjeva, korištenu HTTP metodu, preglednik i operativni sustav korisnika, veza s koje je korisnik upućen i ponekad, korisnikove kolačiće koji koriste za identifikaciju korisnika koji su već posjetili Web mjesto i imaju pohranjene postavke ili druge informacije za personalizaciju sadržaja.

2016-02-01 12:07:14 1.2.3.4 – GET domena.hr/resuri.html – 200 9221 HTTP/1.1 domena.hr Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727) http://www.google.hr
2016-02-01 12:08:25 1.2.3.4 – GET domena.hr/predavanja.html – 200 9221 HTTP/1.1 domena.hr

[3] engl. *pageview. Pregled stranice*
[4] engl. *usage data. Podaci o korištenju*

Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727) http://www.domena.hr/resursi.html
2016-02-01 19:14:14 5.6.7.8 – GET domena.hr/kontakt.html – 200 9221 HTTP/1.1 domena.hr Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1) http://www.nekadrugastranica.hr/korisnapoduzeca.html
2016-02-01 19:15:01 5.6.7.8 – GET domena.hr/onama.html – 200 9221 HTTP/1.1 domena.hr Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1) http://www.domena.hr/kontakt.html

Tablica 1. Primjer poslužiteljskih zapisa. Izvor: autor

Iz gornjeg primjera poslužiteljskog zapisa vidimo iz prve linije datum i vrijeme zahtjeva, IP adresu klijenta, metoda i resurs koji je zatražen, HTTP status zahtjeva. U drugoj liniji je naziv domene na koju korisnik potražuje resurs. Treća linija je *useragent* odnosno preglednik i operativni sustav s kojim klijent pristupa mjestu. Posljednja linija je referenca koja je uputila klijenta na taj resurs.

Iz prvog zapisa vidljivo je da je posjetitelj na Web mjestu domena.hr zatražio stranicu /resursi.html u određenom vremenu, a na tu stranicu je upućen s Google-a. Drugi zahtjev nam pokazuje kako je isti korisnik (identična IP adresa) nakon određenog vremena zatražio drugu stranicu u ovom slučaju predavanja.html.

2.2.2 Kolačići

Kolačići (engl. *cookies*) su podaci koji se pohranjuju u korisnikovom pregledniku i koriste se za identifikaciju korisnika i personalizaciju sadržaja. Prilikom ponovnog posjeta na Web mjesto, podaci pohranjeni u kolačiću se koriste za identifikaciju korisnika i dohvaćanje postavki za personalizaciju. Kolačići pomažu korisnicima jer se u njima pohranjuju podaci o prijavi tako nije potrebno da se korisnik svaki put prilikom osvježavanja stranice mora ponovo prijaviti. Također, kolačići se mogu koristiti i za prikaz specifičnih reklama i ponuda. Ova opcija je zanimljiva ako poduzeće ima novi proizvod ili uslugu pa već postojećim korisnicima žele ponuditi iste. Ovisno o korisnikovim postavkama sigurnosti, kolačići se mogu i odbijati ili brisati svakim zatvaranjem preglednika. U tom slučaju kolačići nisu prigodni jer nisu pouzdani i potrebno je odabrani drugu metodu za identifikaciju korisnika.

2.2.3 Označavanje Web stranica

Metoda označavanja stranica je zapravo ugradnja malog programskog koda, najčešće u JavaScriptu na svakoj stranici koju želimo pratiti. Označavanje Web stranica je najpreciznija metoda jer je moguće prikupiti najviše informacija o korisničkim aktivnostima na stranici. Tako je moguće dobiti informacije o razlučivosti zaslona, vrsti preglednika, svaki potez pokazivača preko sadržaja i puno više. Jedan od primjera takvog koda je Google Analytics.

```
<!-- Google Analytics -->
<script>
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
})(window,document,'script','https://www.google-analytics.com/analytics.js','ga');

ga('create', 'UA-XXXXX-Y', 'auto');
ga('send', 'pageview');
</script>
<!-- End Google Analytics -->
```

Slika 1. Google Analytics kod za označavanje Web stranica. Izvor: developers.google.com

2.2.4 Hibridne metode

Nedostaci i prednosti prethodno navedenih metoda prikupljanja podataka o pregledavanju doveli su do razvoja hibridnih metoda. Tako se vrlo rijetko koristi samo jedna metoda, već dvije ili više. Spoj podataka iz poslužiteljskih zapisa i označavanje Web stranica jedna je od hibridnih metoda za identifikaciju korisnika i pregleda stranica. Moguće je i probati jednu metodu pa ako se ona pokaže neuspješnom, prijeći na drugu i tako dalje.

2.3. Proces Web analitike

Proces Web analitike sličan je procesu rudarenja podataka. Analiza Web mjesta je prvi korak. Analizira se struktura i sadržaj Web mjesta, određuju koji podaci su prigodni za prikupljanje, obradu i koji će bit od koristi za donošenje odluka. U ovom koraku određuje se i metoda kojom će se podaci prikupljati, hoće li se koristiti više izvora i određuju se osobe koje će provesti rudarenje podataka. Idući korak je početak prikupljanja podataka, ako su podaci prethodno prikupljeni, nema potrebe to raditi ponovo. U stvarnom svijetu, trendovi se konstantno mijenjaju, stoga nije poželjno

koristiti stare podatke. Nakon što su podaci prikupljeni, pristupa se čišćenju podataka. Poslužiteljski zapisi se čiste od nepotrebnih informacija, podaci različitih formata i prikupljeni iz nekoliko izvora se formatiraju u prikladne formate i sjedinjuju u jednu bazu podataka, skladište podataka ili u oblik koji je potreban za odabrane metode rudarenja. Nakon što su podaci obrađeni i pohranjeni, pristupa se procesu rudarenja podataka. U ovom koraku koriste se metode i algoritmi kao i u klasičnom rudarenju podataka. Osim klasičnih algoritama za rudarenje po podacima postoje i specijalizirani algoritmi koji se koriste u posebnim slučajevima, ako to ciljevi analize zahtijevaju. Informacije dobivene nakon procesa rudarenja se pohranjuju, interpretiraju i vizualiziraju. Rezultate će ekspertna osoba ili tim analizirati i donijeti odluku ili će znanje prezentirati nadređenima koji će onda donijeti odluku.

Posljednji korak u procesu rudarenja Weba je implementacija znanja dobivenog iz procesa Web analitike. Nakon implementacije znanja, na primjer, procesom Web analitike otkriven je podatak da je početna stranica često i izlazna stranica. Odnosno, korisnici nakon početne stranice odustaju od pregledavanja. Razlog tome može biti loš sadržaj, nepregledna struktura stranice, neispravno prikazivanje na starim ili mobilnim preglednicima. Procesom Web analitike utvrđene su slabe točke Web mjesta te su iste dorađene promjenom sadržaja, nadogradnji sučelja i slično. Nakon što su preinake ugrađene, proces Web analitike kreće ponovo s prikupljanjem podataka, obradom, rudarenjem i dobivanjem znanja. U drugom krugu, moguće je usporediti jesu li preinake bile uspješne, odnosno, ostaju li posjetitelji na Web mjestu i nakon početne stranice.

Slika 2. Proces Web analitike. Izvor: autor

3. Rudarenje podataka

„Rudarenje podataka (engl. *data mining*), odnosno otkrivanje znanja u bazama podataka (engl. *knowledge discovery in databases*) je netrivijalan postupak pronalaženja novih, valjanih, razumljivih i potencijalno korisnik oblika podataka. Pod oblikom podataka misli se na neko otkrivenu pravilnost među podatkovnim varijablama.“[5] Na primjer, prikupljanjem podataka s Web mjesta moguće je pouzdano pratiti uspješnost marketinške kampanje ili proizvoda. „Ako se otkrivene pravilnosti odnose na sve podatke, radi se o otkrivenom modelu, a ako se pravilnost odnosi samo na dio populacije podataka, radi se o otkrivenom uzorku (engl. *pattern*).“[6]

Rudarenje podataka vrši se nad velikom količinom podataka. Tradicionalno, podaci su pohranjeni u relacijskoj bazi podataka. Za potrebe otkrivanja uzorka, modela i analizu ogromne količine podataka, podaci su pohranjeni u skladištu podataka. Podaci nad kojima se vrše metode rudarenja mogu biti nestrukturirani i polustrukturirani.

Rudarenje podataka je relativno novo i mlado multidisciplinarno područje koje uključuje: baze podataka, statistiku, ekspertne sustave, umjetnu inteligenciju, teoriju informacija i još druga područja. Veliku primjenu pronalazi upravo u zadnjih nekoliko godina gdje se rudarenje podataka počelo koristiti svakodnevno. Rudarenje podataka se danas koristi za: praćenje i prognoziranje vremena, marketing, optimizaciju sustava, praćenje gospodarskih i ekonomskih trendova, društvene mreže, pametne kuće, IoT[7] , učinkovitost lijekova i u posljednje vrijeme za otkrivanje terorističkih prijetnji analiziranjem ponašanja korisnika na društvenim mrežama.

3.1. Metode rudarenja podataka

Postoji puno metoda rudarenja podataka, neke su za specifične zadatke i ovise o vrsti analize koju želimo provesti i o formatu u kojem se podaci prikupljaju. Najčešće korištene metode za rudarenje podataka su: asocijativna pravila (*association rules*), klasifikacijske metode (*classification rules*), metode klasteriranja (*clusters*) i stabla odlučivanja (*decision trees*).

---

[5] Panian Željko i suradnici,Poslovna inteligencija, Narodne novine, Zagreb, 2007., str. 148.
[6] Loc. cit.
[7] IoT, engl. *Internet of things.* Mreža raznih uređaja koji imaju senzore i prikupljaju informacije o okolišu

3.1.1 Asocijativna pravila

Asocijativnim pravilima pronalazimo zakonitosti o kupnji artikala u trgovinama, slijedu posjeta korisnika na Web mjestu i slično. Otkrivanje asocijativnih pravila dobivamo vjerojatnost kojim se proizvodi kupuju zajedno ili posjećuju određene stranice na Web mjestu.

1	Početna, o nama, kontakt
2	Početna, kontakt
3	Ponuda, početna, o nama, kontakt
4	Početna
5	O nama, početna, kontakt
6	Ponuda, kontakt,
7	Ponuda, početna, o nama, kontakt

Tablica 2. Primjer transakcija. Izvor: autor

Ako odredimo minimalnu podršku 30% i minimalno pouzdanje 80%, asocijativno pravilo:

Početna, O nama → Kontakt

vrijedi jer: podrška = 4/7 = 0,57 * 100 = 57% i pouzdanje = 4/4 = 1 * 100 = 100%

Dobivena podrška je 57% što je i preko zadane minimalne podrške što zadovoljava prvi uvjet, a to je da se stranice „Početna“ i „O nama“ u preko 30% slučajeva nalaze u istoj sesiji korisnika. Pouzdanje dobiveno pouzdanje je u ovom slučaju 100% jer se u odabranim sesijama Početna, O nama i Kontakt nalaze zajedno. Iz ovog pravila možemo zaključiti da početna stranica Web mjesta funkcionira jer korisnici nastavljaju pregledavati druge stranice i ostvaruje se cilj, a to je slanje upita ili kontaktiranje.

3.1.2 Klasifikacijske metode

Cilj klasifikacijske metode je odrediti zajedničke atribute objektima i svrstati ih u grupe ili klase. Objekti se grupiraju prema zadanom atributu.

3.1.3 Metode klasteriranja

Metodom klasteriranja objekti sličnih atributa se grupiraju u klastere. Za razliku od

klasifikacije, kod klasteriranja, prethodno nisu poznati atributi po kojima će objekti biti svrstani u klastere. Najčešće korišteni algoritmi za klasteriranje su K-means i hijerarhijsko klasteriranje.

3.1.4 Stabla odlučivanja

„Stablima odlučivanja simbolički se prikazuju situacije pri odlučivanju. Iz podataka poznatih situacija i poznatih odluka konstruira se stablo očekivanja, koje se kasnije može upotrebljavati u novim situacijama. U tablici 2. prikazani su podaci već razmatranih zahtjeva za kreditom: spol tražitelja kredita, njegova starost u godinama, broj godina na sadašnjoj adresi, podatak je li tražitelj vlasnik ili korisnik nekretnine u kojoj živi, zanimanje, broj godina radnog staža, koliko je komitent banke, koliki su njegovi mjesečni troškovi, i u zadnjem stupcu odluka banke o zahtjevu za kreditom.“[8]

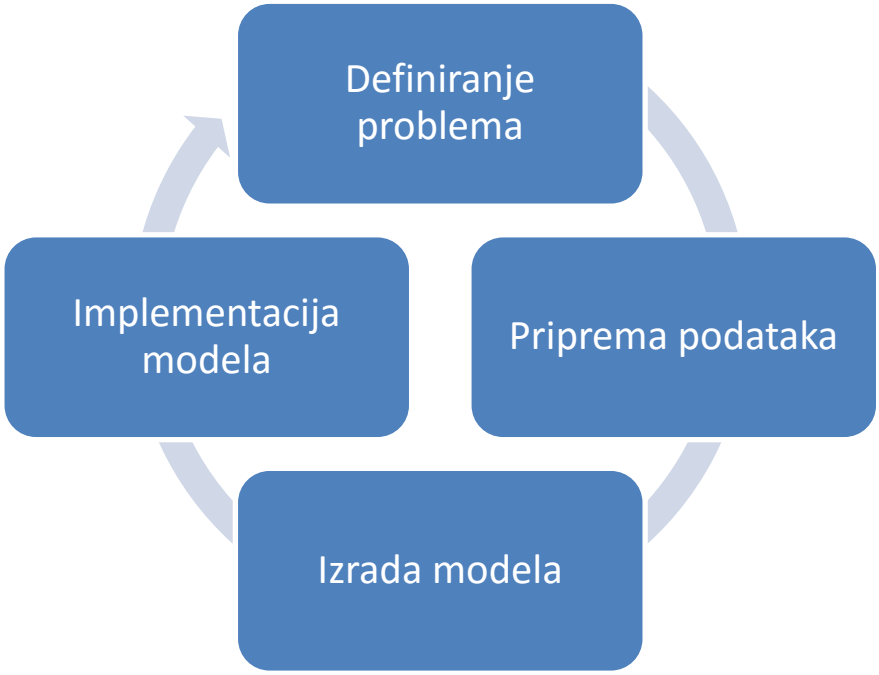
Spol	Starost u god.	Broj god. na adresi	Vlasnik nekretnine	Zanimanje	Godina staža	Komitent u god.	Mjesečni troškovi	Zahtjev
M	50	0,5	Vlasnik	Nezaposlen	0	0	145	Odbijen
M	19	10	Korisnik	Radnik	0,8	0	140	Odbijen
Ž	52	15	Vlasnik	Pisac	5,5	14	000	Odobren
M	22	2,5	Korisnik	Učitelj	2,6	0	000	Odobren
M	29	13	Vlasnik	Vozač	0,5	0	228	Odbijen
Ž	16	0,3	Vlasnik	Nezaposlen	0	1	160	Odbijen
M	23	11	Vlasnik	Ekonomist	0,5	1	100	Odobren
Ž	27	3	Vlasnik	Menadžer	2,8	1	280	Odbijen
Ž	19	5,4	Vlasnik	Kuhar	0,3	0	080	Odbijen
Ž	27	0,3	Vlasnik	Menadžer	0,1	1	272	Odbijen
M	34	4	Korisnik	Konobar	8,5	7	195	Odobren
M	20	1,3	Korisnik	Radnik	0,1	0	140	Odbijen
M	34	1,3	Vlasnik	Prodavač	0,1	0	440	Odbijen
…	…	…	…	…	…	…	…	…

Tablica 3. Stablo odlučivanja. Izvor: Informacijska tehnologija u poslovanju

3.2. Proces rudarenja podataka

Proces rudarenja podataka sastoji se od nekoliko koraka, vrijeme potrebno za svaki korak ovisi o podacima i zadanom problemu. Proces počinje definiranjem problema, drugi korak je priprema podataka a uključeni zadaci su: selekcija, transformacija uzorkovanje i evaluacija podataka. Treći korak je modeliranje podataka, a posljednji korak je implementacija modela, odnosno interpretacija i korištenje dobivenih rezultata.

[8] Čerić Vlatko, Varga Mladen, Informacijska tehnologija u poslovanju, Element, Zagreb, 2004., str. 215.



Slika 3. Proces rudarenja podataka. Izvor: autor

3.2.1 Definiranje problema

Definiranje problema prvi je korak u procesu rudarenja podataka. Na primjer, politička stranka želi osvojiti što više glasova na parlamentarnim izborima. Cilj je izraditi model koji će predviđati grupe građana koji su zainteresirani i upoznati sa strankom, ali i za one potencijalne glasače. Dobiven model građana upotrijebit će se u oglašavanju, propagandi, nadogradnju Web mjesta. Radna skupina zadužena za rudarenje podataka sastoji se od analitičara, odgovorne osobe iz odjela promidžbe i osoba iz vodstva koja će definirati daljnje ciljeve.

3.2.2 Priprema podataka

Priprema podataka oduzima najviše od cijelog procesa rudarenja podataka. Izvori podataka mogu biti Web, društvene mreže, telefonske ankete, promet na Web mjestu i slično. Iako se podaci najčešće pohranjuju u relacijskim bazama podataka, u vrijeme društvenih mreža, podaci mogu biti strukturirani i polustrukturirani. Nastavak prethodnog primjere o političkoj kampanji. Uz ostale ciljeve rudarenja podataka, žele se dodatno prikupiti komentari s društvenih mreža i dnevnih portala kako bi se dobio opći dojam stranke na internetu. Prikupljeni podaci s društvenih mreža i dnevnih portala su nestrukturirani podaci, a uz tekst komentara, prikupljene su i informacije o korisničkim profilima. Dobiveni podaci se zatim čiste i transformiraju u format koji je

prigodan za alate za rudarenje po podacima. Uzorkovanje podataka je uzimanje slučajnog uzorka iz cijelog skupa podataka. Broj podataka za uzorak ovisi o problemu i metodama koje je potrebno izvesti. U fazi evaluacije podataka izbacuju se podaci koji nisu prigodni za rudarenje podataka.

3.2.3 Izrada modela

Za izradu modela glasača koriste se metode rudarenja podataka koje se mogu staviti u tri osnovne kategorije: klasifikacija, previđanje i asocijacija. Odabir metode ovisi o cilju rudarenja podataka. Za prethodni primjer, potrebno je otkriti segment pa će se koristiti metoda otkrivanja. Ako je to moguće, testira se nekoliko metoda i odabere najbolja.

3.2.4 Implementacija

Posljednja faza je implementacija i korištenje rezultata procesa rudarenja podataka. Za prikaz rezultata izrađuju se izvješća a po potrebi se koriste dodatni alati za vizualizaciju. Rezultati procesa trebaju biti jednostavni za čitanje i primjenu. U primjeru s političkom strankom, rezultati rudarenja i analize komentara na društvenim mrežama su tablice u kojima je vidljiv pozitivan, neutralan ili negativan komentar o stranci uz prikaz demografskih podataka kao što su: spol, dob, lokacija, interesi i slično. Rezultati se prezentiraju vodstvu koje će donijeti odluku o marketinškoj strategiji i ostalim ciljevima.

4. Rudarenje Weba

Rudarenje Weba je proces pronalaženja, zahvaćanja i analiziranja podataka s Interneta. Postupcima rudarenja Weba od “sirovih“ podataka, uz pomoć odgovarajućih alata, dobivaju se vrijedne informacije i znanja koja se mogu koristiti kao pomoć pri donošenju odluka. Tri su najvažnija pristupa rudarenju Weba: rudarenje strukture, rudarenje obrazaca ponašanja posjetitelja Web mjesta i rudarenje sadržaja.

4.1. Rudarenje strukture

Rudarenje strukture otkriva korisno znanje iz hiperpoveznica na Web mjestu. Primjerice, prateći poveznice s jednog Web mjesta možemo doći do drugog i tako otkriti kako su Web mjesta povezana. Tražilice koriste ovu metodu kako bi otkrile koliko drugih relevantnih Web mjesta ima upućenu vezu prema drugom i koliko su te veze jake. Na temelju toga, tražilice raspoznaju relevantne i one manje relevantne stranice.

„Istraživanjem, odnosno rudarenjem globalne strukture Weba može se doći do brojnih spoznaja o tome tko se s kime povezuje, na koji način, kako često Web mjesta komuniciraju, koje se vrste informacijskih sadržaja razmjenjuju, gdje dolazi do problema, itd., a potom i analizirati zašto to jest, jest takvo kako jest ili zašto se nešto zbiva tako kako je utvrđeno rudarenjem Weba.“[9]

„Istraživanje, odnosno rudarenje lokalne strukture Web mjesta važan je prvi korak prema razumijevanju kako se Web mjesto koristi i kako ga je moguće poboljšati, odnosno učiniti privlačnijim za korisnike. Mnoge poznate metrike korištene za ocjenu učinkovitosti pojedinačnih Web stranica nemaju previše smisla ako se ne zna kako ta Web stranica funkcionira u svojoj neposrednoj okolini, tj. u strukturi cjelokupnog Web mjesta koja može obuhvaćati na stotine ili tisuće sličnih ili različitih Web stranica. Za ilustraciju, adhezivnost ili “ljepljivost“ (engl. Stickiness) stranica je pokazatelj koji se odnosi na vrijeme koje posjetitelj Web mjesta “provodi“ na jednoj stranici. Općenito, što je to vrijeme duže, to se Web stranica smatra zanimljivijom.“[10]

Razlikujemo navigacijske i destinacijske Web stranice. Glavni cilj navigacijskih stranica je uputiti korisnike na lokaciju na kojoj se nalazi informacija koja je njemu potrebna.

---

[9] Čerić Vlatko, Varga Mladen, Informacijska tehnologija u poslovanju, Element, Zagreb, 2004., str. 154
[10] loc. cit.

Navigacijske stranice upućuju na destinacijske na kojima se nalazi tražena informacija ili resurs. Tako je poželjno što manje vremena zadržavanja na navigacijskim stranicama, a suprotno tome, što dulje vrijeme zadržavanja na destinacijskim stranicama. Alati za rudarenje i analitiku Weba mogu utvrditi vrijeme koje korisnik provodi na određenoj stranici. lako, dulje vrijeme provedeno na određenoj stranici može upućivati na nelogičnu strukturu stranice pa je korisniku potrebno više vremena za pronaći traženu informaciju.

„Lokalna struktura Web mjesta u velikoj je mjeri ovisna o njegovoj namjeni. To je podjednako vrijedi za sva Web mjesta od onih najjednostavnijih koja se, poput online kataloga, sastoje od svega nekoliko sličnih stranica, pa do onih vrlo kompleksnih i hijerarhijskih ustrojenih, kao što su Web mjesta konzultantskih tvrtki ili velikih korporacija. Zato i ne postoji nikakav općenit, jednostavan recept koji bi definirao nešto poput optimalne ili univerzalno primjenjive strukture Web stranica. Takva, naime, lokalna struktura naprosto se postoji pa će od slučaja do slučaja trebati pronalaziti najbolja rješenja. Upravo ovdje mogućnosti rudarenja strukture Weba dolaze do punog izražaja.“[11]

4.2. Rudarenje ponašanja posjetitelja Web mjesta

Rudarenje ponašanja posjetitelja Web mjesta je primjena metoda rudarenja podataka nad prikupljenim podacima o korisnikovom ponašanju kako bi se otkrili korisni uzorci. Svaka se korisnikova akcija evidentira u obliku HTTP zahtjeva na poslužitelju, u kolačićima ili uz poseban programski kod namijenjen za praćenje svakog poteza korisnika. Rezultati rudarenja ponašanja posjetitelja Web mjesta služe kao pomoć poduzećima u utvrđivanju vrijednosti kupaca, promidžbu, itd.

4.3. Otkrivanje i analiza uzoraka iz posjeta Web mjesta

Vrsta analize koja će se izvodit na dobivenim i pročišćenim podataka ovisi o krajnjem cilju analize i željenom rezultatu.

4.4. Analiza sesija i posjetitelja

Statistička analiza je najčešća metoda analiziranja prikupljenih podataka. U ovom

[11] Ibidem. str. 155,

slučaju, podaci su agregirani u predodređene jedinice kao što su dani, sesije ili korisnici. Uobičajene statističke tehnike se ovdje mogu upotrijebiti kako bi se dobilo znanje o posjetiteljima. Ovaj pristup koriste alati za analizu zapisa s poslužitelja. Takvi izvještaji sadrže informacije kao što su: često posjećene stranice, prosječno vrijeme provedeno na svakoj stranici, prosječni tijek pregledavanja Web mjesta, ulazne i izlazne stranice. Iako ovim podacima nedostaju detalji koje može dati detaljno rudarenje i otkrivanje uzoraka, potencijalno znanje može biti korisno za donošenje poslovnih odluka.

Online Analytical Processing (OLAP) pruža veću fleksibilnost i detaljniju obradu podataka od samih statističkih metoda. Izvor podataka za OLAP je najčešće multidimenzionalno skladište podataka koje integrira korištenje Web mjesta, sadržaj, transakcijske podatke na različitim razinama agregacije za svaku dimenziju. OLAP alati omogućuju promjenu razine agregacije po svakoj dimenziji u bilo kojem trenutku za vrijeme analize. Rezultati OLAP upita se mogu koristiti kao ulazne vrijednosti za alate za rudarenje i vizualizaciju podataka.

4.5. Analiza klastera i segmentacija posjetitelja

Klasteriranje je metoda rudarenja podataka koja grupira objekte sličnih atributa u grupe (klastere). Klasteri korisnika ili klasteri stranica su dvije grupe koje se koriste pri analizi Web mjesta. Klasteriranje korisničkih podataka, koje uključuju sesije ili transakcije dovodi do otkrivanja grupa korisnika koji imaju slično ponašanje prilikom pregledavanja Web mjesta. Dobiveno znanje se može upotrijebiti za e-trgovinu i sustave preporuka. Transakcijski klasteri mogu predstavljati korisnike ili segmente posjetitelja na temelju navigacije i transakcija. Međutim, sami transakcijski klasteri nisu pogodni za generiranje detaljne slike korisničkih uzoraka.

4.6. Analiza asocijacija i korelacija

Analiza pravila pridruživanja i analiza statističke korelacije mogu pronaći grupe proizvoda ili stranica kojima se često pristupa zajedno ili se zajedno kupuju. Krajnji rezultat toga je mogućnost da Web mjesta efikasnije postave sadržaj ili kako bi se ponudile dodajte preporuke prilikom kupovine.

4.7. Analiza sekvencijalnih i navigacijskih uzoraka

Metoda otkrivanja navigacijskih uzoraka teži pronaći stranice koje se uobičajeno

pregledavaju i time predvidjeti buduće posjetitelje na istim stranicama kako bi im se tamo ponudile reklame i druge usluge.

5. Web metrike

U Web analitici se mogu razlikovati dva tipa metrike, to su brojevi (*count*) i omjeri (*ratio*). Iz metrika su isključeni roboti tražilica, tako da sve dimenzije dobivene od posjetitelja. Četiri su osnovne metrike: jedinstveni posjetitelji (*unique visitors*), pregledi stranica (*pageviews*), događaji (*events*) i posjete (*visits/sessions*).

Jedinstveni posjetitelji su metrika izražena u obliku broja koja govori koliko je osoba posjetilo Web mjesto u zadanom vremenom periodu. Identifikacija jedinstvenih posjetitelja se odrađuje uz korištenje kolačića. Posjeta Web mjestu je svaki HTTP zahtjev za pregledom jedne stranice, a traje sve dok korisnik ne napusti stranicu, zatvori preglednik ili nakon 30 minuta neaktivnosti. Pregledi stranice predstavljaju koliko je određena stranica pregledana od strane korisnika. Uz metriku pregleda stranici po posjeti, dobivaju se podaci o broju pregledanih stranica od strane posjetitelja za vrijeme trajanja posjete Web mjestu. Događaj je svaka akcija koju korisnik izvršni na stranici. Pregled multimedije, obavljanje transakcije, stavljanje proizvoda u košaricu, prijava na newsletter, slanje kontakt forme, interakcija s oglasima i slično.

Ulazna stranica je Web stranica koju korisnik prvu vidi pri posjeti Web mjestu. Izlazna stranica je posljednja stranica koju korisnik vidi na kraju sesije. Trajanje posjeta je metrika koja mjeri korisnikovo vrijeme provedeno na Web mjestu, a koristi se za analizu kvalitete sadržaja. Vrijeme korisnikove aktivnosti se računa pregledom poslužiteljskih zapisa u kojima se analizira vrijeme svakog HTTP zahtjeva prema poslužitelju.

Problem nastaje kod izlaznih stranica jer nije moguće odrediti koliko vremena je korisnik pregledavao izlaznu stranicu jer na izlazu nije upućen HTTP zahtjev prema poslužitelju. Zato je pri analizi najbolja praksa isključiti sesije koje obuhvaćaju pregled samo jedne stranice. Izvor prometa ili kanali pokazuju izvor kojeg je upućen korisnik na određeno Web mjesto. Razlikujemo: unutarnji i vanjski izvor prometa, tražilice, društvene mreže i direktan promet.

Novi posjetitelji je metrika koja mjeri broj jedinstvenih posjetitelja koji su pregledavali Web mjesto u zadanom vremenu. *Returning visitor* je metrika koja mjeri koliko je jedinstvenih posjetitelja pregledavalo Web mjesto u zadanom vremenskom periodu, a uvjet je taj da je njihov prvi posjet bio prije zadanog vremena. *Repeat visitor* je metrika

koja prikazuje broj korisnika koju su pregledavali Web mjesto dva ili više puta u zadanom vremenskom periodu.

Bounce rate je omjer sesija koje su prestale nakon jednog prikaza stranice i ukupnog broja posjeta. Visok *bounce rate* govori da sadržaj na stranici nije „zanimljiv“, može ukazati na problem učitavanja stranice, bilo to neispravno prikazivanje na mobilnim uređajima ili sporo vrijeme učitavanja fotografija, sadržaja i ostalog.

6. Analiza i rudarenje posjeta Web mjesta na primjeru

Weka je program za rudarenje podataka i vizualizaciju rezultata. Izvor podataka za primjer je proizvoljan i pohranjen u *arff* formatu. U *arff* formatu razlikuju se dvije sekcije. Prva je *header* ili glava, gdje se nalaze informacije o nazivu relacije i popis atributa i tip podataka. Nakon glave dolazi *data* sekcija koja sadrži podatke.

```
@relation web_log

@attribute posjetitelj real
@attribute stranica {/, /o-nama, /ponuda, /kontakt, /ponuda/usluge, /ponuda/proizvodi, /blog, /social}
@attribute red real
@attribute referral {organic, social, direct}

@data
1,/,0, organic
1,/o-nama,1, organic
1,/ponuda,2, organic

2,/,0, direct
2,/o-nama,1, direct
2,/ponuda,2, direct
2,/,3, direct
2,/o-nama,4, direct

3,/,0, organic
3,/o-nama,1, organic

4,/,0, organic
4,/o-nama,1, organic
4,/ponuda,2, organic
4,/,3, organic

5,/,0, organic

6,/,0, organic
6,/ponuda,1, organic
6,/ponuda/usluge,2, organic
6,/kontakt,3, organic

7,/,0, direct
7,/ponuda,1, direct
7,/ponuda/proizvodi,2, direct
7,/o-nama,3, direct
```

Slika 4. Uzorak podataka. Izvor: autor

Atributi podatka iz primjera su sljedeći:

@attribute posjetitelj real	Jedinstveni identifikator posjetitelja u obliku realnog broja.
@attribute stranica {/, /o-nama, /ponuda, /kontakt, /ponuda/usluge, /ponuda/proizvodi, /blog, /social}	Atribut stranice koji može imati vrijednosti / - početna stranica /o-nama – standardna o nama stranica /ponuda – stranica s ponudama /kontakt – stranica s kontakt podacima /ponuda/usluge – podstanica stranica ponuda /ponuda/proizvodi – podstranica stranice ponuda /blog – blog stranica /social – stranica s poveznicama za društvene mreže

@attribute red real	Atribut red. Broj koji označava posjetiteljevu navigaciju kroz Web mjesto. Počinje od nule koja označava prvu stranicu.
@attribute referral {organic, social, direct}	Atribut referer koji može imati jednu od tri vrijednosti. Organic za posjete putem organskog pretraživanja. Social za promet s društvenih mreža i direct za direktan promet.

Tablica 4. Popis i opis atributa. Izvor: autor

Podaci su pohranjeni u sljedećem formatu.

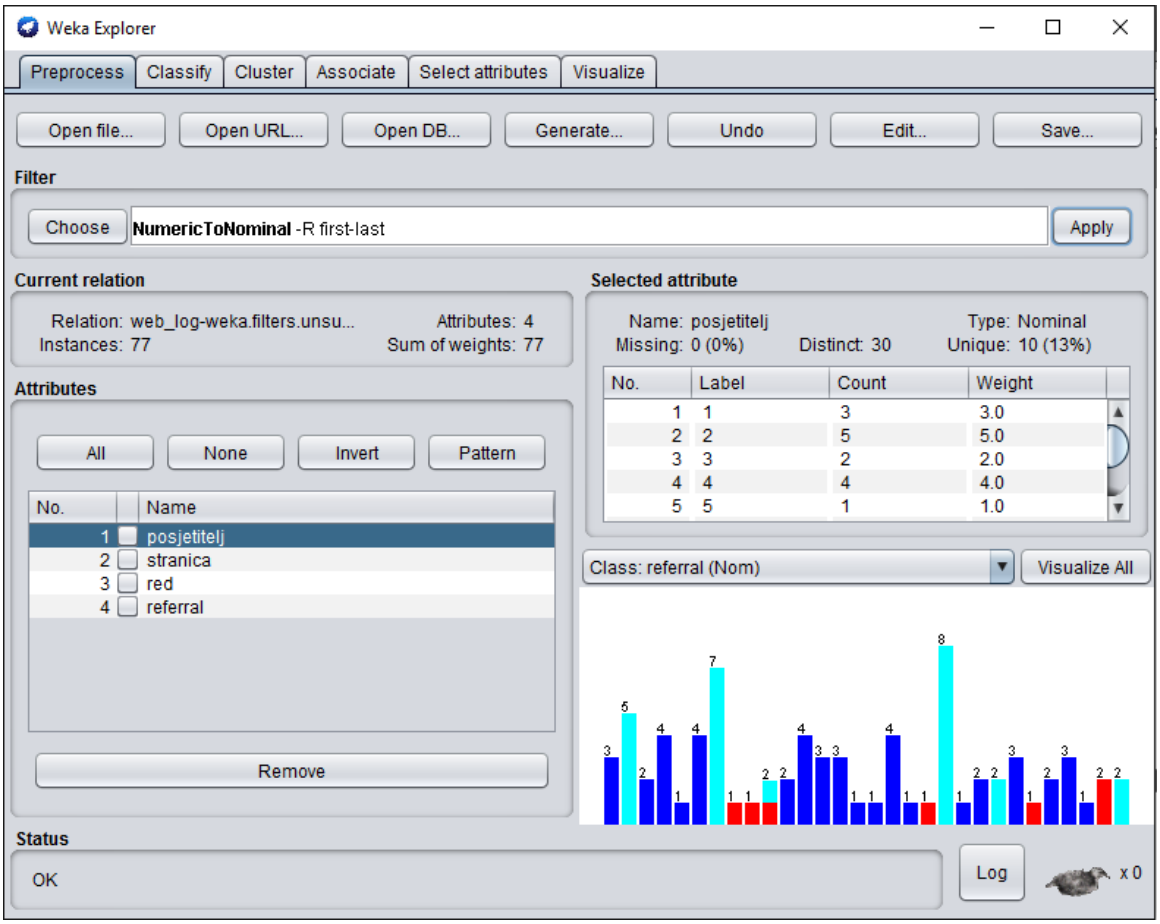
```
13  2,/,0, direct
14  2,/o-nama,1, direct
15  2,/ponuda,2, direct
16  2,/,3, direct
17  2,/o-nama,4, direct
```

Slika 5. Sesija. Izvor: autor

Prvi atribut je identifikator posjetitelja, drugi atribut je posjećena stranica, treći atribut je redoslijed navigacije i posljednji atribut je izvor prometa. Za ovaj primjer, posjetitelj broj dva je posjetio prvo početnu stranicu, zatim */o-nama*, */ponuda*, ponovo početnu stranicu i na kraju izlazna stranica je */o-nama*. Izvor prometa je direktan, odnosno posjetitelj je direktnim ukucavanjem Web adrese došao na Web mjesto.

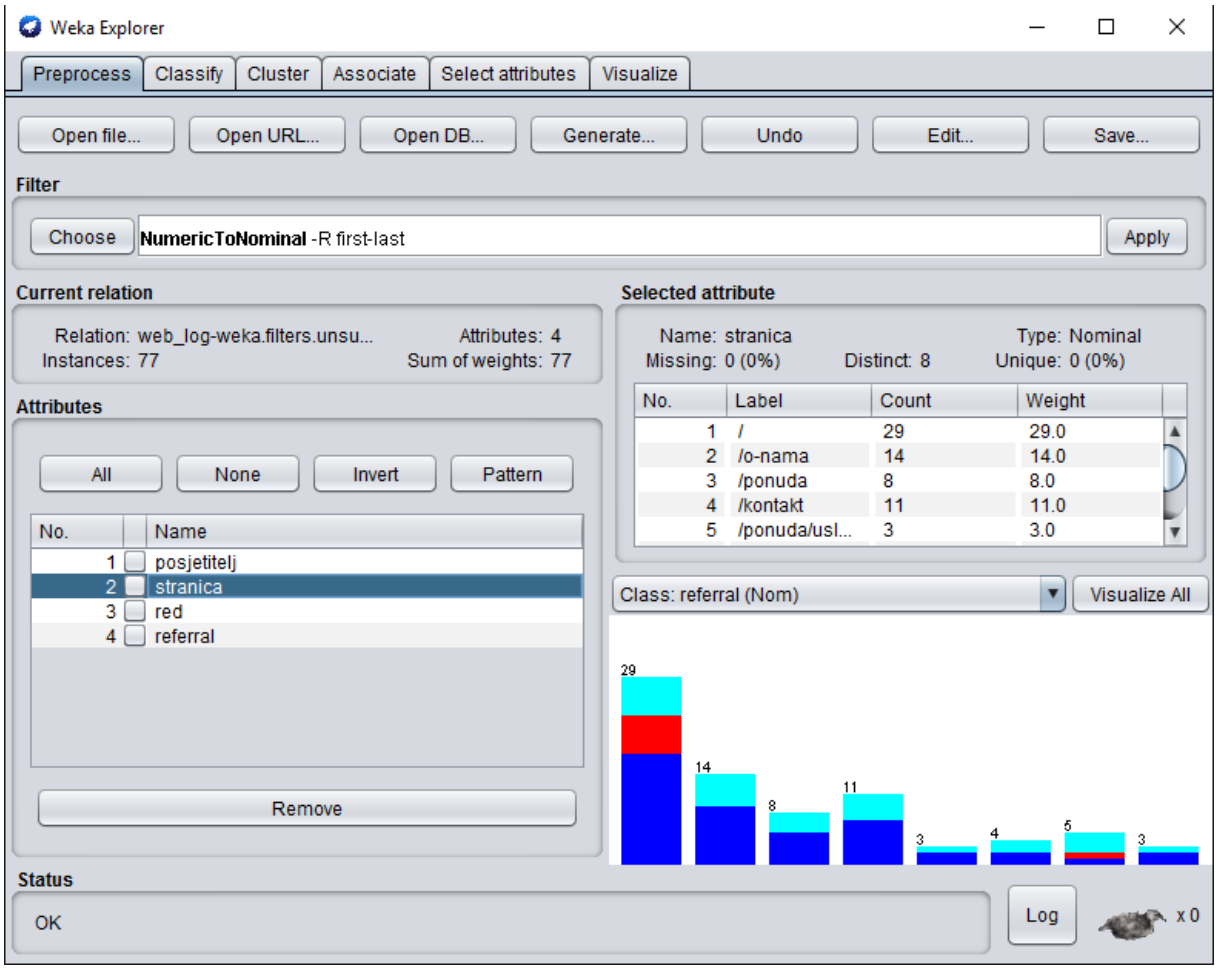
Podaci koji će se koristiti za rudarenje podataka u ovom primjeru sadrže 77 instanci za ukupno 30 posjetitelja. Ostavljeni su osnovne informacije jer cilj rudarenja podataka je pronaći uzorke posjetitelja i ostale korisne informacije.

Nakon učitavanja podataka u Weka-u, potrebno je izvršiti filtriranje nad atributom posjetitelj koristeći filter *NumericToNominal.*



Slika 6. Učitani podaci u Weka-u i pregled atributa posjetitelji. Izvor: autor

Korisne informacije o podacima su vidljive na *preprocess* tabu. Broj instanci je 77, a jedinstven broj instanci je 30, odnosno, broj jedinstvenih posjetitelja. Za svaki označen atribut, program vizualizira podatke ako je to moguće. Atribut posjetitelj govori o broju stranica koje je posjetitelj pregledavao za vrijeme jedne sesije. Minimalni broj za svaku sesiju je jedan i ne može biti nula.

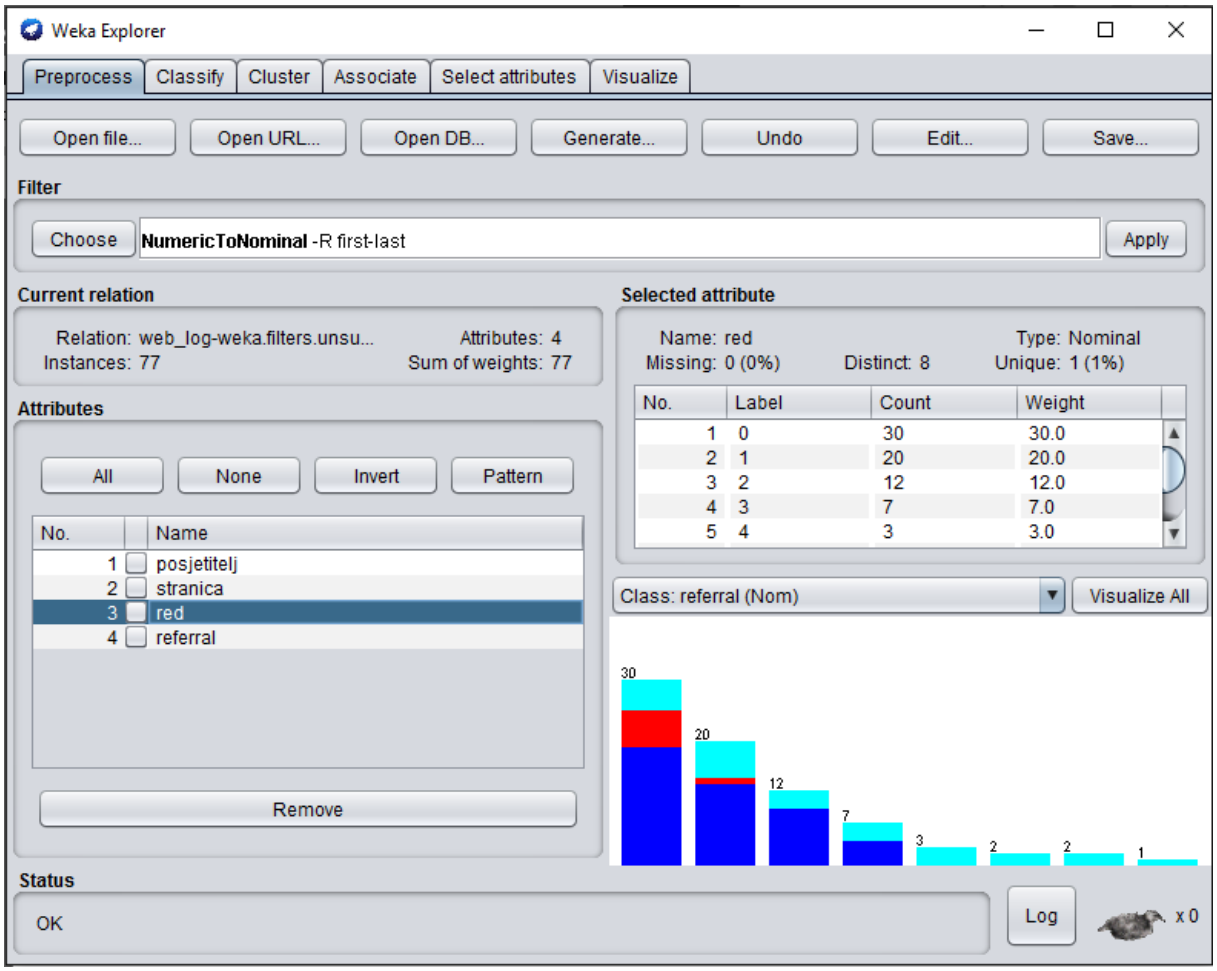


Slika 7. Atribut stranica. Izvor: autor

Atribut *stranica* govori broj posjeta na pojedinoj stranici. Najviše posjeta je naravno na početnoj stranici, iako je broj posjeta na početnoj stranici 29, odnosno jedan manje od ukupnog broja posjeta. Može se sa sigurnošću zaključiti da je ulazna stranica ujedno i početna stranica. S obzirom da je iduća stranica ima upola manje posjeta, možemo zaključiti da dobar dio korisnika napusti Web mjesto odmah nakon pregledavanja početne stranice. Razlog tome može biti da je posjetitelj zaključio da se ovdje ne nalaze informacije koje su mu potrebne, a uz druge metrike kako što su vrijeme učitavanja stranice i izgleda na različitim zaslonima, broj ispadanja sa stranica može biti nepregledan sadržaj, loš izgled Web mjesta i ostalo.

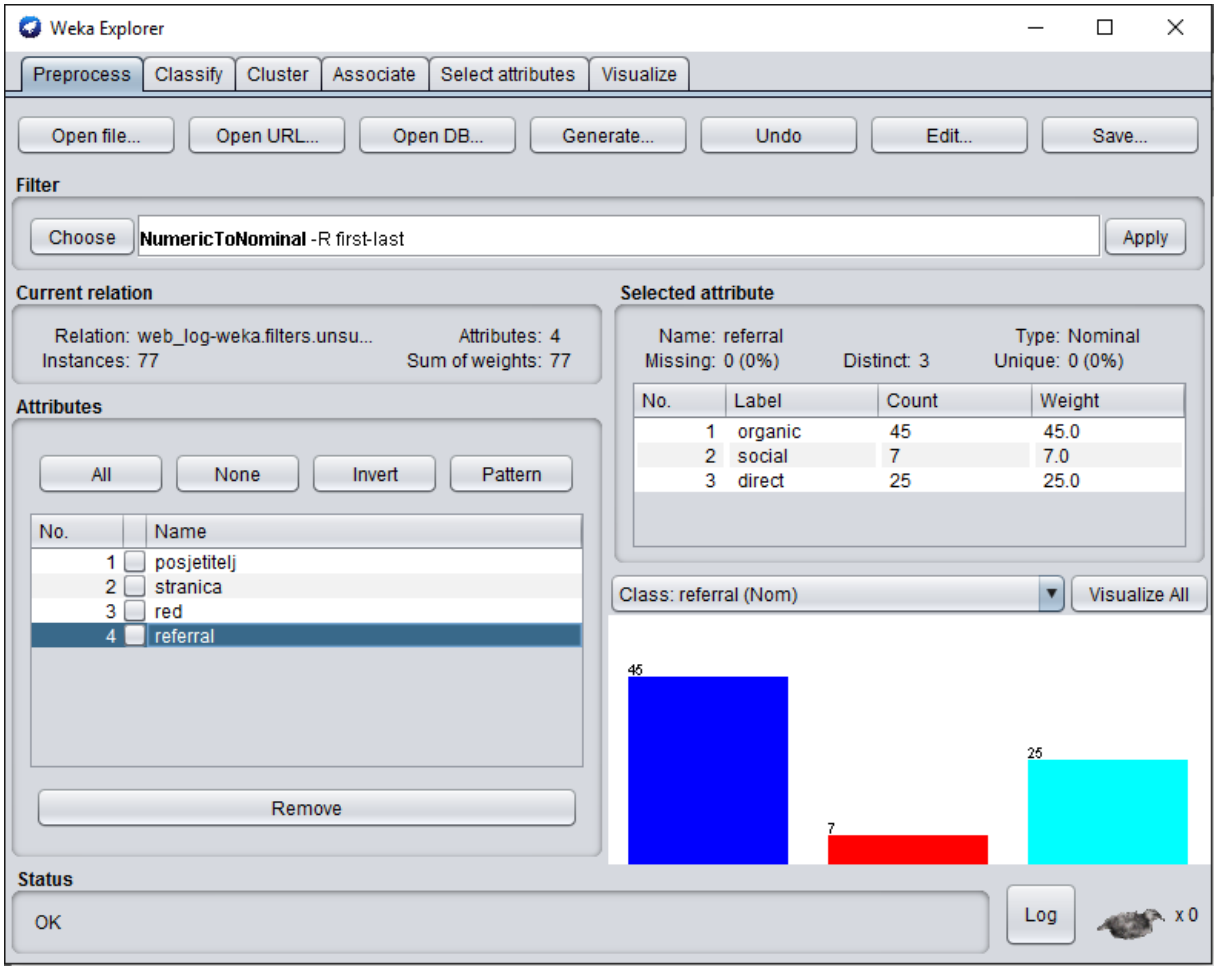
Broj	Stranica	Broj posjeta
1	/	29
2	/o-nama	14
3	/ponuda	8
4	/kontakt	11
5	/ponuda/usluge	3
6	/ponuda/proizvodi	4
7	/blog	5
8	/social	3

Tablica 5. Broj posjeta Web mjesta. Izvor: autor



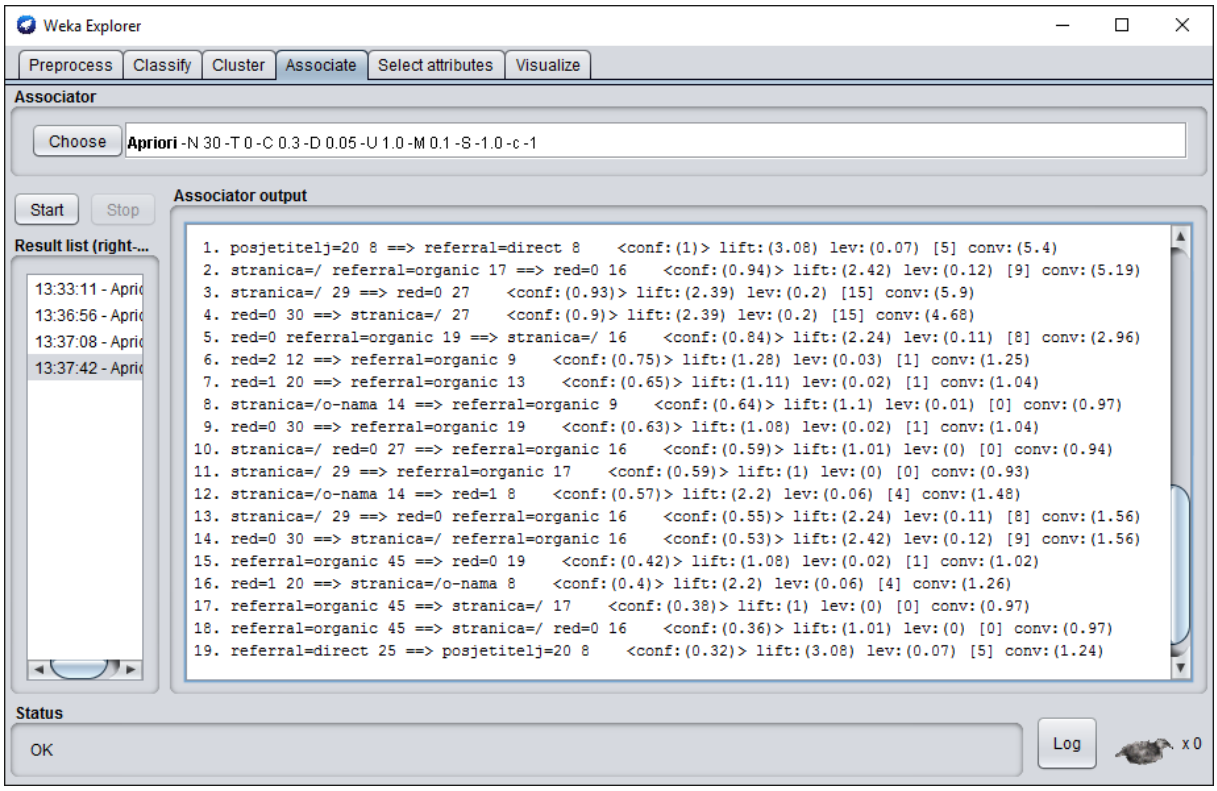
Slika 8. Pregled atributa red. Izvor: autor

Iz atributa *red* vidljivi su podaci o redoslijedu navigacije korisnika na Web mjestu. Razlika između reda nula i jedan iznosi deset. Odnosno, deset je korisnika nakon pregledne prve stranice napustio Web mjesto.



Slika 9. Atribut referral. Izvor: autor

Atribut *referral* pokazuje izvor prometa za svaku stranicu. Primarni izvor prometa je putem organskog pretraživanja, zatim direktan promet i na kraju promet upućen s društvenih mreža.



Slika 10. Apriori algoritam. Izvor: autor

Za otkrivanje pravila pridruživanja koristi se Apriori algoritam. Minimalna podrška je postavljena na 0.3. Algoritam je otkrio ukupno 19 pravila prema zadanim parametrima. Kao korisna pravila mogu se izdvojiti:

2. stranica=/ referral=organic 17 ==> red=0 16 <conf:(0.94)> lift:(2.42) lev:(0.12) [9] conv:(5.19)

Ulazna stranica je početna stranica i izvor prometa je organsko pretraživanje s podrškom od 0.94.

3. stranica=/ 29 ==> red=0 27 <conf:(0.93)> lift:(2.39) lev:(0.2) [15] conv:(5.9)

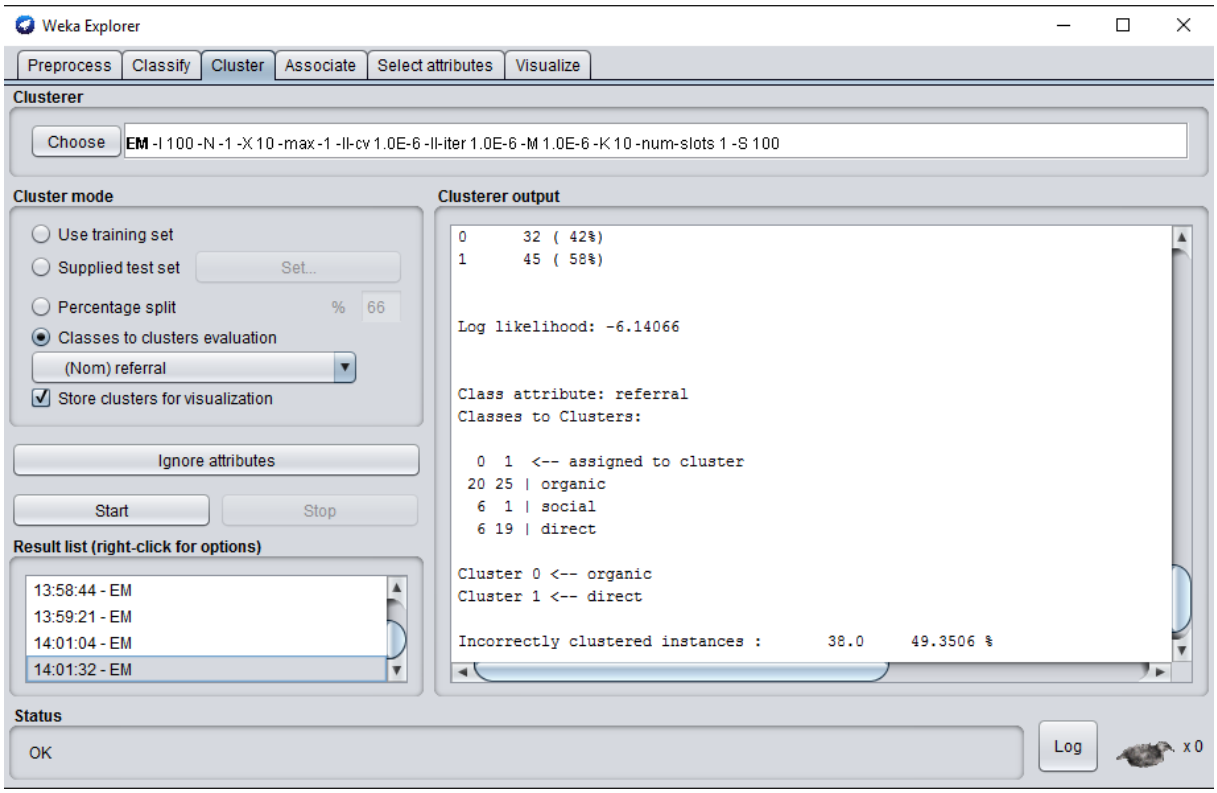
Početna stranica je ulazna stranica s podrškom 0.93

8. stranica=/o-nama 14 ==> referral=organic 9 <conf:(0.64)> lift:(1.1) lev:(0.01) [0] conv:(0.97)

Izvor prometa za stranicu /o-nama je organsko pretraživanje s podrškom 0.64

16. red=1 20 ==> stranica=/o-nama 8 <conf:(0.4)> lift:(2.2) lev:(0.06) [4] conv:(1.26)

S podrškom od 0.4, druga stranica koju posjetitelj pregleda je /o-nama



Slika 11. Rezultati EM klasteriranja. Izvor: autor

EM algoritam klasterirao je podatke prema zadanim parametrima. *Referral* je zadana klasa za evaluaciju klastera. Rezultat algoritma su dva klastera. Klaster 0 grupiran je prema organic a klaster 1 prema direkt *refferalu.*

Slika 12. Vizualizacija rezultata klasteriranja. Izvor: autor

Na X osi nalaze se dva prethodno generirana klastera od EM algoritma. Na Y osi su posjećene stranice. Iz grafičkog prikaza zaključuje se da posjetitelji koji su došli preko internetskih tražilica uglavnom posjećuju samo početnu stranicu, dok posjetitelji koji su došli putem izravnog prometa nastavljaju pregledavati i druge stranice.

7. Zaključak

Web je danas novo mjesto trgovine, a trgovine žele znati tko su njihovi posjetitelji i što oni traže i to im ponuditi. S tisuće i tisuće potencijalnih kupaca potrebno je automatizirati proces preporuka. Ovdje dolazi analiza i rudarenje posjeta Web mjesta, odnosno, identifikacija posjetitelja Web mjesta. Korištenjem nekoliko metoda i procesa iz nekoliko grana znanosti, rudarenje i analiza posjetitelja postaje zasebna disciplina.

Znanje koje se dobiva iz procesa analize i rudarenja posjeta Web mjesta može značiti uspješno ili negativno poslovanje. Osim poslovanja, analizom posjetitelja, svako Web mjesto može dobiti uvid u tip posjetitelja. Tradicionalno su se analizirali zapisi na poslužiteljima, međutim, danas se u realnom vremenu prikupljaju podaci iz više izvora, a podaci su uglavnom nestrukturirani.

Pojavljivanje rastuće količine korisničkih i transakcijskih podataka je dovela do potencijalne riznice znanja gdje, uglavnom trgovine, dobivaju uvid u potrebe svojih kupaca.

Primjena metoda i alata za rudarenje podataka pronalazi primjenu u svim granama života. Tako je moguće uz automatizirane procese konstantno i u realnom vremenu pratiti trendove na društvenim mrežama.

Cijeli proces počinje od definiranja problema, odnosno od postavljanja pitanja na koje tražimo odgovor. Na primjer: top turistička destinacija za iduću godinu, otkrivanje kriminalnih aktivnosti na društvenim mrežama, povećanje prodaje, analiza postojeće promidžbene kampanje. Cilj određuje metodu i tehnike koje će se koristiti za otkrivanje odgovora iz prikupljenih i obrađenih podataka. Nakon analize preostaje interpretacija informacija u znanje i donošenja odluka na temelju istih.

LITERATURA

[1] Panian Željko i suradnici, Poslovna inteligencija, Narodne novine, Zagreb, 2007.

[2] Čerić Vlatko, Varga Mladen, Informacijska tehnologija u poslovanju, Element, Zagreb, 2004

[3] Meta S. Brown, Data Mining for dummies, John Wiley & Sons, Inc., Hoboken, New Jersey, 2014

[4] Robert Manger, Miljenko Maručić, Skripta strukture podataka i algoritmi, Zagreb, 2007.

[5] Mohammed J., Wagner M., Data mining and analysis, Cambridge University Press, New York, 2014.

[6] Avinash Kaushik, Web Analytics 2.0, Wiley Publishing, Inc., Indianapolis, Indiana, 2010.

[7] Bing Liu, Web Data Mining, Springer-Verlag, Berlin, 2007.

POPIS SLIKA

POPIS TABLICA

SAŽETAK

Rudarenje posjeta Web mjesta relativno je novo područje koje se sve češće koristi u poslovnim procesima i pri donošenju odluka. Rudarenje podataka koristi se u većini grana: sigurnost, promet, trgovine, medicina, itd. Dok metode i alati za rudarenje Weba pronalaze uporabu za otkrivanje korisničkih uzoraka s Web mjesta ili mobilne aplikacije. Alati i metode proizašle iz rudarenja podataka posebno su prilagođene za probleme rudarenja Weba. Rezultati rudarenja koriste se najčešće u poslovne svrhe a primjenu pronalaze u optimizaciji Web mjesta, u sustavima preporuka i personaliziranom sadržaju.

Najpopularniji alati kao što su Google i Yahoo Analytics omogućili su korisnicima, koji mogu biti mala i srednja poduzeća, uvid u svoje korisnike i mogućnost mjerenja uspješnosti Web mjesta. Ovisno o problemu, rudarenje Weba rade stručnjaci uz posebne alate i često uz analizu sirovih podataka.

Ključne riječi: rudarenje podataka, rudarenje Weba, Web analitika.

SUMMARY

Mining visit Web sites is a relatively new field that is increasingly used in business processes and in decision making. Data mining is used in most branches: security, traffic, commerce, medicine, and so on. While the methods and tools for Web mining find use for finding user patterns with Web sites or mobile applications. Tools and methods derived from data mining are especially adapted to the problems of Web mining. Results mining are used mostly for business purposes and are applied in optimizing Web sites, for recommendations and personalized content.

The most popular tools such as Google and Yahoo Analytics enabled users, who can be small and medium-sized enterprises, access to customers and the ability to measure the performance of Web sites. Depending on the problem, Web mining experts work with special tools and often with the analysis of raw data.

Keywords: data mining, Web mining, Web analytics.