Sustavi poslovne inteligencije

Matošić, Matej

Undergraduate thesis / Završni rad

2015

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:137:794752

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

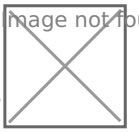
Download date / Datum preuzimanja: **2024-03-29**



Repository / Repozitorij:

Digital Repository Juraj Dobrila University of Pula





Sveučilište Jurja Dobrile u Puli
Fakultet za ekonomiju i turizam
"Dr. Mijo Mirković"

Matej Matošić

"Sustavi poslovne inteligencije"

Završni rad

Pula, 2015.

Sveučilište Jurja Dobrile u Puli
Fakultet za ekonomiju i turizam
"Dr. Mijo Mirković"

Matej Matošić
Matični broj : 0303038662, redovni student
Smjer: Informatika, III stupanj

"Sustavi poslovne inteligencije"

Završni rad

Predmet : Elektroničko poslovanje
Mentor: prof.dr.sc. Vanja Bevanda

Pula, rujan 2015.

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani _____, kandidat za prvostupnika _____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, . . 2015. _____

SADRŽAJ

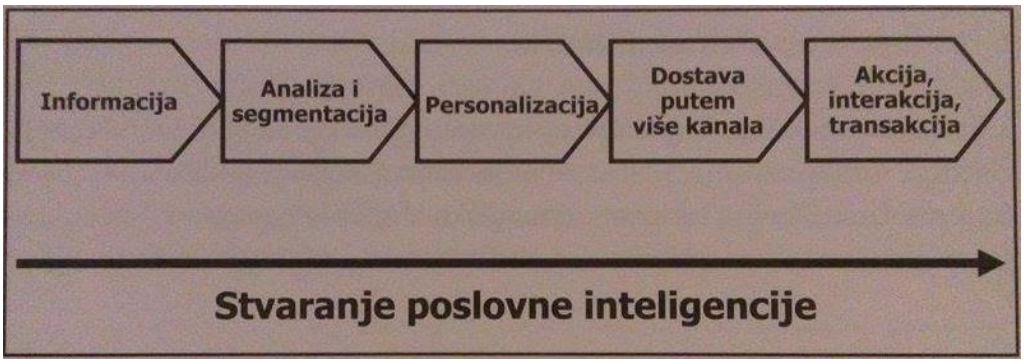
UVOD
1. POSLOVNA INTELIGENCIJA 1
   1.1. Pojam i značajke: 2
   1.2. Arhitektura sustava BI 3
2. SKLADIŠTENJE PODATAKA 5
   2.1. Metode 5
   2.2. Odrednice 6
   2.3. ETL procesi 8
   2.4. Arhitekture skladišta podataka 10
   2.5. Elementi i operacije višedimenzijskog modela podataka 11
3. OLAP- VIŠEDIMENZIJSKI INFORMACIJSKI SUSTAVI 13
4. RUDARENJE PODATAKA 17
   4.1. Proces rudarenja podataka 18
   4.2. Metode rudarenja podataka 19
   4.3. Primjer rudarenja podataka 23
5. IBM Cognos BI 27
   5.1. Usporedba Cognosa i drugog dobavljača – Pentaha 33
6. ZAKLJUČAK 36
7. POPIS LITERATURE 37
POPIS TABLICA 39
POPIS SLIKA 39
SAŽETAK 40

UVOD

U današnje vrijeme u poslovnim poduzećima trebalo bi znati planirati. Zbog velike konkurencije na tržištu, poduzećima je potrebno moći planirati kako bi opstali na tržištu. Kako bi mogla ostvariti svoje ciljeve, poduzeća trebaju razviti dobru strategiju. Znati planirati značilo bi znati i upravljati, a da bi se moglo upravljati poduzećem trebaju se donositi ključne odluke. Poslovna okolina se danas uvelike razlikuje od one prije. Promijenili su se brojni zakoni, došlo je do napretka tehnologije i sve je postalo automatizirano, poboljšale su se tehnike proizvodnje te se povećala kvaliteta proizvoda. Kupci očekuju dostupnost usluga u bilo koje vrijeme i na bilo kojem mjestu. Zbog ovih promjena u poslovnoj okolini, poduzeća su prisiljena smanjiti svoje vrijeme planiranja te donositi odgovarajuće odluke. Donošenje odluke se temelji na raspoloživim informacijama. Da bi se donijela odluka trebalo bi posjedovati točne i nužne informacije. Život u informatičkom dobu omogućio je posjedovanje određenih sustava i alata koji olakšavaju prikupljanje, analiziranje i distribuiranje informacija. U završnom radu proučit će se kako sustavi i alati poslovne inteligencije mogu utjecati na poboljšanje rada poduzeća pri poslovnom odlučivanju. Svako poslovno poduzeće teži za poboljšanjem poslovanja kako bi ostvarilo što veće prihode, a istovremeno smanjilo troškove. Opisati će se sustavi poslovne inteligencije i objasniti princip njihova rada te će se isto tako pobliže opisati i jedan od alata poslovne inteligencije – Cognos. Hipoteza rada je da su sustavi i alati poslovne inteligencije u današnje vrijeme postali ključni faktori uspjeha nekog poduzeća te da im olakšavaju proces donošenja bitnih poslovnih odluka. Odgovorit će se na pitanja može li se i kako jednostavno doći do potrebnih informacija te kako u tome pomažu sustavi poslovne inteligencije.

1. POSLOVNA INTELIGENCIJA

Jedan od važnijih današnjih izvora koji odlučuje o propasti ili opstanku poduzeća je informacija. U prošlom stoljeću nisu postojala računalna sredstva za analiziranje podataka nego su se informacije stjecale iz izvora koji nisu bili automatizirani te su se odluke donosile intuicijski. Uvođenje automatizma je dovelo do povećanja dostupne količine podataka. No, i dalje, zbog nekompatibilnosti i infrastrukture, nije bilo jednostavno prikupiti informacije. Trebali su proći mjeseci kako bi se dobila konačna analiza i izvještaji. Donošenje dugoročnih odluka uvelike su olakšali izvještaji, dok su se kratkoročne i dalje donosile intuicijski. Strateška prednost nekog poduzeća nije se više temeljila na fizičkim resursima nego se počela stjecati količinom znanja i informacija u poduzeću. Prikupljene informacije su se pretvarale u znanje poduzeća. Informacije su postale ključni faktor pri donošenju nekih od najjednostavnijih odluka kao što je povećati ili smanjiti broj nekog proizvoda o čemu su odlučivale informacije o prodanoj količini tijekom tjedna ili stanja u skladištu. Menadžment može odlučivati o vraćanju nekog propalog proizvoda ili uvođenja novog. Za donošenje takvih malo kompliciranijih odluka nije dovoljno samo prikupiti informacije nego bi ih trebalo profiltrirati i povezati kako bi se donijela odluka. To nije jednostavan proces zbog toga što se na tržištu nalazi velik broj informacija, a trebalo bi izabrati one korisne. Tu se pojavljuje tzv. poslovna inteligencija koja obavlja taj proces (prikuplja i analizira podatke).



Slika 1. Komponente modela poslovne inteligencije

(izvor: PANIAN, Ž. I KLEPAC, G. (2003.))

Na slici je prikazano stvaranje poslovne inteligencije. Kako informacija prolazi kroz niz filtera kako bi bila što korisnija.

1.1. Pojam i značajke:

Business intelligence (BI) ili poslovna inteligencija je pojam koji obuhvaća skup alata, aplikacija i metoda kojima se omogućava proces prikupljanja raspoloživih podataka te pretvorba tih podataka u informacije potrebne korisnicima za donošenje poslovnih odluka. Poslovna inteligencija se smatra posebnom disciplinom u koju ulaze računovodstvo, poslovna analiza, informacijske tehnologije i marketing. Pomoću poslovne inteligencije podaci se prikupljaju, analiziraju i distribuiraju kako bi olakšali korisnicima donošenje poslovnih odluka ili rješavanje poslovnih problema. Također omogućuje svim korisnicima pregled poduzeća kako bi dobili informacije koje im trebaju i omogućava poduzećima da se pripreme za situacije koje ih očekuju na tržištu.

Business intelligence ima tri značajke[1]:

- Poslovna inteligencija je proces kojim se podaci i informacije prikupljaju i obrađuju te pretvaraju u znanje poduzeća.
- Usmjerava se na informacije koje mogu predvidjeti buduće događaje ili kretanja.
- Ona ima glavnu ulogu u procesu donošenja odluka.

Prema istom referiranju korisnici od poslovne inteligencije očekuju:

1. Sigurnost- zaštitu podataka od neovlaštenih pristupa.
2. Raspoloživost i dostupnost podataka bilo gdje i bilo kad.
3. Brzu obradu podataka i vidljivost rezultata obrade u poslovnom okruženju.
4. Stabilna memorija – radi podržavanja rasta broja korisnika i podataka
5. Brzi razvoj i jednostavna administracija - alati koji će olakšati poslove vezane uz administriranje BI sustava

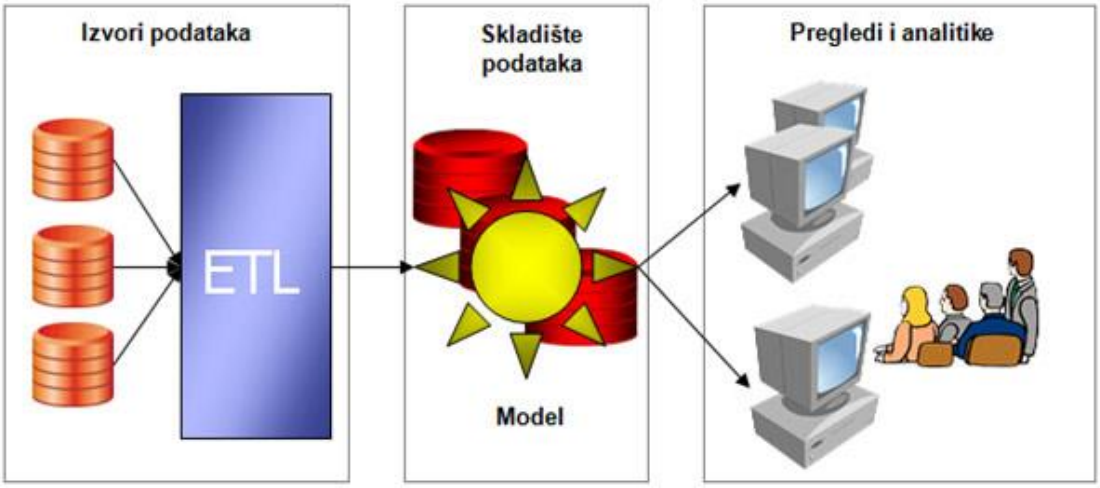
Za opstanak poduzeća na tržištu bitno je raspolagati pravom informacijom u pravo vrijeme. Sustavi poslovne inteligencije skupljaju te informacije pa ih ekstrahiraju, transformiraju i pohranjuju te zahvaljujući definiranim pravilima i povezivanjem nastaju skladišta podataka. Pomoću poslovne inteligencije poduzeće može analizirati ili odrediti kupce i dobavljače, vidjeti koji kupci postižu bolje poslovne rezultate i predvidjeti buduće trendove. Sve više poduzeća uvodi sustavi poslovne inteligencije kako bi konkurirali drugima na tržištu, osim toga koriste ju zbog distribucijskih kanala te ponude robe i usluga koje nadmašuju potražnju. Koristiti se za

[1] JAVOROVIČ, B. i BILANDŽIĆ, M. (2007.)

pronalazak novih kupaca ili zadržavanje postojećih. Ako kupac bude nezadovoljan proizvodima ili uslugama, osim što će izbjegavati određeno poduzeće, može stvoriti i lošu reputaciju. Najveći problem je kako prikupiti podatke relevantne za donošenje odluka kojima bi se povećao prihod, a smanjili troškovi. U poduzećima koja primjenjuju alate poslovne inteligencije smanjilo se vrijeme koje im je potrebno za donošenje poslovnih odluka (s tjedne na dnevnu bazu), štoviše povećalo se vrijeme potrebno za analizu podataka. Poduzeća očekuju bolji marketing, nove proizvode, usmjerenje koncentracije na kupce, povećane prihode, a smanjene troškove.

1.2. Arhitektura sustava BI

Informatički gledano poslovna inteligencija se sastoji od procesa prikupljanja, spremanja i analiziranja podataka. Podaci koji su potrebni za donošenje odluka su pohranjeni na različitim mjestima i sustavima, tablicama u Excelu, internetu i drugdje. Jedna od važnijih faza izgradnje sustava poslovne inteligencije je definiranje mjesta gdje su podaci te izvlačenje istih.
Izvlačenje, transformiranje, čišćenje i spremanje podataka se vrši ETL (engl. ETL- Extract-Transform-Load) procesima. Podaci se pohranjuju u organizirana skladišta podataka odakle se mogu analizirati. Izrađuju se sustavi u kojima korisnici sami mogu pristupiti potrebnim podacima te se tako olakšava rad korisnicima. Sastavnice sustava poslovne inteligencije su infrastruktura, funkcionalnost, organizacija i poslovanje. U Infrastrukturi se nalaze skladišta podataka, alati potrebni za ETL procese te operativna skladišta podataka. Pod funkcionalnost pripadaju BI platforme, Data Mining, BI aplikacije (operativne, strateške, analitičke). Organizacija se sastoji od mjerenja učinkovitosti, korporativne kulture, BI metodologije, centra poslovne inteligencije (sjedinjuje znanja i vještine). Poslovanje daje podatke o uspješnosti, novim trendovima i transparentnostima. Današnje organizacije se ne koriste gotovim rješenjima nego zahtijevaju specifična rješenja i alate koji se prilagođavaju potrebama organizacije, ali i korisnika.



Slika 2. Arhitektura sustava BI

(izvor: http://alfatec.hr/wp-content/uploads/2011/11/DTW1.jpg , 7.8.2015.)

ETL (Izdvajanje (engl. Extract), Transformacija (engl.Transform), Punjenje (engl.Load) to je niz procesa kroz koje podaci prolaze iz izvorišnih sustava te se spremaju se u skladište podataka. Kroz više metoda korištenja podataka iz skladišta korisnici dolaze do svojih podataka.

Zadnjih par godina u Hrvatskoj je došlo do proširenja i napretka tržišta koje se obogatilo izborom hardwera, softwera i raznih usluga, što omogućava potencijalnim kupcima širok raspon izbora. Kupci u Hrvatskoj se ne mogu žaliti na količinu izbora zbog toga što se na natječaje javlja desetak i više različitih dobavljača. Tržište je specifično po tome što je većina kupaca koncentrirana na OLAP (engl. OLAP- Online Analytical processing) tehnologije i izvještavanje te na analizu podataka prikupljenih u CRM (engl. CRM-Customer relationship management) sustavima, dok je veoma slaba potražnja za programe rudarenja podataka. Dobavljači su podijeljeni u dvije grupe: 1) veliki dobavljači IBM, Oracle i Microsoft čije baze služe kao temelj za izgradnju skladišta podataka. Platforma se koristi kao osnovica i klijentima se nudi front-end alati za krajnjeg korisnika. 2) veliki izvorni BI dobavljači koji su prisutni na hrvatskom tržištu su SAS, Cognos, Business Object i Micro Strategy. Bazirani su na front-end alate za krajnje korisnike koji se mogu vezati na bilo koju platformu za bazu podataka, osim SAS-a koji bi trebao biti povezan s lokalnim uredom.

2. SKLADIŠTENJE PODATAKA

Skladište podataka (engl. data warehouse) je skup podataka organizacije na kojem se temelji sustav potpore odlučivanju. Skladište podataka bi trebalo podacima potpuno "pokriti" jedno ili više poslovnih područja (npr. nabave, prodaje), podaci u skladištu trebaju biti sveobuhvatni, tj. "integrirani" od unutarnjih podataka organizacije, ali i podataka iz njenog okruženja. Podaci bi trebali obuhvatiti duži vremenski period (pet, deset ili više godina), jer su vremenske analize poslovno vrlo značajne. Orijentiranost prema poslovnim analizama ne zahtijeva od skladišta da se podaci promptno ažuriraju kao u bazi podataka[2]. Zaključak je da skladište predstavlja podršku menadžmentu pri donošenju poslovnih odluka. Skupljanje više podataka u jednu bazu i nije tako jednostavan proces. Skladište podataka je mjesto gdje su svi podaci na jednom mjestu, sigurni, definirani i prilagođeni za razne analize koje će se obavljati. Skladište podataka predstavlja jako dobar izvor podataka u organizaciji te su u njemu dostupni svi stari i novi podaci. Skladišta su stvorena kako bi se poboljšao proces donošenja odluka i kako bi se korisnicima omogućilo lakše pristupanje i pregledanje podataka iz više različitih kutova.

2.1. Metode

Postoje tri metode korištenja podataka iz skladišta[3]: sumarni izvještaj, analitička obrada (OLAP), dubinska analiza podataka ili rudarenje podataka. Sumarni izvještaj (engl. reporting) predstavlja najjednostavniju metodu korištenja skladišta podataka te zahtijeva minimalno znanje o strukturi skladišta. Izrada sumarnih izvještaja vrlo je slična radu s transakcijskom bazom podataka, ali je razlika u tome što skladište podataka sadrži mnogo veću količinu podataka i pruža kvalitetniji uvid u podatke. Analitička obrada (OLAP) je interaktivna dinamička analiza velike količine podataka. Upiti se postavljaju na temelju podataka dobivenih iz prethodnih upita tj. njihova struktura nije unaprijed poznata. Alati za analitičku obradu u potpunosti su prilagođeni višedimenzijskom modelu podataka. Zbog toga korisnik bi trebao poznavati strukturu skladišta podataka te imati dovoljno znanja za rad s alatima za analitičku obradu. Analitičkom obradom podataka bave se poslovni analitičari i menadžeri organizacija.

---

[2] *http://www.pravinshetty.com/StudyMaterials/Concepts/Wiley%20&%20Sons%20-%20Building%20The%20Data%20Warehouse.%20Third%20Edition.pdf (7.8.2015)*

[3] GARAČA, Ž. i ĆUKUŠIĆ, M. (2011.)

Dubinska analiza podataka ili rudarenje podataka (engl. Data Minig) je proces kojim se otprije nepoznati obrasci i odnosi među podacima u bazi podataka automatski otkrivaju. Alati za rudarenje podataka koriste vrlo složene, sofisticirane algoritme kako bi unutar velikog niza podataka identificirali odnose i veze za koje se nije znalo da postoje. Osoba koja provodi dubinsku analizu trebala bi imati vrlo široko znanje o skladištima podataka, izuzetno dobro poznavati strukturu skladišta čiji sadržaj analizira i razumjeti sve algoritme koje ima na raspolaganju kako bi se mogli primijeniti na pravi način i u pravilnom redoslijedu. Za dubinsku analizu podataka potrebno je stručno obrazovanje i iskustvo. Poslovna analitika treba biti jednostavna za korisnika. Korisnik bi trebao moći komunicirati sa sustavom, postavljati razne izvještaje i dobivati rezultate kroz web aplikacije bez pomoći neke treće osobe. Iako se proces prikupljanja podataka čini jednostavan, to i nije baš tako. Trebalo bi uključiti eksterne podatke koji se integriraju i otklanjaju im se nedostaci te se na kraju pohranjuju kako bi bili dostupni svima.

2.2. Odrednice

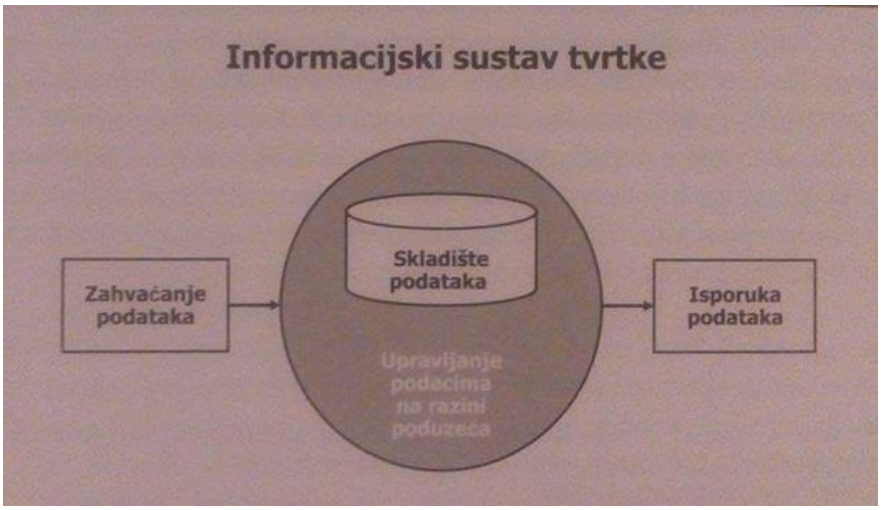
Poznati autor Wililiam Inmon dijeli skladište na četiri ključne odrednice: subjektna orijentiranost, integriranost, postojanost i vremenska dinamičnost.[4]

- Subjektna (tematska) orijentiranost - podaci u skladištu podataka su organizirani oko područja koji su bitni za njeno poslovanje (klijenti, proizvodi i sl.), a transakcijski sustavi su organizirani prema aplikacijama koje organizacija koristi.
- Integriranost - najvažnije u skladištu podataka je da su podaci integrirani. To znači da imaju definirane nazive, jedinice, strukture, ključeve i dr. Kada se izrađuje skladište radi se na cijeloj organizaciji, ali svaki odjel ima svoju bazu podataka. Podaci se prenose u skladište iz više izvora i iz različitih baza te se tako podaci trebaju integrirati s drugim podacima. Trebalo bi definirati načine definiranja podataka, ključeva, atributa.
- Postojanost ili nepromjenjivost skladišta - u transakcijskim bazama stalno se rade izmjene i brisanje podataka, kada se unesu u skladište oni se ne mijenjaju ako je unos ispravan. Nemogućnost brisanja i mijenjanja stvara sigurnost tj. sprječava narušavanje sigurnosti podataka.

[4] GARAČA, Ž. i ĆUKUŠIĆ, M. (2011.)

- Vremenska dinamičnost skladišta – označuje to da je svaki podatak bio aktualan u nekom vremenskom periodu. Baze podataka čuvaju podatke najviše nekoliko mjeseci, dok skladište podataka sadrži podatke stare od 5 do 10 godina.

Iz navedenih odrednica skladišta vidljivo je da koncept skladišta podataka udovoljava brojnim zahtjevima što se postavljaju pred skladište podataka. U skladištu podataka je sadržana velika količina detaljnih podataka zbog toga što tvrtka ima pregled svega što se događalo tijekom svih marketinških kampanja kako bi mjerila njihovu učinkovitost. Marketing je nepredvidiv proces u kojem se promjene brzo događaju, a raspolaganje detaljnim podacima omogućuje brzu reakciju tvrtke na te promjene. Skladište podataka često ažurira podatke izvršenih poslovnih transakcija i to je najbolje obavljati odmah nakon završetka neke transakcije ili procesa. Osim toga, skladište služi velikom broju ljudi te je oblikovano tako da može poslužiti svakoj (čak i nepredvidivoj) svrsi. Uvijek je raspoloživo svima te proširivo i zaštićeno (omogućuje trajnost i zaštitu osjetljivih podataka).



Slika 3. Pozicija skladišta podataka u informacijskom sustavu poduzeća
(Izvor: PANIAN, Ž. I KLEPAC, G. (2003.))

Na slici je prikazano kako je skladište podataka u sredini informacijskog sustava tvrtke. Skladište podataka upravlja podacima tako što ih dohvaća, obrađuje i isporučuje.

Funkcije skladišta podataka[5] se očituju u tome da je skladište odraz poslovnih pravila primjenjivih u poduzeću. Trebalo bi biti oblikovano tako da se jednostavno i brzo prilagodi svim

[5] PANIAN, Ž. I KLEPAC, G. (2003.)

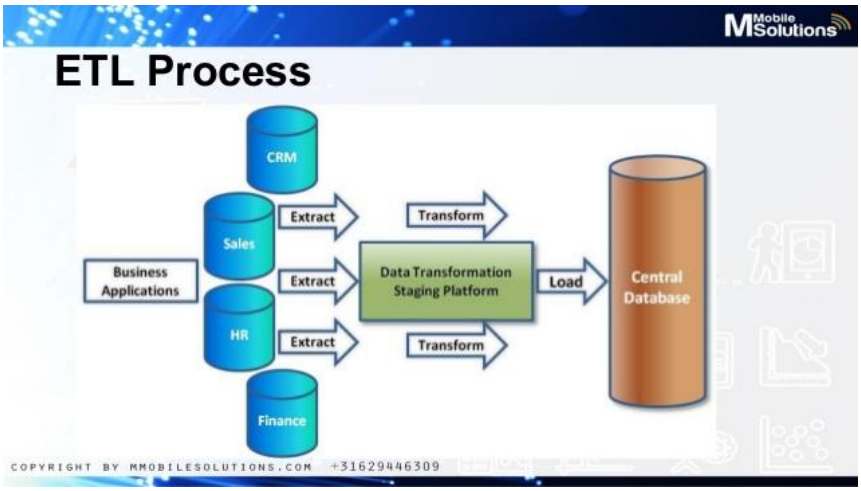
promjenama u poslovnim pravilima što obuhvaća sposobnost prihvaćanja novih podataka, promjene u hijerarhijskim i logičkim odnosima među podacima. Predstavlja točku u kojoj se prikupljaju podaci koji služe za stvaranje integriranih informacija. Pri modeliranju skladišta trebaju se primjenjivati tehnike koje podržavaju predmetnu orijentaciju i omogućuju prilagodljivost. U njemu se nalaže strateške informacije čija se povijest može povezati s podacima i odnosima među njima što iziskuje primjenu tehnike modeliranja koja omogućuje ugradnju povijesne perspektive u spremljeni sadržaj. Skladište podataka je izvor informacija koje se izvode iz spremišta podataka (engl. Data Mart). Spremišta se koriste za istraživanje, rudarenje, postavljanje upita i analitičku obradu. Ona iziskuju potrebu za modelom koji može agregirati i sintetizirati podatke. Predstavlja izvor stabilnih podataka koji su neovisni o promjenama u procesima te treba model neosjetljiv na utjecaj procesa.

2.3. ETL procesi

ETL procesi su skup procesa koji ekstrahiraju tj. zahvaćaju i vade (engl. Extract), transformiraju tj. preoblikuju (engl. Transform) i pune (engl. Load) podatke iz jednog ili više transakcijskih sustava u skladište podataka.[6] ETL procesi započinju pripremama koje uključuju reformatiranje, usklađivanje i čišćenje izvornih podataka. Reformatiranje podataka služi kako bi se podaci iz različitih datoteka, koji podržavaju različite formate, prikazali u jedinstvenom formatu koji se koristi u daljnjoj obradi. Vrši se usklađivanje nepotrebnih podataka zbog toga što postoje isti podaci na više mjesta u informacijskim sustavima koji su nedosljedni što znači da im se razlikuju vrijednosti na mjestima gdje se javljaju. Na kraju bi trebalo očistiti podatke od pogrešaka koje nastaju kao posljedica rada informacijskog sustava (računalni virusi, sabotaže, itd.). Alati za ekstrakciju podataka biraju se gledajući iz perspektive izvršnih sustava u kojima je najprimjereniji način ekstrakcije podataka kopiranje cijelog sadržaja operativnih datoteka i transakcijskih baza podataka (kopije se proslijede stručnjacima za obradu podataka, ali često su velikih dimenzija) ili gledajući iz perspektive poslovne inteligencije u kojoj je najprimjereniji način ekstrahiranja podataka sortiranje, filtriranje, čišćenje i agregiranje podataka u jednom koraku (utječe na transakcijske sustave i dovodi do poremećaja pri obavljanju redovitih poslova). Rješenje ovih problema nalazi se u kompromisu između dviju perspektiva: izabrati alate koji omogućuju što učinkovitije obavljanje ETL procesa uz što brže zahvaćanje potrebnih podataka.

[6] Ibid, str. 86.

Najviše vremena ETL procesa oduzima transformacija (80%). Problemi koji usporavaju transformaciju i tako ju odužuju su: nekonzistenti primarni ključevi (nepodudarnost primarnih ključeva u izvornim datotekama i oni koje pretpostavljaju alati i aplikacije BI), nekonzistentne vrijednosti podataka (kopiranje i dupliciranje podataka), različiti formati podataka (npr. broj računa nije onakvog formata koji pretpostavlja alat poslovne inteligencije), netočne vrijednosti podataka (logika čišćenja podataka koja ispravlja netočne vrijednosti), ugrađena procesna logika (korištenje zastarjelih programa koji dobro rade, ali korisnici ne poznaju sistem rada), sinonimi i homonimi (onemogućavaju jednostavno prepoznavanje redundantnih podataka).[7] Bez transformacije podataka nema ni unosa podataka te je ovaj proces potrebno pažljivo i savjesno obaviti. Većina podataka koji prolaze kroz ETL procese su aktivni, no postoje i oni arhivirani povijesni podaci potrebni za analize te zbog toga treba upotrijebiti tri vrste ETL programa: programi za inicijalno punjenje (početno punjenje baza podataka potrebnih za stvaranje poslovne inteligencije aktivnim podacima), programi za punjenje povijesnih podataka (početno punjenje baza podataka arhiviranim povijesnim podacima) i programi za inkrementalno punjenje (slijedno punjenje baza podataka tekućim operativnim podacima).



Slika 4: Koncept ETL procesa

(Izvor: http://www.slideshare.net/mondyholten/etl-overview-training, 8.8.2015.)

Prikaz ETL procesa u kojem se podaci prvo ekstrahiraju iz izvora, zatim preoblikuju tj. usklađuju se formati podataka te se vrši unos podataka u centralnu bazu.

[7] Ibid, str. 87.

Kako bi se ispravno odlučilo koji tip ETL alata treba koristiti, tvrtka bi trebala izvršiti vrednovanje raspoloživih opcija koje se provodi sistematičnim i temeljitim poduzimanjem sljedećih koraka: provedbom analize troškova i koristi, stvaranjem popisa ETL proizvoda i ponuđača, usporedbom ETL proizvoda i ponuđača s unaprijed postavljenim zahtjevima, objektivno ocjenjivanje svakog pojedinačnog ETL proizvoda, provjera referentne liste klijenata svakog ponuđača ETL softvera, sužavanje popisa ETL proizvoda i ponuđača, organiziranje demonstracije proizvoda, testiranje odabranih proizvoda.[8]

2.4. Arhitekture skladišta podataka

Razvijene su i koriste se tri arhitekture skladišta podataka[9]:

- Dvoslojna arhitektura s jednim zajedničkim skladištem podataka
- Dvoslojna arhitektura s većim brojem nezavisnih lokalnih spremišta podataka
- Troslojna arhitektura s jednim zajedničkim skladištem podataka i većim brojem povezanih lokalnih spremišta podataka

Dvoslojna arhitektura s jednim zajedničkim skladištem podataka je tradicionalan način korištenja skladišta pri kojem jedinstveno, centralno skladište podataka tvrtke zahvaća podatke iz mnogobrojnih izvora (ranije razvijeni, naslijeđeni sustavi koji održavaju baze podataka starijih arhitektura; operativni podaci kao oni iz online sustava za obradu transakcija; vanjski izvori podataka). Zbog toga što su ovakva skladišta centralizirana te velika i kompleksna, teško je kvantificirati povratak ulaganja u ovakvu implementaciju te je to najčešći razlog zamjene ovog tipa skladišta nekim novijim. Izgradnja većeg broja nezavisnih spremišta podataka je postala alternativa jedinstvenom, centraliziranom skladištu. U praksi je dokazano da taj povećani komfor rezultira dugoročno povećanim troškovima. Neke od poteškoća koje su svojstvene ovom skladištu su: povećano radno opterećenje, ograničena proširivost platforme, otežana implementacija novih aplikacija, poteškoće u komunikaciji među organizacijskim jedinicama tvrtke. Ovi nedostaci rezultiraju razvojem novog rješenja problema skladištenja podataka.

Nakon nedostataka prošlih tipova skladištenja podataka, došlo je razvitka najučinkovitijeg rješenja u kojem veći broj lokalnih spremišta podataka povezuje jedno zajedničko skladište podataka. U skladištu podataka se pohranjuju sve bitne informacije iz djelatnih sustava tvrtke da

[8] PANIAN, Ž. I KLEPAC, G. (2003.).
[9] Ibid, str. 94.

bi se onda isporučivale lokalnim spremištima u jedinstvenom formatu. Lokalna spremišta zatim agregiraju i sažimaju podatke prema kriterijima koje postavljaju lokalne aplikacije. Prednosti ovog tipa skladišta su smanjeno radno opterećenje informatičara, otvaranje mogućnosti dodavanja novih aplikacija, povećavanje proširivosti platforme, točnost i preciznost informacija, otklanjanje poteškoća u komunikaciji, ali i dalje ostaje problem visokih troškova uspostavljanja zajedničkog skladišta podataka na razini cijele tvrtke. Taj se problem donekle umanjuje primjenom odgovarajuće metodologije uspostavljanja troslojne arhitekture – inkrementalne izgradnje sustava skladištenja podataka. Ova metoda osigurava dizajn lokalnih spremišta koji će omogućiti njihovo uklapanje u model podataka sposoban prevladati ograničenost spremišta na pojedinačne aplikacije i organizacijske jedinice tvrtke. Umjesto da se cijelo okruženje skladišta podataka gradi odjednom, ono se gradi postupno te nadograđuje. Nakon što se sva planirana spremišta dovedu u funkciju, nastaje potpun sustav skladištenja podataka na razini cijele tvrtke.

2.5. Elementi i operacije višedimenzijskog modela podataka

Dimenzijsko modeliranje definirara se kao oblikovanje i implementacija standardnog koncepta pomoću kojeg se omogućuje brže pretraživanje podataka. Konceptualni model definira poslovne procese kroz neke vrste dimenzija koje se organiziraju hijerarhijski. Prema višedimenzijskom modelu definirane su: činjenice, mjere, dimenzije i hijerarhije.[10]

Činjenica (engl. fact) predstavlja polaznu točku poslovne analize. Ona je središnja točka interesa u procesu donošenja poslovne odluke. Mjere (engl. measures) su opisne varijable s neprekidnim skupom vrijednosti koje opisuju činjenicu. Kako skladište podataka sadrži golemi broj zapisa, zbrajanjem ovih numeričkih pokazatelja mogu se dobiti vrijednosti zanimljive za analizu. Glavna područja interesa poslovne analize razlikuju se prema tipu organizacije i načinu poslovanja. Dimenzije (engl. dimensions) su nezavisni parametri koji opisuju činjenicu, a sastoje se od atributa diskretnog skupa vrijednosti. Uobičajeni primjeri dimenzija za činjenicu prodaje proizvoda u skladištu podataka maloprodajnog lanca trgovina su vrijeme, prodavaonica i prodani proizvod. Dimenzijska hijerarhija (engl. dimensional hierarchy) je pojam koji se odnosi na funkcijski ovisne dimenzijske atribute jedne dimenzije, kojima je činjenica opisana na različitoj razini granularnosti odnosno detaljnosti. Dimenzijska struktura je najčešće korištena struktura unutar skladišta podataka te je njezina velika prednost to što se podaci mogu virtualizirati.

[10] GARAČA, Ž. i ĆUKUŠIĆ, M. (2011.)

Najčešće se koriste tri dimenzije, te se podaci prikazuju u podatkovnim kockama u kojima dimenzija kocke predstavlja parametar. Dimenzije imaju svoje pozicije čije vrijednosti predstavljaju atribute. Rotacija ili pivotiranje predstavlja važnost neke dimenzije tako što će se isticati ili će se staviti u pozadinu. Selekcija je odabiranje jedne ili više dimenzija gdje je moguće odabrati pozicije različitih dimenzija. Problem je fiksiranje jedne dimenzije dok se ostale mijenjaju. Granularnost ili zrnatost se odnosi na detaljnost podataka u skladištu. Najnižu zrnatost imaju najdetaljniji podaci, a najvišu zrnatost imaju podaci koji su najmanje detaljizirani.

3. OLAP- VIŠEDIMENZIJSKI INFORMACIJSKI SUSTAVI

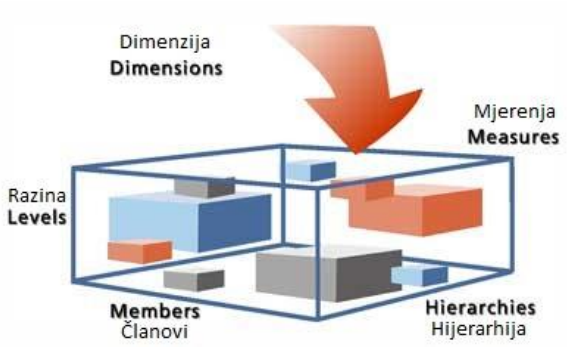
OLAP (engl. On Line Analytical Processing) je softverska tehnologija koja omogućava analitičarima, menadžerima i radnicima znanja uvid u podatke pohranjene u skladištu podataka kroz brz, konzistentat, interaktivan pristup i širok spektar mogućih pogleda na informacije transformirane iz sirovih podataka, a koje reflektiraju stvarnu dimenzionalnost organizacije razumljivu korisnicima. Olap je informacijski sustav za pristup i manipulaciju višedimenzijskim podacima koji dolaze s različitih izvora, a spremljeni su u skladište podataka. Također se definira kao brza analiza dijeljenih višedimenzijskih informacija (engl. Fast Analysis of Shared Multidimensional Information). Navedena definicija sadrži pet ključnih pojmova koji karakteriziraju OLAP koncept:[11]

- Brza – označava da je informacijski sustav dizajniran kako bi korisnicima dostavio tražene rezultate u što kraćem vremenskom roku, u prosjeku manje od 5 sekundi, pri čemu jednostavnije analize ne zahtijevaju više od 1 sekunde, a one složenije više od 20 sekundi.
- Analiza - označava da informacijski sustav može podržati poslovnu logiku te matematičku i statističku analizu koja je potrebna korisniku. Svakako je važno korisniku omogućiti i definiranje novih pravila tj. samostalno definiranje odgovarajućih izvještaja.
- Dijeljenih – označava da informacijski sustav može istovremeno koristiti više korisnika te da osigurava sve sigurnosne zahtjeve od tajnosti do upravljanja konkurentnih pristupom na bilo kojoj razini podataka.
- Višedimenzijskih – informacijski sustav ima potrebu pružati višedimenzijski konceptualni pogled na podatke i hijerarhiju podataka jer to predstavlja najlogičniji način za provođenje poslovne analize. Kada bi OLAP trebalo definirati jednom riječju to bi bilo upravo višedimenzionalnost.
- Informacija – kvalitetna informacija danas predstavlja najvažniji resurs za donošenje kvalitetnih odluka. OLAP sustavi korisnicima trebaju pružiti pravovremene, točne i razumljive informacije.

[11] GARAČA, Ž. i ĆUKUŠIĆ, M. (2011.)

OLAP se zasniva na sustavu višedimenzijske analize, pa se tako podaci mogu promatrati kroz više dimenzija. Bitna karakteristika OLAP-a je velika brzina rada, postavljanje upita i dobivanje odgovora u što kraćem roku. Druga bitna karakteristika je analiziranje velikog broja dimenzija, u kvalitetnim alatima može biti i više od 10 dimenzija. Također OLAP je vrlo interaktivan. U klasičnom izvještaju korisnik čita i analizira, dok OLAP sustav pretpostavlja interakciju korisnika i sustava. Prednost OLAP-a proizlazi iz mogućnosti strukturiranja podataka na način koji je u skladu s načinom na koji ljudi prirodno analiziraju pojave. Najvažnije od struktura su dimenzije i hijerarhija podataka. Dimenzija je pripadnost entiteta, na kojega se odnose podaci, određenoj kategoriji.[12] Svi članovi neke dimenzije (npr. proizvodi ili kupci) prema nekim obilježjima se svrstavaju u skupine. Pripadnost skupini se određuje na dva načina: podatke o članovima bi trebalo moći uspoređivati te agregirati. Skupovi podataka koji su strukturirani po dimenzijama mogu se podvrgavati analizi po tim dimenzijama ili dimenzijskoj analizi. Osnovna metoda dimenzijske analize podataka je metoda raslojavanja i presijecanja (*slice and dice*). Višedimenzijski skupovi podataka mogu se analizirati horizontalno i vertikalno. Osnovna kvaliteta metode *slice and dice* je ta što je korisnicima lako shvatljiva. Suvremeni OLAP alati ovu metodu jako brzo obavljaju i korisnicima nude privlačna i razumljiva grafička sučelja (kocke ili zvjezdasti dijagrami). Osim toga, korisnike privlači mogućnost rotacije ili pivotiranja dimenzija. OLAP kocka se može po volji rotirati oko osi. OLAP sustavi omogućuju i organiziranje podataka u hijerarhijske strukture kako bi se podaci mogli agregirati u nove podatke višeg stupnja apstrakcije. Hijerarhijska organizacija svake dimenzije omogućuje korisnicima da krenu s vrha hijerarhije i spuštaju se do nižih razina podataka kako bi pronašli zanimljive uzorke ili anomalije (taj postupak se naziva *drill down*).

[12] PANIAN, Ž. I KLEPAC, G. (2003.)



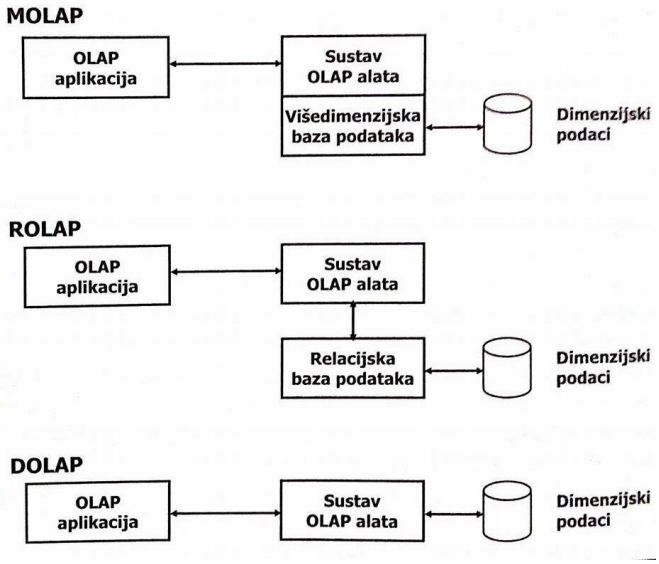
Slika 5: Struktura OLAP sustava

(Izvor: http://www.esri.com/news/arcuser/0206/olap1of2.html, 10.8.2015.)

Prikaz odnosa među komponentama strukture OLAP-a koja se sastoji od dimenzija i hijerarhije podataka, razina dimenzije, članova određene razine dimenzije i provedenih mjerenja.

Dva temeljna oblika OLAP-a s obzirom na to kako su podaci spremljeni: višedimenzijski (OLAP-MOLAP) i relacijski (OLAP- ROLAP). U MOLAP-u su podaci spremljeni u višedimenzijskim bazama podataka i dijeli se u tri sloja. Prvi sloj je podatkovni sloj, drugi se odnosi na aplikacijski, dok treći predstavlja prezentacijski sloj. MOLAP-ov engine koji se nalazi u aplikacijskom sloju daje višedimenzijski pogled na podatke. Sustavi za upravljanje bazama(engl. Database Managment System, DBMS) daju mogućnost čitanja podataka iz višedimenzijskih baza podataka. Specifično je to što nisu izrađeni na nekom općem modelu nego na relacijskom modelu pa je zbog toga MOLAP teško prilagodljiv. Relacijski OLAP ili ROLAP zadržava model informacijskog sustava gdje se analize provode u relacijskom bazama, a podaci se spremaju u obliku redaka i stupaca. Ovako su podaci spremljeni u poslovnim dimenzijama. Kako bi struktura ostala skrivena korisniku, a ipak da podaci budu višedimenzionalni, kreiran je semantički sloj. Sloj meta podataka postavlja dimenzije u relacijske tablice te sažima i agrerira. Sastavni su dio ROLAP modela i mogu se pohraniti u relacijske baze. ROLAP također ima troslojnu arhitekturu kao i MOLAP. Kada se bira između korištenja MOLAP-a ili ROLAP-a trebalo bi paziti na sljedeće: MOLAP alati imaju u sebi predizračunate gotove setove podataka koji predstavljaju podatkovne kocke iz kojih se izvlače mogući odgovori na postavljena pitanja. Vrlo brz je odaziv, tj. odgovor na pitanje, a nedostatak je to što postane velik uslijed dodavanja novih dimenzija i detaljnijih podataka. MOLAP se preporučava korisnicima koji imaju ograničene upite npr. iz područja financija jer se redovito ažuriraju podaci, a korisnik postavlja

pitanja na dnevnoj, mjesečnoj, godišnjoj bazi. ROLAP ne koristi predizračunate podatkovne kocke i pruža mogućnost postavljanja bilo kojih pitanja. Nedostaci su sporo vrijeme odgovora i moguća ograničenja skalabilnosti. Prilagođeni su korisnicima koji će svakodnevno postavljati različite upite.



Slika 6 : Inačice OLAP alata

(Izvor: PANIAN, Ž. I KLEPAC, G. (2003.))

Prikaz inačica i usporedba OLAP alata – višedimenzijskog MOLAP-a, relacijskog ROLAP-a i OLAP alata za stolna računala DOLAP

HOLAP alati (engl. Hybrid OLAP) su kombinacija MOLAP i ROLAP alata te predstavljaju hibridni alat uz pomoću kojega se simultano obavljaju višedimenzijske analize iz podataka koji su spremljeni u relacijskim bazama i višedimenzijskoj bazi.

OLAP alati za stolna računala (engl. Dekstop OLAP, DOLAP) su jednostavniji alati prilagođeni stolnim računalima i manje zahtjevnim korisnicima. Podržavaju jednostavnije upite koje će zadovoljiti pojedinog korisnika.

4. RUDARENJE PODATAKA

Rudarenje podataka (engl. *data mining*) se može definirati kao značajan proces otkrivanja uzoraka u velikim skupovima podataka koji uključuje metode koje se nalaze na raskrižju umjetne inteligencije, strojnog učenja, statistike i baze podataka. Sveukupni cilj procesa rudarenja podataka je izvući podatke iz skupa podataka i pretvoriti ga u razumljivu strukturu za daljnju upotrebu. Osim same analize, proces uključuje aspekte baze podataka, upravljanja podataka, pred i post-obradu podataka, vizualizaciju i online ažuriranja. Ipak, u sustavima poslovne inteligencije rudarenje podataka nije nužno podržano skladištima podataka, a jednostavno se može definirati kao pronalaženje zakonitosti u podacima. Pojam rudarenja podataka se može shvatiti na dva različita načina. U prvom, širem značenju, pojam se odnosi na cijeli proces pronalaženja znanja iz podataka na raspolaganju, dok se drugo, uže značenje, odnosi na definiranu fazu obrade podataka. Proces rudarenja podataka uključuje određivanje ciljeva prema potrebama korisnika, pripremu podataka, rudarenje podataka, ocjenu i upotrebu otkrivenog znanja. Definirana faza obrade podataka uključuje određivanje najbližeg susjeda, grupiranje, stablo odlučivanja, umjetne neuronske mreže i asocijativna pravila. Rudarenje podataka se oslanja na visoke domete informatike, matematike i statistike, pri otkrivanju odnosa i struktura među varijablama i kreiranju novog znanja. Podaci se često nalaze u skladištima podataka koja sadrže nepromjenjive podatke tolerante na pogreške prilikom transakcija. Skladišta često pohranjuju podatke iz mnogih izvora, podaci se čiste, spajaju te vežu uz varijable i sažimaju na niz korisnih načina.

Metode rudarenja postoje godinama, no tek u posljednjem desetljeću su postale strateški bitne u poslovnim organizacijama. Primjene rudarenja podataka se odnose na područje organizacije sa svojim postojećim ili budućim kupcima i klijentima[13], a to su:

- Upravljanje odnosima s kupcima
- Direktni marketing (ponude se šalju samo kupcima za koje postoji najveća vjerojatnost odaziva)
- Izrada profila kupaca (utvrđuje se uzorak ponašanja pojedinog kupca)

[13] GARAČA, Ž. i JADRIĆ, M. (2011.)
.

- Istraživanje povezanosti prodaje različitih proizvoda
- Predviđanje financijskog rizika
- Traženje potencijalnih zloupotreba u poslovanju
- Planiranje infrastrukture, transport, farmacija, itd.

Smisao rudarenja podataka je primijeniti niz tehnika i metoda koje predstavljaju alternativni pristup u procesu otkrivanja znanja. Primjeri primjene pokazuju da je rudarenjem podataka moguće tretirati niz problema, od financijskih i poslovnih do medicinskih, računalnih i slično.

4.1. Proces rudarenja podataka

Danas u rudarenju podataka postoje dvije dominantne metodologije: CRISP (*Cross Industry Standard Process for data mining* - metoda koja je nastala, inicijativom proizvođača i korisnika softvera za rudarenje, kako bi se proces standardizirao) i SEMMA (*Sample, Explore, Modify, Model, Assess* - metoda koju je razvio SAS institut zbog olakšavanja primjene statističkih i vizualizacijskih tehnika, odabira i transformacije prediktivnih varijabli, modeliranje varijabli za predviđanje izlaza te potvrđivanje vjerodostojnosti modela).[14]

1) CRISP metodologija podrazumijeva razumijevanje problema, razumijevanje podataka, pripremu podataka, modeliranje, evaluaciju modela i razvijanje modela. Kod razumijevanja problema trebalo bi odrediti cilj istraživanja, procjenu situacije i razvoj plana. Razumijevanje podataka čini početno prikupljanje, opis podataka, istraživanje te provjera kvalitete podataka. Kod pripreme podataka vrši se probiranje, čišćenje te transformiranje podataka u odgovarajući oblik. Modeliranje se razlikuje od tehnike do tehnike. Rudarenje podataka virtualizacijom i klaster analizom je prikladno za početnu analizu. Bolje razumijevanje podataka omogućava razvijanje detaljnijeg modela prilagodbe podacima. Evaluaciju modela čine rezultati modela koji se vrednuju u kontekstu ciljeva postavljenih u prvoj fazi što dovodi do identifikacije drugih potreba. Prilikom razvijanja modela rudarenja podataka, modeli se mogu koristiti za potvrđivanje ranije postavljenih hipoteza ili za otkrivanje znanja i predikciju.

[14] GARAČA, Ž. i JADRIĆ, M. (2011.)

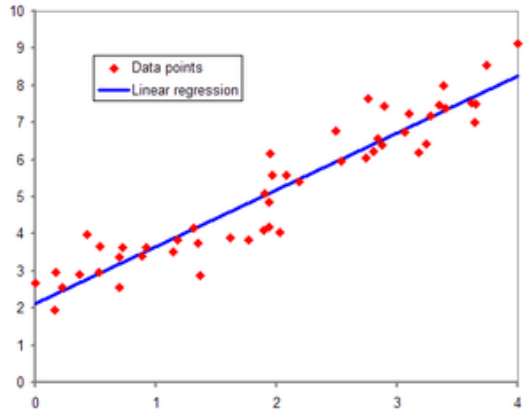
2) SEMMA (Sample Explore Modify Model Assess)

Uzorkovanje (engl. Sample) kreira tablice s podacima iz prikupljenih podataka. Uzorak bi trebao biti dovoljno velik da sadrži informacije, ali i dovoljno malen kako bi se rudarenje moglo provesti u što manjem vremenskom roku. Istraživanje (engl. Explore) traži očekivane veze i neočekivane anomalije u podacima radi njihova boljeg razumijevanja i pronalaska mogućih rješenja problema. Modificiranje (engl. Modify) uključuje modifikaciju podataka kreiranjem, odabiranjem i transformacijom varijabli na koje se istraživač fokusira prilikom modeliranja. Modeliranje (engl. Model) kombinira analitičke alate radi pronalaska kombinacije koja pouzdano predviđa ciljanu varijablu. Procjena (engl. Assess) podrazumijeva procjenjivanje podataka ocjenom njihove upotrebljivosti i pouzdanosti rezultata iz procesa rudarenja.

4.2. Metode rudarenja podataka

Cilj rudarenja podataka i statističkih analiza je opisivanje podataka i otkrivanje znanja unutar njih pomoću odgovarajućih metoda i tehnika te izgradnja modela koji predviđaju prošle događaje, a mogu se koristiti i za predviđanje budućih. Postoje različite metode za analiziranje prikupljenih podataka, a sve za cilj imaju prikaz dosadašnjih kretanja u podacima te izvođenje zaključaka što podaci znače u širom kontekstu. Ako se koriste različite metode rudarenja podataka, mogu se analizirati isti podaci pomoću različitih algoritama i sagledati iz različitih perspektiva. Svaka metoda se sastoji od niza pravila, funkcija i jednadžbi za otkrivanje značajnih uzoraka u podacima i predviđanju kretanja. Metoda bi trebala biti u skladu s tipovima podataka i tehnikom modeliranja kako bi provedeno rudarenje bilo što uspješnije. Uspješno rudarenje ovisi i o kvaliteti podataka te preciznom definiranju problema koji se istražuje. Proces pretprocesiranja olakšava izbor odgovarajuće metode. Poželjno je prilikom rudarenja koristiti različite metode kako bi se dobiveni rezultati mogli usporediti. Najčešće korištene metode rudarenja podataka su: regresijska analiza (logistička i linearna - normalna regresija za predikciju, logistička regresija za klasifikaciju) - jedna od najstarijih i najčešće korištenih. Koristi se za objašnjavanje karakteristika određenih pojava te predviđanja vrijednosti zavisne varijable pri zadanim vrijednostima nezavisnih varijabli. Logistička regresija se najčešće primjenjuje u biologiji i medicini. Regresija se općenito može podijeliti na jednostruku (postoji samo jedna zavisna i samo jedna nezavisna varijabla) i višestruku (postoji jedna zavisna i više nezavisnih varijabli). Linearnom regresijom se modelira odnos između jedne ili više zavisnih ili nezavisnih varijabli, a naziv tehnike upućuje

kako je zavisna varijabla linearna kombinacija parametara nezavisnih varijabli. Polazi se od prvotne pretpostavke koja kaže da je veza između promatranih varijabli linearna tj. da se jedna varijabla može iskazati kao linearna funkcija druge. Logistička regresija je izrazito popularan klasifikacijski algoritam koji se koristi kod varijabli binarnih, nominalnih ili ordinalnih obilježja podataka, pri čemu se matematički istražuje povezanost nekoliko nezavisnih varijabli s jednom zavisnom koja je dihotomnog obilježja (može poprimiti vrijednosti obilježja 0 i 1). Popularnost joj se temelji na logističkoj funkciji koja ima oblik S krivulje.

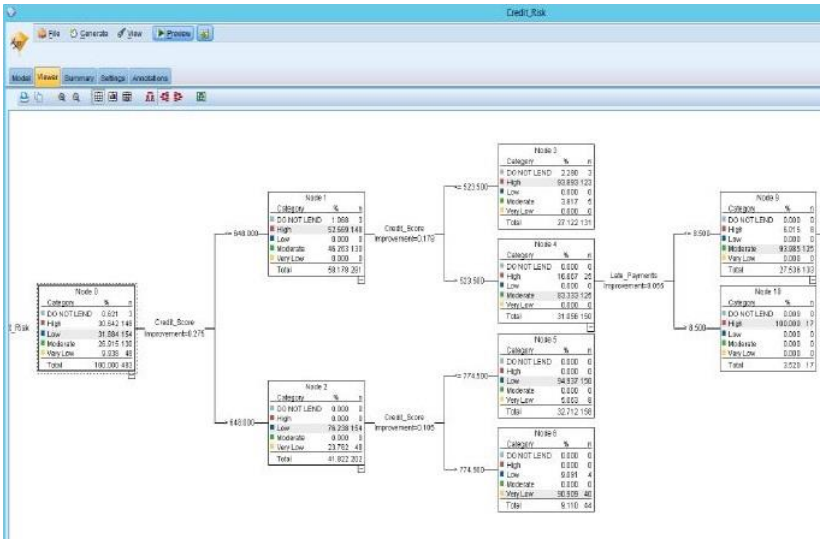


Slika 7: Primjer linearne regresije s jednom nezavisnom varijablom
(Izvor: https://hr.wikipedia.org/wiki/Linearna_regresija, 12.8.2015.)

Na slici je prikaz jednostruke regresijske linearne analize u kojoj postoji jedna zavisna i jedna nezavisna varijabla u kojoj je zavisna varijabla linearna kombinacija parametara nezavisne varijable.

Stabla odlučivanja: koriste se za klasifikaciju atributa s obzirom na ciljnu varijablu te procjenu ciljne varijable. Glavni algoritam stabla odlučivanja izvodi operacije kojima nastoji podijeliti podatke pomoću nezavisnih varijabli u podgrupe koje su homogene prema određenim obilježjima. Stabla odlučivanja su izrazito popularna zbog mogućnosti grafičkog prikaza rezultata, ali i zbog činjenice da su visoko fleksibilna . Klasični algoritmi koji se koriste u stablima odlučivanja su CLS, ID3, 14.5, CART i dr. Oblik stabla se formira pomoću čvorova odnosno grana koje proizlaze iz glavnog čvora, korijena, s tim da se čvorovi iz kojih se pružaju grane nazivaju testni čvorovi, a krajnji, proizašli iz njih, čvorovi odluke odnosno listovi. Neki od navedenih algoritama izgrađuju binarna stabla, dok neki razgranatija, a također se razlikuju i po načinu na koji se vrši grananje – *bottom up* ili *top down* pristup. Za razliku od regresijske

metode može raditi sa svim tipovima podataka, uključujući i nedostajuće vrijednosti. Ova metoda može poslužiti prilikom analize (ne)sklonosti potrošača kupnji nekog proizvoda ili pri segmentaciji tržišta.

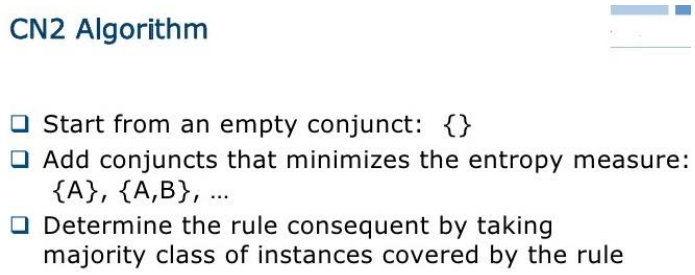


Slika 8: Izgled rezultata stabla odlučivanja

(Izvor: http://www.skladistenje.com/primjena-metoda-rudarenja-podataka-na-primjeru-procijene-kreditnog-rizika-dio-2/, 12.8.2015.)

Prikaz izgleda rezultata dobivenih metodom stabla odlučivanja. Oblik stabla čine čvorovi, grane i listovi. Grane su testni čvorovi, listovi čvorovi odluke.

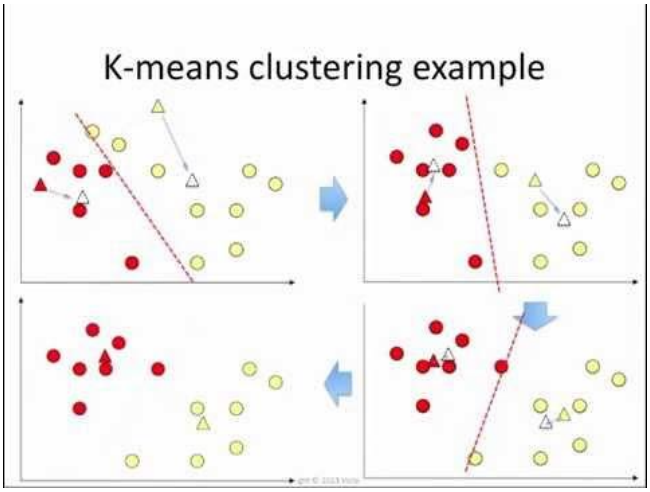
CN2 pravila: CN2 je hibridni algoritam koji kombinira pravila sa stablom odlučivanja. Vrlo moćna i popularna tehnika modeliranja koja se koristi za rješavanje klasifikacijskih i predikcijskih problema. Ako se usporede sa stablima odlučivanja, pravila su lakša za razumijevanje jer svako pravilo se odnosi na dio znanja koje se može interpretirati odvojeno, bez potrebe da se u interpretaciju uvode ostala pravila, ali s tim da pravilo odluke bi trebalo biti interpretirano kao cjelina.



Slika 9: Prikaz CN2 algoritma
(Izvor: http://www.slideshare.net/pierluca.lanzi/machine-learning-and-data-mining-12-classification-rules, 12.8.2015.)

Na slici je prikaz CN2 algoritma pomoću kojeg se rješavaju klasifikacijski i predikcijski problemi prilikom rudarenja podataka. Algoritam započinje tako što treba krenuti od prazne veze {}, dodaje se veza koja smanjuje mjere entropije {A}, {A,B}, određuju se pravila posljedica uzevši u obzir većinu instanci koje su obuhvaćene pravilima.

Klasteriranje: označava postupak pronalaska grupa sa sličnim obilježjima unutar podataka. Cilj je pronaći grupe koje se razlikuju jedna od druge, a elementi unutar grupa su slični jedni drugima. Nedostatak ove metode je činjenica da je klaster analiza jednostavniji statistički postupak te se ne treba na nju osloniti prilikom analize podataka. Osim toga, struktura klaster analize je labilna zbog osjetljivosti na odstupajuće vrijednosti. Postoji veliki broj algoritama za klasteriranje od kojih su osnovni: k - means algoritam i hijerarhijski algoritam. Osnovna karakteristika k – means metode je dijeljenje osnovne populacije na k segmente, pri čemu svaki od segmenata sadrži n sličnih elemenata. Sličnost elemenata algoritam procjenjuje na temelju funkcije udaljenosti. Hijerarhijsko klasteriranje polazi od grupiranja objekata u stablo klastera. Dijeli se na aglomerativno i divizijsko hijerarhijsko klasteriranje, ovisno o smjeru particioniranja koje se može kretati od vrha prema dnu ili obrnuto. Glavni nedostatak hijerarhijskih algoritama je nemogućnost ponavljanja procesa klasifikacije na istoj razini stabla nakon što je jednom izvršeno dijeljenje populacije u klastere.



Slika 10: Primjer K-means klasteriranja

(Izvor: https://www.youtube.com/watch?v=_aWzGGNrcic, 12.8.2015.)

Prikaz podataka dobivenih metodom klasteriranja. Cilj je pronaći grupe koje se razlikuju jedna od druge, a elementi unutar grupa su slični jedni drugima.

4.3. Primjer rudarenja podataka

Prikazat će se jednostavan primjer iz rudarenja podataka[15] na izmišljenim vremenskim podacima koji pomažu u donošenju odluke hoće li se igrati neka nespecifična igra ili ne. Instance u skupu podataka se karakteriziraju po vrijednostima značajki ili atributa kojima se mjere različiti aspekti instanci. U ovom slučaju postoje 4 atributa: outlook (prognoza), temperature (temperatura), humidity (vlaga) i windy (vjetrovito). Ishod je odluka igrati ili ne igrati. Sva četiri atributa su vrijednosti koje su simbolički kategorije, a ne brojevi (Tablica 1). Prognoza može biti sunny (sunčano), overcast (oblačno) ili rainy (kišovito) vrijeme. Temperatura može biti hot (vruće), mild (blaga), cool (hladno). Vlažnost može biti high (visoka) ili normal (normalna). Vjetar može biti true (istinito) ili false (lažno). To stvara 36 mogućih kombinacija od kojih je 14 prisutno u skupu ulaznih primjera.

[15] http://searchbusinessanalytics.techtarget.com/feature/Simple-data-mining-examples-and-datasets , 4.9.2015

Tablica 1: Podaci za rudarenje

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

(Izvor: http://searchbusinessanalytics.techtarget.com/feature/Simple-data-mining-examples-and-datasets, 4.9.2015.)

Set pravila iz ovih informacija izgleda ovako:

If outlook = sunny and humidity = high then play = no

If outlook = rainy and windy = true then play = no

If outlook = overcast then play = yes

If humidity = normal then play = yes

If none of the above then play = yes

Pravila se okidaju sekvencijalno. Pravila koja se objašnjavaju u primjeru nazivaju se popisom odluka. Pravilima se primjeri iz tablice klasificiraju točno. Promatra li se primjer pojedinačno i izvan konteksta, neka pravila su onda netočna. (npr. pravilo if humidity = normal, then play = yes rezultira jednim netočnim pravilom). Značenje pravila je povezano s načinom na koji se interpretiraju.

U nešto kompliciranijem obliku, prikazano u tablici 2, dva atributa – temperatura i vlažnost – imaju numeričke vrijednosti. Ovo znači da bi bilo koja metoda trebala kreirati nejednakosti koje uključuju ove atribute umjesto jednostavnih ravnopravnih testova. Ovaj problem se naziva problem numeričkih atributa, točnije u ovom slučaju problem miješanih atributa jer nisu svi atributi numerički.

Tablica 2: Mješovite vrijednosti atributa

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	85	85	False	No
Sunny	80	90	True	No
Overcast	83	86	False	Yes
Rainy	70	96	False	Yes
Rainy	68	80	False	Yes
Rainy	65	70	True	No
Overcast	64	65	True	Yes
Sunny	72	95	False	No
Sunny	69	70	False	Yes
Rainy	75	80	False	Yes
Sunny	75	70	True	Yes
Overcast	72	90	True	Yes
Overcast	81	75	False	Yes
Rainy	71	91	True	No

(Izvor: http://searchbusinessanalytics.techtarget.com/feature/Simple-data-mining-examples-and-datasets, 4.9.2015.)

Sada prvo pravilo mijenja oblik:

If outlook = sunny and humidity > 83 then play = no

Malo složeniji proces je potreban za stvaranje pravila koja uključuju numeričke testove.

Pravila koja su dosad viđena su klasifikacijska pravila: ona predviđaju klasifikaciju primjera u smislu hoće li igrati ili ne. Moguće je i zanemariti klasifikaciju i samo tražiti bilo koja pravila koja snažno povezuju različite vrijednosti atributa. Nazivaju se pravilima pridruživanja. Primjeri pravila pridruživanja koja se mogu izvesti iz navedenih podataka su :

If temperature = cool *then humidity = normal*

If humidity = normal and windy = false *then play = yes*

If outlook = sunny and play = no *then humidity = high*

If windy = false and play = no *then outlook = sunny and humidity = high.*

Sva ova pravila su 100% točna na danim podacima, ne čine lažna predviđanja. Prva dva se odnose na četiri primjera u skupu podataka, treće na tri primjera, a četvrto na dva primjera. Postoje mnoga druga pravila, ustvari gotovo 60 pravila pridruživanja se mogu pronaći i primijeniti na dva ili više primjera vremenskih podataka i potpuno su točna na ovim podacima. Ako se traže pravila koja su točna manje od 100% tada će se pronaći puno više pravila. Postoji ih jako puno zato što za razliku od klasifikacijskih pravila, pravila pridruživanja mogu predvidjeti bilo koji od atributa, a ne samo određenu klasu te mogu predvidjeti više od jedne stvari (npr. četvrto pravilo predviđa da će vrijeme biti sunčano i da će vlažnost biti visoka).

5. IBM Cognos BI

Cognos Incorporated (prvobitno nazvana *Quasar*) je kompanija koja se nalazi u sklopu IBM-a, a bavi se proizvodnjom softvera za poslovnu inteligenciju te upravljanjem rezultatima poslovanja. Cognos je i ime platforme koja se bavi analizom poslovnih baza podataka nekog poduzeća. Od poslovne inteligencije te financijskih rezultata i strategija upravljanja do analize aplikacija, Cognos softver pruža ono što organizaciji treba za postizanje vrhunskih rezultata. Dizajniran je kako bi svatko u organizaciji mogao donositi odluke kojima se postižu bolji poslovni rezultati. Osim toga, povećava povrat investicije i smanjuje ukupne troškove posjedovanja.

Cognos platforma omogućava:

- Učinkovitost i nadogradnju,
- Međusobno djelovanje velikih podataka,
- Zajednički pregled podataka iz svih sučelja,
- Fleksibilnu implementaciju,
- Arhiviranost sadržaja,
- Eliminaciju grešaka,
- Vlasništvo nad podacima,

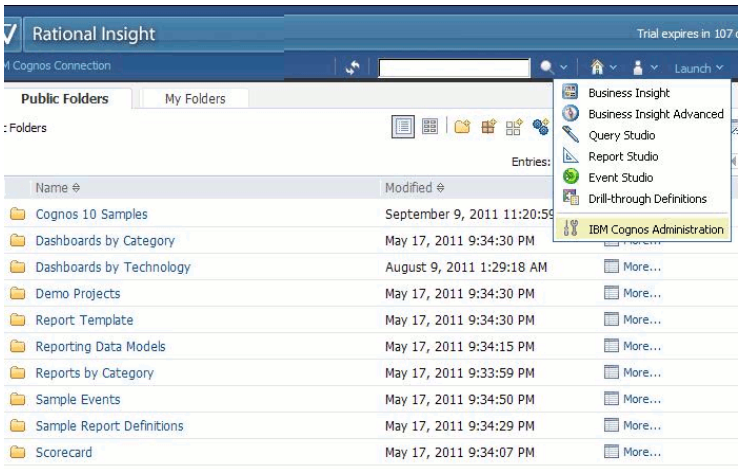
IBM je 2010. je predstavio Cognos 10, novu verziju ranijeg Cognos 8 BI koja između ostalog omogućava i korištenje aplikacije putem interneta pomoću mobilnih uređaja. Cognos sadrži sljedeće alate: IBM Cognos 8 BI, IBM Cognos Planning, IBM Cognos TM1, IBM Cognos 8 Controller, IBM Cognos TM1 Executive Viewer, IBM Cognos Finance, IBM Cognos 8 Buisness Viewpoint. Cognosov sustav poslovne inteligencije pruža jedinstveni radni prostor za poslovnu inteligenciju i analitiku koji može koristiti cijela organizacija za odgovaranja na ključna poslovna pitanja i nadmašivanje konkurencije. S ovim sustavom korisnici mogu:

- Jednostavno pregledati, sastaviti i prilagoditi podatke
- Istražiti sve tipove informacija iz svih kutova kako bi se procijenila trenutna poslovna situacija
- Analizirati činjenice i predvidjeti taktičke i strateške implikacije
- Komunicirati i koordinirati zadatke kako bi se angažirali pravi ljudi u pravo vrijeme

- Korištenjem mobilnih uređaja bilo gdje pristupiti informacijama

Osnovne komponente sustava su:

Cognos Connection – web portal za IBM Cognos BI. To je polazna točka za pristup svim funkcijama u sklopu softvera. Uz pomoć portala može se pretraživati sadržaj u obliku izvješća ili tablica sa rezultatima te upravljati njime, strukturirati ili ga prikazati. Osim toga, portal se koristi za kreiranje zadataka, raspored i distribuciju izvješća, stvaranje prečaca, URL-ova itd.

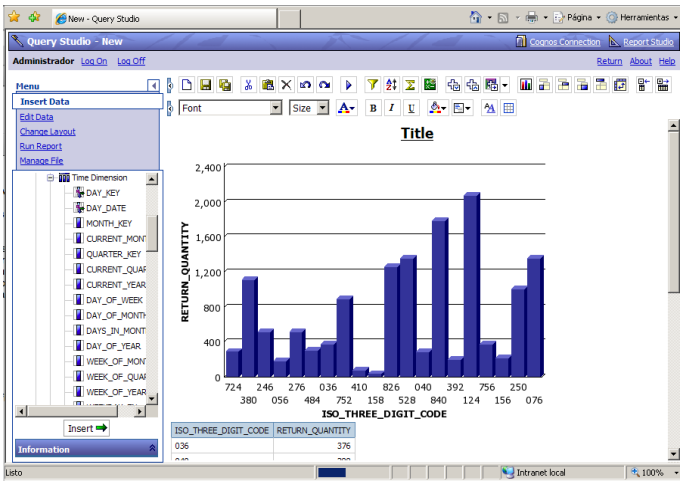


Slika 11: Prikaz IBM Cognos Connectiona

(Izvor: https://jazz.net/help-dev/rational-insight/index.jsp?topic=%2Fcom.ibm.rational.raer.install.doc%2Ftopics%2Ft_sample_datasourcecon.html 16.8.2015.)

Na slici je prikaz IBM Cognos web portala putem kojega se može pretraživati, strukturirati ili prikazivati sadržaj u obliku izvješća ili tablica.

Query Studio – omogućava jednostavne upite i izvješća na jednostavna poslovna pitanja. Izgled izvješća se može prilagoditi i podaci se mogu filtrirati i sortirati, također je podržano oblikovanje i stvaranje dijagrama.



Slika 12: Prikaz dijagrama Query studija

(Izvor: http://www.dataprix.com/en/introduction-ibm-cognos-bi-business-intelligence-suite-ibm, 16.8.2015.)

Na slici je prikaz izvješća Query studija koje je oblikovano u obliku dijagrama.

Report Studio – koristi se za izradu izvješća uprave koja se mogu slagati putem interneta. Izvještaji su većinom statični i u formatima Excel, PDF (ne omogućuje daljnje analize), HTML, XML.

Analysis Studio - Korisnici mogu stvarati analize velikih izvora podataka i tražiti pozadinske informacije o događaju ili aktivnosti. Višedimenzionalna analiza omogućuje identificiranje trendova i razumijevanje anomalija ili odstupanja koje nisu očite u drugim vrstama izvješća. U analizu mogu biti uključene drag and drop značajke, elementi i ključni pokazatelji uspješnosti , redci i stupci, OLAP funkcionalnosti (za dublje razumijevanje izvora informacija koje se koriste u analizi).

Event Studio – informira o događajima unutar poduzeća u realnom vremenu. Obavijesti se šalju putem e-maila, objavom na portalu ili pokretanjem izvješća.

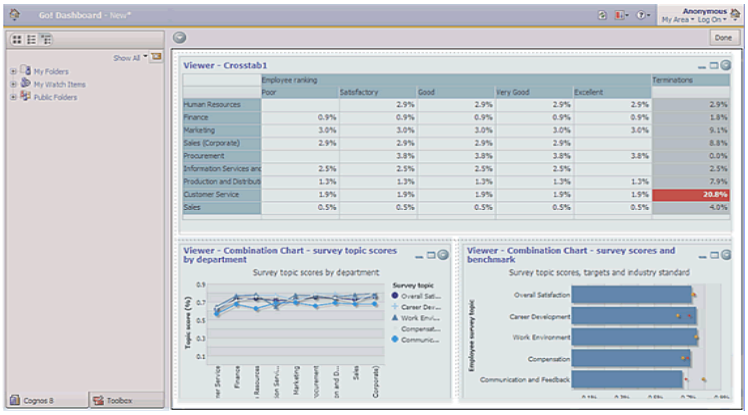
Dodatne komponente sustava su:

Go! Office – omogućuje korisnicima rad sa sadržajem IBM Cognos u svom poznatom Microsoft Office okruženju. Komponenta omogućuje pristup svim sadržajima izvješća IBM Cognosa, uključujući podatke, metapodatke, zaglavlja i podnožja te dijagrame. Korisnici mogu koristiti predefinirana izvješća ili stvoriti nove sadržaje pomoću Query Studio, Analysis Studio ili Report Studio. Uvozom sadržaja u Microsoft Excell korisnici mogu koristiti značajke oblikovanja,

proračuna i prikaza. Kreirani dokumenti se onda mogu uvesti pomoću Cognos Connection te objaviti kako bi bili dostupni i drugim korisnicima.

Go! Search – u Cognos Connectionu može se pretražiti tekst koji se nalazi u izvješćima, analizama, pločama i događajima. Pretraživanje cijelog teksta u IBM Cognos Go! Search je povezano sa pretragom u redovitim tražilicama kao što je Google. Korisnici mogu pretraživati operatere kao što su +, -, ili koristiti navodnike za promjenu zadanog postupka pretrage pomoću više riječi. Rezultati pretraživanja su razvrstani u silaznom redoslijedu, unos sa najvećom količinom relevantnih metapodataka se prikazuje na vrhu popisa.

Go! Dashboard – pomoću IBM Cognos Go! Dashboard mogu se kreirati interaktivne ploče koje sadrže IBM Cognos sadržaje i vanjske izvore podataka kako bi odgovarale individualnim potrebama za informacijama. Stavke koje mogu biti dodane u nadzornu ploču su: izvješća predmeta (prikazana u Cognos Vieweru), dijelovi izvješća kao što su liste i dijagrami. Cognos Search omogućuje pretraživanje objavljenih sadržaja. Osim toga na ploči se mogu prikazati i web linkovi, web stranice, slike i RSS feedovi. Korisničko sučelje ima dva načina rada: u interaktivnom načinu rada postojeće ploče koje se mogu pregledati i koristiti, dok se kreiranje i uređivanje ploča obavlja u načinu za sastavljanje.

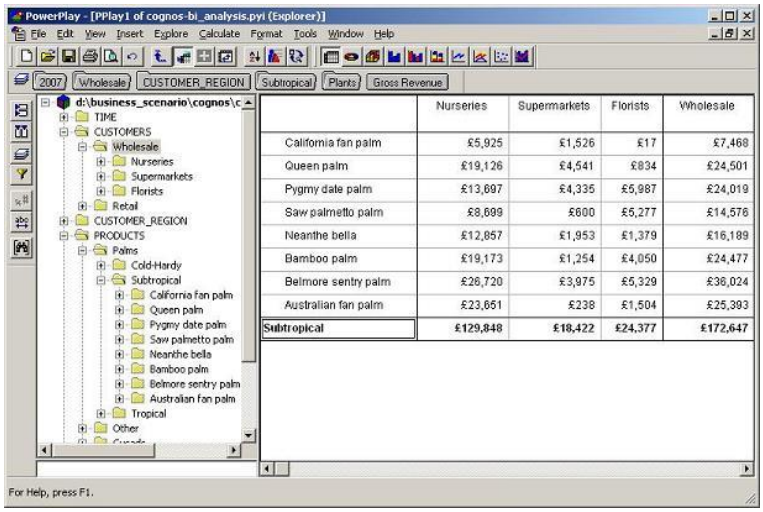


Slika 13: IBM Cognos Go! Dashboard

(Izvor: http://www.ibm.com/developerworks/data/library/cognos/infrastructure/cognos_specific/page497.html, 16.8.2015.)

Na slici je prikaz sučelja IBM Cognos Go! Dashboarda u kojoj se mogu dodavati razni izvještaji i dijagrami te se mogu prikazati i web linkovi i stranice, slike i RSS feedovi.

Cognos je vrlo jednostavan za korištenje, koristi se *drag and drop* (povuci i spusti) funkcijom koju razumiju svi korisnici. Cognosovi alati omogućuju tvrtki brz uvid u podatke, pretvaranje podataka u informacije, prezentiranje podataka i usporedbu ostvarenih pokazatelja s onima planiranim. Kako bi koristili Cognos ne treba imati instalirani program nego mu se može pristupiti preko različitih internetskih preglednika. Cognosom se kreće pomoću poveznica koje vode ka grupiranim podacima. Izvještajima i kockama može se pristupiti na web portalu Cognosa pomoću hiperlinka. Izvještaji su većinom statični i u formatu Excela ili PDF-a. PDF format ne omogućuje daljnje analize dok Excel omogućuje jednostavno prebacivanje u tablice i dorađivanje. Putem izbornika (prompt-a) može se prilikom izrade izvještaja odabrati parametre poput datuma, formata, itd.(parametri unutar izvještaja mogu biti obavezni ili opcionalni). Prompt-ovi se razlikuju ovisno o vrstama podataka i cilja izvještaja. Razlikuju se i zavisni prompt-ovi koji se biraju nakon odabira glavnog prompta. Cognos kocka je interaktivni pristup podacima koji su grupirani u nivoima unutar dimenzija (poslovni aspekti koji omogućuju analizu, odgovaraju na pitanja kada, što, gdje , kako). Pokretanjem kocke dobije se sučelje u kojem se prikazuju kategorije najvišeg nivoa prve dimenzije iz popisa u recima, a najviši nivoi druge dimenzije iz popisa u stupcima. Sve se izražava u mjeri (poslovni pokazatelji, brojevi) koja se nalazi prva na popisu.

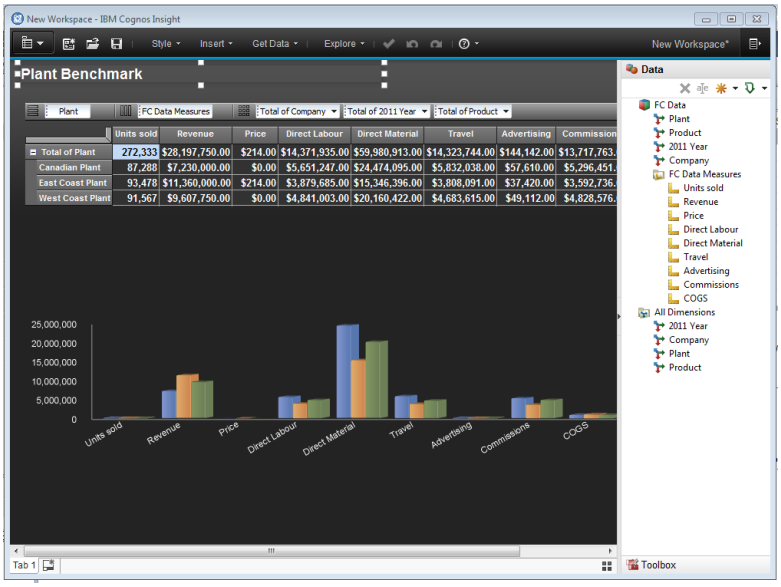


Slika 14: Sučelje Cognos kocke

(Izvor: http://cognosfaq.blogspot.hr/2012/12/cognos-transformer-tutorial.html, 16.8.2015.)

Na slici je prikaz koliko su zaradili rasadnici, supermarketi, cvjećari na određenim vrstama palmi.

Postoje i multidimenzijske Cognos kocke pomoću kojih se pretražuju podaci za dobivanje agregiranih podataka i omogućava se analiza za operativne potrebe tj. detaljne i pojedinačne podatke. Mogu se dobiti agregirani podaci na razini tvrtke ili detaljizirani podaci. Multidimenzijske kocke su oblikovani i optimizirani skupovi podataka u kojima se prikazuje svaka dimenzija u skladu s postavljenim parametrima. Podaci se mogu oblikovati na različite načine. Služenje Cognos kockom se odvija jednostavno pomoću odabira određenih funkcija. Prednost multidimenzijskih kocki je u tome što ih korisnik sam može kreirati kako želi te sačuvati i kasnije po potrebi pristupiti podacima koji mu trebaju.



Slika 15: Multidimenzijske Cognos kocke

(Izvor: http://www.performance-ideas.com/wp-content/uploads/2012/02/Cognos-Insight-Dimensions.png, 16.8.2015.)

Na slici je prikaz izgleda i rada multidimenzijskih Cognos kocki u kojima je svaka dimenzija oblikovana u skladu s postavljenim pravilima. U traci su postavljena pravila koja će se prikazati. Prikazani su rezultati iznosa postrojenja (Kanade, Istočne i Zapadne obale SAD-a) po prihodima, cijeni, materijalu, putovanjima, prodanim jedinicama i slično.

Cognos kockama pronalazi se niz različitih funkcija kojima se omogućuje kretanje po kocki, istraživanje, različite operacije i analiziranje. Sve ove funkcije olakšavaju donošenje poslovnih odluka tako što pružaju korisnicima mogućnost jednostavnog pretraživanja informacija, njihovu organizaciju, analiziranje i na kraju pretvorbu. Neke od važnijih funkcija kocki su: *drag and drop* (povuci i spusti – pomoću koje se pretražuju informacije), *drill down/up* (služi dobivanju

agregiranih ili detaljiziranih informacija), promjene dimenzija ili mjere (zamjena dimenzija u recima ili stupcima te odabir mjere koja korisnicima odgovara), filtracija podataka (ograničavanje podataka na one potrebne korisnicima), chart (izrađivanje grafova i grafičkih prikaza tablica), skrivanje nepotrebnih podataka (prikazivanje samo potrebnih podataka), promjene u izvještajima te spremanje izvještaja, računske operacije i mnoge druge. Pomoću ovih funkcija korisnik sam može jednostavno kreirati kocku i kretati se unutar nje. Kocke pružaju korisnicima jednostavnu pretragu podataka te filtraciju podataka, kojom se dobivaju samo oni potrebni, te raspodjelu tih podataka. Sve ovo ubrzava proces donošenja bitnih poslovnih odluka.



Slika 16: Alatna traka Cognos kocke u kojoj se nalaze osnovne funkcije
(Izvor: http://www.cognos-bi.info/tutorial/measures.htm, 16.8.2015.)

Na slici je prikazana alatna traka Cognos kocke na kojoj se može pristupiti funkcijama koje su potrebne za rad (drag and drop, drill down/up, chart i ostalo).

5.1. Usporedba Cognosa i drugog dobavljača – Pentaha

Pentaho je open source BI poduzeće koje pruža širok raspon alata koji pomažu korisnicima bolje upravljati svojim tvrtkama[16]. Ovi alati uključuju softver za integraciju podataka, alate za rudarenje podataka, dashboard aplikacije, online mogućnosti analitičke obrade podataka i još mnogo toga. Pentaho ima dvije kategorije proizvoda: standardna verzija za poduzeće - proizvod koji dolazi izravno iz Pentaha sa svim prednostima, značajkama i programima koji dolaze uz plaćene aplikacije kao što su analiza usluga, dizajn dashboarda i interaktivno izvješćivanje. Alternativa je korištenje open source verzije. Ova verzija ima svoje prednosti, osim činjenice što je besplatna, puno ljudi radi na ovom projektu kako bi poboljšali njegovu kvalitetu i širinu funkcionalnosti. Pentaho BI platforma (odnedavno preimenovana u Business Analytics – BA Platform) čini jezgru softvera koja sadržava sadržaj kreiran i u samom poslužitelju putem dodataka i podatke objavljene na serveru iz aplikacija na radnoj površini. Uključuje značajke za upravljanje sigurnošću, izvođenje izvješća, korištenje dashboarda, OLAP analizu i raspoređivanje iz skladišta. Pentaho BA platforma izvodi se u serveru Apache Java aplikacije. Modularna

[16] https://en.wikipedia.org/wiki/Pentaho 5.9.2015

arihitektura omogućuje izgradnju sustava tako da se u kratkom roku stvore preduvjeti za potpunu funkcionalnost u svim dijelovima BI funkcionalnosti.[17]



Slika 17. Prikaz sučelja BI platforme Pentaha

(Izvor: http://sourceforge.net/projects/pentaho/ 5.9.2015.)

Na slici je prikazano sučelje i način rada u Dashboardu Pentaho BA platforme.

Cognos softverski paket je najbolji za tvrtke srednjeg do velikog razmjera. Proizvod je osmišljen kako bi pomogao organizacijama da postanu visoko rangirane. Donose programe za različite organizacijske veličine. Ima proizvode za pojedince, grupe, odjele, srednje ili velike tvrtke. Njegov glavni cilj je da svi mogu donijeti bitne poslovne odluke za poboljšanje i rast posla. Koristi ga više od 23.000 korisnika iz cijelog svijeta i zaslužio je povjerenje kupaca. Prednosti Cognos BI platforme su te da je to snažan analitički alat sa sučeljem koje je lako za navigaciju. Softver pruža kompletne BI mogućnosti kao što su *ad hoc* upiti i izvješća u jednoj arhitekturi. Paket nudi različita rješenja ovisno o veličini i potrebama organizacije kako bi svaka ostvarila svoje specifične ciljeve. Najnovija Cognos 10 verzija podržava mobilne platforme, sve od Androida i iOS-a do Windows i Blackberry mobilnih platorma. Nedostaci Cognosa su ti što korisnicima treba duže da nauče kako potpuno koristiti Cognos softver i integrirati ga sa raznim BI alatima. Osim toga, Cognos je oslabio u odnosu na pregled dokumenata jer ne dopušta

[17] http://www.vinteh.hr/rjesenja/pentaho.html 5.9.2015

jednostavan pristup svim izvješćima. Za razliku od Cognosa, Pentaho je savršen za male i rastuće poslovne tvrtke, industrije koje pružaju zdravstvenu zaštitu, maloprodaju, financijske usluge. Prednosti Pentaho BA platforme su tome što podržava vizualizaciju i analizu različitih podataka na više dimenzija. Pentaho nudi istinski mobilno iskustvo za korisnike koji ga mogu koristiti s bilo kojeg mjesta, osim toga pruža cjelodnevnu tehničku podršku korisnicima koji imaju poteškoća s korištenjem. Interaktivna vizualna analiza značajka pruža korisnicima neposredan pristup svim podacima, a poslovni analitički softver nudi 30 – dnevnu besplatnu probnu verziju za korisnike. Nedostaci Pentaho BA platforme su to što uključuje Weka modul za rudarenje podataka koji je manje sposoban i napredan u odnosu na druge slične alate. Osim toga, dokumentacija Pentaho BA platforme je slaba u usporedbi s drugim sličnim alatima i može korisnicima izazvati poteškoće pri korištenju. [18]

[18] http://business-intelligence.softwareinsider.com/compare/139-345/Cognos-vs-Pentaho-Business-Analytics 5.9.2015

6. ZAKLJUČAK

Promjene u poslovnoj okolini te sve jača konkurencija zahtijevaju od poduzeća promjene u načinu poslovanja. Za opstanak na tržištu potrebno je prilagoditi se novom vremenu i pratiti trendove u poslovanju. Da bi to mogla činiti poduzeća trebaju posjedovati prave informacije. Posjedovanje pravih i potrebnih informacija može učiniti veliku razliku u poslovanju poduzeća dok posjedovanje loših informacija dovodi do krivih poslovnih odluka i lošeg poslovanja. Korištenje informacijske tehnologije i njenih alata olakšava donošenje pravih odluka i omogućuje opstanak na tržištu. Sustavi poslovne inteligencije ubrzavaju proces prijenosa informacija, povećavaju njihovu učinkovitost; prikupljaju, analiziraju, spremaju i obrađuju informacije. Na taj način pomažu poduzećima pri donošenju poslovnih odluka. No, nije sve na samom sustavu. Kako bi ga korisnici uopće mogli koristiti trebali bi se educirati i znati s njim "surađivati" pri donošenju odluke. Sustavi poslovne inteligencije koriste brojne alate za uvid u podatke kao što su opisani OLAP (brz i efikasan uvid u podatke) i Data mining (proces otkrivanja uzoraka u velikim skupovima podataka). Da bi se ovi alati mogli primjenjivati, trebala bi postojati mjesta u koja se pohranjuju podaci, a nazivaju se skladišta podataka. Zbog napretka tehnologije i konkurencije na tržištu poduzeća su bila prisiljena početi koristiti ovakve sustave poslovne inteligencije kao što je opisani Cognos. Cognos omogućava jednostavnije korištenje, analiziranje podataka, mjerenja efikasnosti i poboljšanje funkcionalnosti te samim time pomaže tvrtkama pri poslovnom odlučivanju. Dolazi se do zaključka da je posjedovanje informacija ključ uspješnosti poduzeća. Bitno je posjedovati pravu informaciju u pravo vrijeme, a to se danas ne bi moglo bez pomoći sustava poslovne inteligencije. Potrebno je još dosta educiranja korisnika ovih sustava kako bi uvidjeli njihovu stvarnu korist i pomoć u poslovanju.

7. POPIS LITERATURE

Knjige

1. PANIAN, Ž. I KLEPAC, G. (2003.) *Poslovna inteligencija*. Zagreb. MASMEDIA d.o.o.
2. JAVOROVIČ, B. i BILANDŽIĆ, M. (2007.) *Poslovne informacije i business intelligence.* Zagreb. Golden marketing – Tehnička knjiga.
3. GARAČA, Ž. i JADRIĆ, M. (2011.) *Rudarenje podataka: Različiti aspekti informacijskog društva.* Split. Ekonomski fakultet u Splitu
4. GARAČA, Ž. i ĆUKUŠIĆ, M. (2011.) *Višedimenzijski informacijski sustavi – Skladištenje i analitička obrada podataka*. Split. Ekonomski fakultet u Splitu.
5. BANEK, M. (2005.) *Oblikovanje skladišta podataka iz polustrukturiranih izvora* (magistarski rad). Zagreb. Fakultet elektrotehnike i obrazovanja.
6. INMON, W.H. (2005.) *Building the Data Warehouse*. 4th Edition. Indianapolis. John Wiley & Sons.
7. PONNIAH, P. (2001.*) Data Warehousing Fundamentals: A Comprehensive Guide for IT Proffessionals*. New York. John Wiley & Sons, Inc.

Internet

1. http://www.pravinshetty.com/StudyMaterials/Concepts/Wiley%20&%20Sons%20-%20Building%20The%20Data%20Warehouse.%20Third%20Edition.pdf, [pristupljeno 7.8.2015.]
2. http://ekonomskiportal.com/poslovna-inteligencija-zasto-sto-i-kako/ [pristupljeno 7.8.2015.]
3. http://web.efzg.hr/dok/INF/Ceric/itup_knjiga/skladista_podataka.pdf, [pristupljeno 10.8.2015.]
4. http://searchbusinessanalytics.techtarget.com/feature/Simple-data-mining-examples-and-datasets, [pristupljeno 4.9.2015]
5. http://www.vinteh.hr/rjesenja/pentaho.html pristupljeno, [pristupljeno 4.9.2015.]
6. http://business-intelligence.softwareinsider.com/compare/139-345/Cognos-vs-Pentaho-Business-Analytics, [pristupljeno 4.9.2015.]

7. https://en.wikipedia.org/wiki/Pentaho, [pristupljeno 5.9.2015.]
8. http://www.skladistenje.com/pojmovi/ [pristupljeno 5.9.2015.]

POPIS TABLICA

1. Tablica br. 1 - Podaci za rudarenje
2. Tablica br 2 - Mješovite vrijednosti atributa

POPIS SLIKA

1. Slika br. 1 - Komponente modela poslovne inteligencije
2. Slika br. 2 - Arhitektura sustava BI
3. Slika br. 3 -Pozicija skladišta podataka u informacijskom sustavu poduzeća
4. Slika br. 4 - Koncept ETL procesa
5. Slika br. 5 - Struktura OLAP sustava
6. Slika br. 6 - Inačice OLAP alata
7. Slika br. 7 - Primjer linearne regresije s jednom nezavisnom varijablom
8. Slika br. 8 - Izgled rezultata stabla odlučivanja
9. Slika br. 9 - Prikaz CN2 algoritma
10. Slika br. 10 - Primjer K-means klasteriranja
11. Slika br. 11 - Prikaz IBM Cognos Connectiona
12. Slika br. 12 - Prikaz dijagrama Query studija
13. Slika br. 13 - IBM Cognos Go! Dashboard
14. Slika br. 14 - Sučelje Cognos kocke
15. Slika br. 15 - Multidimenzijske Cognos kocke
16. Slika br. 16 - Alatna traka Cognos kocke u kojoj se nalaze osnovne funkcije
17. Slika br. 17 - Prikaz sučelja BI platforme Pentaha

SAŽETAK

U današnje vrijeme zbog napretka tehnologije i jake konkurencije na tržištu, poduzeća su počela koristiti sustave poslovne inteligencije u poslovnom odlučivanju. Uspješnost se počela temeljiti na posjedovanju pravih informacija. Pomoću ovih sustava unapređuje se poslovanje, lakše i brže se dolazi do informacija koje se analiziraju i pohranjuju u skladišta te se olakšava donošenje poslovnih odluka. Poduzeća koriste razne alate BI kao što su OLAP i rudarenje podataka ili aplikacijama kao što je Cognos. Korištenje sustava BI postalo je ključno za uspješno poslovanje.

Technological advances and market competitiveness are the main reasons for the increasing use of business intelligence systems. Possesion of proper information is considered as the foundation of a successful business. BI transforms data into useful information which are analyzed and stored in data warehouses. With the use of BI we can create new business opportunities and simplify the making of important business decisions. BI use tools like OLAP and Data Mining or applications such as Cognos.