



UNIVERZITET SINGIDUNUM
DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

STUDIJSKI PROGRAM
SAVREMENE INFORMACIONE TEHNOLOGIJE

– MASTER RAD –

**PREGLED PRIMENE INTELIGENTNIH SISTEMA,
POSLOVNE INTELIGENCIJE
I ANALIZA POTROŠAČKE KORPE
(Market Basket Analyze)**

Mentor:

Prof. dr Milan Milosavljević

Student:

Marko Štalmatović

410291/2014

Beograd, 2016.



UNIVERZITET SINGIDUNUM
Departman za poslediplomske studije

STUDIJSKI PROGRAM
SAVREMENE INFORMACIONE TEHNOLOGIJE

– MASTER RAD –

**PREGLED PRIMENE INTELIGENTNIH SISTEMA,
POSLOVNE INTELIGENCIJE
I ANALIZA POTROŠAČKE KORPE
(Market Basket Analyze)**

Mentor:

Prof. dr Milan Milosavljević

Student:

Marko Štalmatović

410291/2014

Beograd, 2016.

Pregled primene inteligentnih sistema i poslovne inteligencije u elektronskom poslovanju

Sažetak:

Predmet ovog rada predstavlja pregled primene inteligentnih sistema i definisanje pojma poslovne inteligencije kroz karakteristike savremenih Business Intelligence (BI) sistema. Sa osvrtom na elektronsko poslovanje, elektronsku trgovinu i na njihove osnovne komponente Enterprise Resource Planning (ERP), Client/Customer Relationship Management (CRM). U nastavku je opisana personalizacija, praćenje potreba i interesovanja korisnika za izradu korisničkih profila i analiza pojedinih elemenata arhitekture sistema poslovne inteligencije. Zatim je prikazan sistem poslovne inteligencije u elektronskom poslovanju kroz opis metoda i tehnika proučavanja sadržaja kao i primena mašinskog učenja u sistemima za preporučivanje proizvoda na primeru sajta "Amazon". Praktičan deo rada predstavlja analizu konkretnog seta podataka kroz tehniku Market Basket analize u softveru R studio i definisanje analize kao tehniku Data Mining-a. Kao i opis izabrane analize, sa osvrtom na pravila asocijacije i principe Apriori algoritma, opis interfejsa, glavnih karakteristika i opcija R studija.

Ključne reči: inteligentni sistemi, poslovna inteligencija, elektronsko poslovanje, elektronska trgovina, BI, ERP, CRM, Market Basket, pravila asocijacije, Apriori algoritam.

Overview of intelligent systems and business intelligence in e-Business

Abstract:

The subject of this work is overview of the intelligent systems and definition of the concept of business intelligence through features of modern Business Intelligence (BI) systems. With regard to e-business, e-commerce and their basic components Enterprise Resource Planning (ERP), Client/Customer Relationship Management (CRM). The following describes personalization, needs tracking and interests of users to create user profiles and analysis of elements architecture of business intelligence systems. In a separate part of this work presents the system of business intelligence in e-Business through the description of the methods and techniques study of the content and application of machine learning systems to recommend products to the example site "Amazon". Special part of this work is the analysis of specific data set through the technique of Market Basket Analysis in R studio software and defining the Market Basket analysis as Data Mining technique. The beginning of work description chosen analysis with Association Rules and principle of Apriori algorithm. Sequel describe the main features and options of R studio as used software in analysis.

Key words: Intelligent System, Business intelligence, e-Business, e-Commerce, BI, ERP, CRM, Market Basket Analyze, Association rules, Apriori algorithm.

Sadržaj:

2	Definisanje pojma poslovne inteligencije i njenih osnovnih komponenti....	4
2.1	Definisanje pojma poslovne inteligencije.....	4
2.2	Sistemi za upravljanje odnosa sa klijentima – CRM.....	6
2.3	Sistemi za planiranje poslovnih resursa – ERP	8
2.4	Implementacija ERP rešenja.....	10
3	Opis karakteristika elektronskog poslovanja i elektronske trgovine	11
3.1	Elektronsko poslovanje.....	11
3.2	Elektronska trgovina.....	12
4	Skladištenje i analiza podataka	13
4.1	Skladištenje i analiza podataka.....	13
4.2	Rudarenje podataka - Data Mining.....	14
5	OLAP sistemi	15
5.1	OLAP sistemi.....	15
5.2	Arhitektura OLAP sistema	16
6	Opis metoda i tehnika preporučivanja sadržaja na sajtu "Amazon"	19
6.1	Struktura sajta "Amazon"	19
7	Market Basket analiza.....	22
7.1	Uvod u Market Basket analizu	22
7.2	Pravila asocijacije	23
7.3	Apriori algoritam	24
8	Opis softvera	25
8.1	R studio.....	25
9	Studija slučaja.....	30
9.1	Uvod u analizu.....	30
9.2	Analiza seta podataka “Namirnice” [8]	30
10	Zaključak.....	36
	Literatura	37
	Lista slika	38

1 Uvod

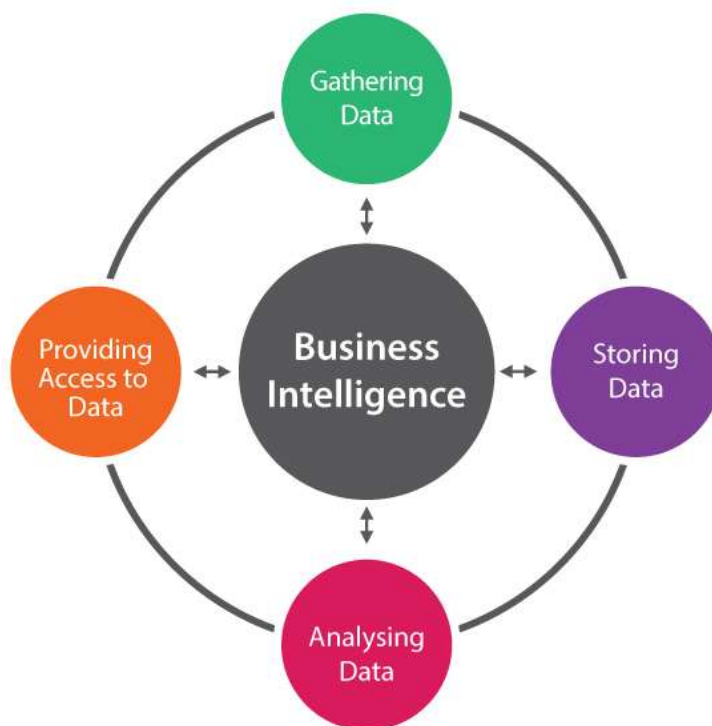
Ovaj rad je organizovan u deset poglavlja, počevši od ovog uvodnog dela. U drugom poglavlju je opisan istorijat, pojam i osnovne komponente poslovne inteligencije, kao što su ERP i CRM. Zatim sledi definisanje elektronskog poslovanja i elektronske trgovine, sa opisom vrsta elektronske trgovine i njihovom primenom. Naredno poglavlje bavi se skladištenjem i analizom podataka. Opisuje osnovne elemente i metode Data Warehousing-a i Data Mining-a. OLAP sistemi su predmet sledećeg poglavlja sa opisom arhitekture takvih sistema. Zatim sledi opis metoda i tehnika preporučivanja sadržaja korisnicima kroz primer sajta "Amazon". U sledećem poglavlju je opisan pojam, značaj i primena Basket Market analize. U nastavku sledi opis pravila asocijacije i Apriori algoritma. Zatim je opisan korišćeni softver, u ovom slučaju to je R studio. Opisane su njegove osnovne opcije i karakteristike. Poslednje poglavlje pre zaključka bavi se analizom konkretnog seta podataka kroz R studio. Objašnjavaju se ciljevi analize i radnje koje su preduzete da bi se analiza obavila i dala tražene rezultate.

2 Definisanje pojma poslovne inteligencije i njenih osnovnih komponenti

2.1 Definisanje pojma poslovne inteligencije

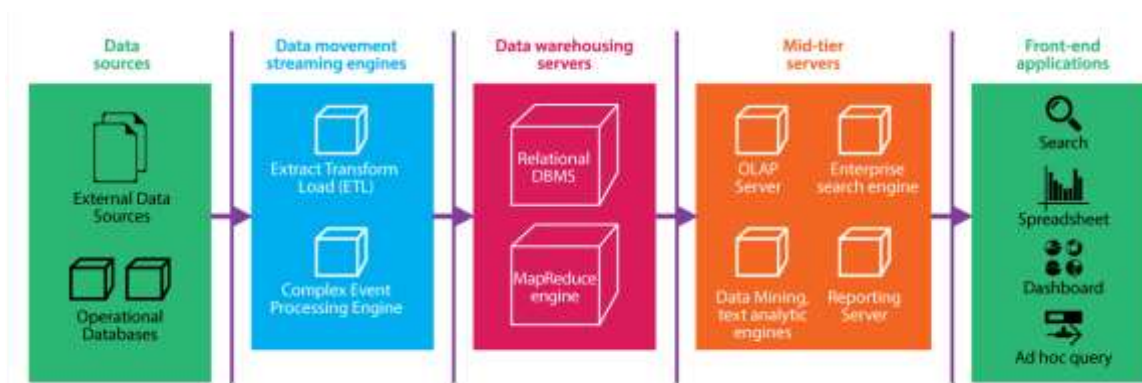
Pionir poslovne inteligencije je Hans Peter Lun, koji je bio IMB-om istraživač na polju kompjuterskih nauka. Prvi put spominjajući poslovnu inteligenciju 1958. godine i definišući je kao: “Sposobnost da se shvate relacije između prezentiranih podataka i pomaganja aktivnostima da dođu do određenih ciljeva.” Njegovi izumi našli su primenu u različitim oblastima kompjuterskih nauka, tekstilnoj industriji, lingvistici i informatici. Najveća dostignuća su mu algoritam Lihn, način indeksiranja ključnih reči u kontekstu KWIC (Key Words In Context) i idea za selektivno širenje informacija SDI (Selektiv Dissemanitoin of Information). Današnji SDI sistemi duguju Lunu koji ih je opisao kao: “Automatski metod za pružanje aktuelnih usluga o podizanju svesti naučnika i inženjera.” Od prvih koraka obrade podataka može se pratiti razvoj poslovne inteligencije do današnjih dana.

Enterprise systems (ES) imaju za cilj da podrže sve poslovne procese kompanije koristeći internet tehnologije. Oni integrišu informacije u okviru poslovanja, kao i sa spoljnim saradnicima, kupcima, dobavljačima i partnerima.



Slika 1 Business Intelligence

Glavne vrste sistema su Enterprise Resource Planning (ERP) koji se bavi unutrašnjom proizvodnjom, distribucijom i finansijskim procesima, zatim Customer Relationship Management (CRM) koji se bavi odnosima sa klijentima. Knowledge Management (KM) služi za prikupljanje i korišćenje podacima o kupcima. Sledeći je Supply Chain Management (SCM) koji je zadužen za marketing, prodajne procese i snabdevanje, zatim Supplier Relationship Management (SRM) koji služi za pronalaženje i upravljanje procesima kupovine, skladištenja robe i usluga. Druge vrste ES-a su Product Lifecycle Management (PLM), finansijsko upravljanje i upravljanje ljudskim resursima. Tu su i programi koji pomažu pri odlučivanju (EIS - Enterprise Information Systems i DSS - Decision Support System). Moderna poslovna inteligencija je prešla dug put u razvoju od svog osnivanja. Poslovna inteligencija (Business Intelligence - BI) predstavlja skup metodologija, analiza, koncepata i distribuciju informacija uz pomoć različitih softverskih alata. Ona predstavlja i sposobnost za prikupljanje, razumevanje, sortiranje i upravljanje podacima radi dobijanja prednosti nad konkurencijom. Što je od izuzetne važnosti za rad i rast preduzeća u savremenoj digitalnoj i globalnoj ekonomiji. Ranije donošenje odluka koje je bilo bazirano na intuiciji, promenjeno je razvojem tehnologije. Poslovna inteligencija je od ključnog značaja za vođenje kompanije kako bi se stekla prednost nad konkurencijom, uz pomoć upravljanja svojim podacima, informacijama i dobijenim znanjem. [1]



Slika 2 Tipična arhitektura poslovne inteligencije

Primena sistema poslovne inteligencije danas je u ogromnoj meri, što poboljšava performanse kompanije i unapređuje njene usluge, prodaju i produktivnost. Uz njeno korišćenje mogu se predvideti budući poslovni trendovi, pomoći korisnicima pri odlučivanju, održati kompetentnost na tržištu, kao i proširivanje tržišta. Uvođenje ovakvih sistema je dug process, koji zahteva dosta predhodne analize i investicija. Savremeni poslovni (BI) sistemi uključuju poslovne modele, skladištenja baza podataka, OLAP analize, BI platforme, aplikacije, ETL alate i alate za izveštavanje. Osnovna podela poslovne inteligencije u odnosu na izvore podataka su tržišna inteligencija (Market Intelligence) i unutrašnja inteligencija (Internal Intelligence).

2.2 Sistemi za upravljanje odnosa sa klijentima – CRM

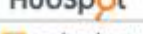
Danas trgovci putem interneta imaju na raspolaganju karakteristike koje su nekada imali klasični trgovci u prodavnicama. Počevši od uvida u kretanje robe do praćenja ponašanja klijenata odnosno potrošača i sa dobijenim podacima formiraju vlastite baze podataka. Koncept CRM-a je uspostavljanje, razvoj i interaktivno održavanje odnosa sa klijentima putem stalne veze i razmene informacija. To se ostvaruje brzo i lako uz današnju tehnologiju, brze veze, softvere i uređaje. [Lovreta] CRM se oslanja na ERP sisteme, integriše metodologije, strategije i softverski sistem, pri čemu pruža pomoć kompanijama da na organizovan način upravljaju odnosima sa klijentima i na taj način sakupljaju više informacija o njihovim potrebama i ponašanju. Na taj način dolazi do boljeg razumevanja, predviđanja i koncipiranja adekvatnog odgovora na ponašanje klijenata, pri čemu se stvaraju jače veze između njih. [2] CRM prikuplja informacije iz svih izvora unutar kompanije na jedno mesto što pruža pojednačnu sliku o svakom klijentu u realnom vremenu. Osnovni elementi CRM sistema su: operativni – služi za unošenje podataka u informacioni sistem preko aplikacije/softvera za praćenje klijenata, kolaborativni – služi za uspostavljanje kontakta sa klijentima putem različitih medija (od ličnog kontakta, putem telefona ili imejla i interneta) i analitički – kao najkompleksniji deo CRM sistema zadužen je za kreiranje profila o klijentu kao i o njegovim željama i potrebama kroz detaljne analize velikog broja podataka.



Slika 3 CRM sistem

Osnovne podkategorije CRM softvera su modifikovani softveri za male i srednje kompanije - CRM lite ili small business CRM, zatim softveri ili grupa alata za veće kompanije CRM Enterprise Suite i m-CRM za mobilne platforme. CRM softveri za mobilne platforme postaju sve popularniji kao i usluge preko ovih uređaja (slanje podataka, multimedijalnih sadržaja i dokumenata) jer su mobilni telefoni uvek uz korisnike sa bežičnim internetom, GPS-om (Globalni pozicioni sistem) i aplikacijama, što je veoma važno za današnje poslovanje. Ovakvi softveri su obično deo nekog već postojećeg softverskog paketa. Danas postoji velikih broj kompanija koje nude različite softvere za CRM, pa je zbog toga dosta teško kompanijama da izaberu odgovarajuće rešenje koje bi bilo najbolje za njih. Lideri iz ovih oblasti su Salesforce i Siebel (Oracle). Salesforce je više orijentisan ka manjim i srednjim preduzećima sa oko 30.000 klijenata i 650.000 korisnika, dok je Siebel (Oracle) za veće klijente i ima preko 5.000 klijenata sa 4.6 miliona korisnika. Postoje i mnoga druga softverska rešenja kao što su Microsoft Dynamics CRM, Surado CRM, Infor ERP XA CRM, Aplicor, Oncontact i SuperOffice. [Lovreta]

Takođe na tržištu se nalaze i besplatni (open-source) softverski paketi za upravljanje odnosa sa klijentima. Najpoznatiji open-source paketi su: SuiteCRM, CapsuleCRM, Insightly, FreeCRM, i SugarCRM. Postoji mogućnost i rentiranja softverskog paketa putem interneta. Ključne razlike između kupljenog softvera i open-source rešenja su bezbednost podataka i brzina rada zbog udaljenih servera. Softveri za odnos sa klijentima treba da poseduju lak korisnički interfejs (UI) na globalnom nivou.

Poređenje besplatnih CRM rešenja					
Kompanija	Besplatni kontakti	Besplatni korisnici	Besplatno skladištenje	Besplatna podrška	Cena nadogradnje
 Capsule	250	2	10 Mb	—	\$12/user/month
 insightly	2500	2	200 Mb	Email	\$12/user/month
 really simple systems	Neograničeno	2	100 Mb	Full Service	\$15/user/month
 FreeCRM	100000	100	—	—	\$24 ⁹⁰ /user/month
 ZOH CRM	Neograničeno	10	Neograničeno	Full Service	\$12/user/month
 Zurmo	Neograničeno	Neograničeno	Neograničeno	Online	\$32/user/month
 vtiger	Neograničeno	Neograničeno	Neograničeno	Online	\$12/user/month
 Bitrix24 ⁹	Neograničeno	12	5 Gb	—	\$25/user/month for 12 users
 raynet ⁹	150	2	50 Mb	Full Service	\$19/user/month
 suiteCRM	Neograničeno	Neograničeno	Neograničeno	Online	\$16 ⁴⁰ /user/month
 HubSpot	Neograničeno	Neograničeno	Neograničeno	Full Service	\$10/user/month
 salesbox	Neograničeno	2	Neograničeno	Full Service	\$19/user/month

Slika 4 Poređenje besplatnih CRM sistema

2.3 Sistemi za planiranje poslovnih resursa – ERP

Većina softverskih paketa je šezdesetih godina omogućavala samo kontrolu zaliha. Kasnije (70-tih godina) su uvedene MRP (*Material Requirements Planning*) rešenja koja su korišćena za raspoređivanje proizvodnje. Kasnije su ovi sistemi unapređeni i poboljšani uvođenjem novih alata za planiranje prodaje, obradu porudžbina i planiranje kapaciteta. Osamdesetih godina ovi sistemi su još više uznapredovali i stvaraju novu/dругu generaciju ovakvih sistema - MRP II. Druga generacija dodaje sisteme za vođenje računovodstva i finansija. Devedesetih godina na tržište izlaze prvi ERP sistemi koji integrišu sve informacije i tokove informacija jedno kompanije.

ERP poslovni informacioni sistemi su modularno orijentisani procesi u kojima se planiraju poslovni resursi i koji podržavaju sve poslovne funkcije sa mogućnostima prilagođavanja konkretnim uslovima kompanije. ERP strateški alati omogućavaju brz rad i pristup sa kupcima, dobavljačima i poslovnim partnerima na globalnom tržištu putem interneta. ERP integralni poslovni sistemi pomažu planiranju i izvršavanju poslovnih procesa. Ovaj sistem integriše sve poslovne procese i informacije u jednu celinu, čime omogućava bolji i brži uvid u sve segmente poslovanja i efikasno praćenje na svim nivoima, čime smanjuje troškove kompanije. Brz odziv na kompleksne zahteve klijenata i tržišta, čime raspolaže ovakav sistem, pojednostavljuje ulaz na regionalno i svetsko tržište. Veliki broj analiza i izveštaja sa brzim i jednostavnim pristupom informacija omogućava lakše donošenje operativnih i poslovnih odluka. [2] Koristi od ERP sistema mogu se videti u raznim delovima poslovanja, kao što su distribucija, računovodstvo, proizvodnja i usluga dobara.



Slika 5 ERP sistem

ERP sistemi su grupisani po sledećim oblastima i modulima poslovanja:

- metode i tehnike preporučivanja sadržaja i primena mašinskog učenja u sistemima za preporučivanje proizvoda
- upravljanje finansijama
- upravljanje prodajom i marketingom
- upravljanje nabavkom
- upravljanje magicinskim poslovanjem
- upravljanje proizvodnjom
- planiranje resursima
- upravljanje servisom i održavanjem
- kadrovska evidencija

Najčešći moduli jednog ERP paketa su: upravljanje finansijama, lanci snabdevanja (SCM), odnosi sa klijentima (CRM), elektronsko poslovanje (e-Business), upravljanje uslugama (SM), distribucija, alati za optimizaciju poslovanja, prodaja i marketing.

U izvršavanju procesa uključen je jedan on poslovnih informacionih sistema, odnosno komercijalni softver za velike kompanije. ERP softverski paketi čine jednu od najbrže rastućih tržišta u softverskoj branši. Softverski paket bi trebao da imitira poslovne procese u samoj kompaniji. U slučaju da se to ne dešava, trebalo bi promeniti poslovne procese ili postojeći softver. Ni jedno od ova dva rešenja nije lako postići. Menjajući poslovne procese kompanije i ljudi se moraju prilagođavati novim i prihvatiti ih u potpunosti. U protivnom softver neće postići željene rezultate. Dok menjanje softvera zahteva nadogradnju i ponovni rad na njegovim podešavanjima. Takvi softveri/programi imaju mogućnost prilagođavanja, prema konkretnim zahtevima i potrebama kompanije. Najpoznatiji paketi su: ORACLE – people soft, Pulls, BAAN, MDS – nav, AX, GR i SI. Sigurno najpoznatiji sistem R/3 (baziran na klijent/server tehnologiji) predstavio je SAP (*Systems, Applications, Products*), jedna od najpoznatijih softverskih kompanija iz Nemačke. Takođe postoje i besplatna (*open-source*) rešenja kao što su Apache OFBiz, xTuple, Odoo, ErpNext i drugi.

2.4 Implementacija ERP rešenja

Danas većina organizacija i kompanija razmatra mogućnost implementacije jednog od ERP rešenja, što nije nimalo lak zadatak. Kompleksnost uvođenja jednog ovakvog sistema zavisi od veličine same kompanije, njenom broju procesa, odeljenja koja su na raspolaganju i potrebnih modula. Zato se ceo proces implementacije deli u više jednostavnih procesa.

Prvi zadatak kompanije je određivanje jasnih ciljeva i specifičnih zahteva, kako bi se dobila jasnija slika o potrebnom sistemu. Uvođenje ovakvih sistema zahteva detaljnu analizu svih poslovnih procesa, uvođenje novih radnih procedura i obuku zaposlenih. Zato postoje tri strategije implementacije ERP sistema u kompaniju, a to su postupna (step by step), potpuna (big bang) i implementacija po modelu (roll out).

Faze uvođenja ERP-a su:

- izbor paketa i modula
- tehnička instalacija
- prilagođavanje
- kalibracija
- uvođenje u red
- održavanje

Na osnovu RFP-a (Request For Proposal) i razumevanju uslova prodavac predlaže određeni vid implementacije. Ovakav predlog treba da sadrži hardverske i softverske zahteve, module i broj radne snage koja je potrebna za uvođenje ERP rešenja. Postoji velika verovatnoća za potrebom novih hardverskih i softverskih komponenti radi normalnog funkcionisanja budućeg ERP sistema. Najčešće su to serveri i rashladni sistemi. Kako bi se izbegli dodatni troškovi za nabavku nove infrastrukture, može se preći na Cloud rešenja. Ovom vrstom rešenja kompanija radi na servisima koja su OnLine i čine infrastrukturu. Potrebno je samo izvršiti pretplatu za ovakvu vrstu usluge.

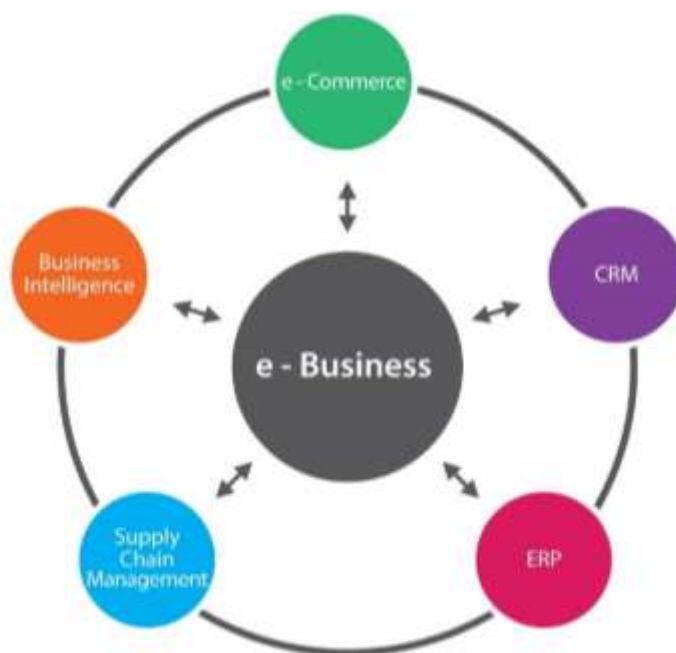
U početnim fazama implementacije poželjno je angažovanje konsultanta koji bi obučavao IT sektor i drugo osoblje. Na kraju instalacije rešenja potrebno je obaviti testiranje pre nego što se sistem stavi u praktičnu upotrebu i pokrene.

Svako uvođenje ovakvih sistema je popriličan izazov jer sadrži veliki broj specifičnih zahteva koji zavise od same kompanije i njenih procesa.

3 Opis karakteristika elektronskog poslovanja i elektronske trgovine

3.1 Elektronsko poslovanje

Elektronsko poslovanje (e-Business) obuhvata sve oblike procesa transakcija i razmena informacija koje za svoj rad i obavljanje funkcija koriste informacione i komunikacione tehnologije - najčešće internet. Osnovu elektronskog poslovanja čine ERP sistemi obezbeđujući funkcije *front-office*-a, čime se omogućava klijentima pristup praćenja svojih porudžbina putem interneta. Elektronsko poslovanje se tačnije definiše od elektronske trgovine (e-Commerce) jer uključuje niz akcija. Od prodaje i kupovine proizvoda i usluga, preko pružanja servisa i podrške klijentima / poslovnim partnerima pa do korišćenja elektronskog učenja (e - Learning) i transakcija u okviru kompanije.



Slika 6 Elektronsko poslovanje

Termin elektronsko poslovanje ili e-biznis prvi put spominje direktor IMB-a Lou Gerstner 1997. IMB definiše elektronsko poslovanje na sledeći način: „Proces korišćenja Web tehnologije u službi pomaganja procesa poslovanja, sa ciljem da se poveća produktivnost i rast. Što omogućava kompanijama lakšu komunikaciju sa klijentima, partnerima, prodavcima i potrošačima.“ [3] Elektronsko poslovanje je deo internet ekonomije, koji sadrži sve aktivnosti kompanije preko interneta.

3.2 Elektronska trgovina

Elektronska trgovina (e-commerce) je sveopšte poslovanje putem kompjuterskih mreža, kao što je internet. Za njeno obavljanje ona koristi komunikacione tehnologije (pre svega internet), ekstranet, elektronsku poštu, elektronske knjige, baze podataka, mobilne telefone i tablete. Takođe sadrži sve što sadrži i klasičan vid trgovine, od distribucije, kupovine, prodaje, servisiranja i ostalih usluga. Ova vrsta trgovine uključuje elektronski transfer novca, upravljanje lancem snabdevanja, elektronski marketing, elektronsku razmenu i sakupljanje podataka. Elektronska trgovina je značajan napredak u poslovanju, jer omogućava poslovanje bez geografskih, vremenskih i nacionalnih ograničenja. Glavne osobine elektronske trgovine su sveprisusnost, globalna dostupnost, univerzalni standardi, bogastvo, interakcija, gustina informacija i prilagođavanje zahtevima odnosno personalizacija.

Vrste elektronske trgovine: [3]

- B2B - kompanija ka kompaniji (business to business)
- B2C - kompanija ka potrošaču (business to consumer)
- B2E - kompanija ka zaposlenima (business to employee)
- C2B - potrošač prema kompaniji (consumer to business)
- C2C - potrošač ka potrošaču (consumer to consumer)
- B2G - kompanija ka vladi / ministarstvu - javni sektor (business to government)
- P2P - pojedinac prema pojedincu (peer to peer)
- m-commerce - mobilna trgovina (mobile commerce)

B2B trgovina je tip elektronske trgovine u kojoj kompanije trguju između sebe, odnosno sa drugim kompanijama. Ovakav vid e-trgovine predstavlja njenu najveću formu.

B2C trgovina predstavlja najčešći oblik e-trgovine u kome kompanije putem interneta prodaju svoje proizvode individualnim kupcima. Ova vrsta e-trgovine ima konstantan porast od svog nastanka. B2C trgovina poseduje više podkategorija: portali, internet prodavnice, kreatori tržišta, transakcioni brokeri, provajderi usluga i provajderi sadržaja.

B2E trgovina koristeći intra-biznis mrežu omogućava kompanijama da obezbedi proizvode i usluge svojim zaposlenima. To uključuje korporativna saopštenja, specijalne ponude zaposlenima i snabdevanje resursima.

C2B trgovina je oblik e-trgovine gde kranji potrošač ili korisnik, pravi proizvod ili uslugu koju kompanija koristi sa ciljem da se završi poslovni proces i stekne konkurentsku prednost. Jedan od primera za ovakav vid poslovanja je sajt za zapošljavanje UpWork (bivši Elance), takođe Google AdWords i Amazon Affiliates.

C2C trgovina predstavlja oblik e-trgovine u kome potrošač prodaje proizvod drugom potrošaču. On proizvod postavlja na jedan od sajtova za takvu vrstu prodaje ili aukcije, sa svim obezbeđenim alatima koja ovakva vrsta sajta (market mejer) može posedovati. Od pretrage, postavke, do kompletne transakcije (naplate i dostave) proizvoda. Najpoznatiji su eBay, kupujem-prodajem, limundo, kupindo i drugi.

B2G trgovina je izvedena od B2B trgovine i predstavlja usluge na različitim nivoima vlasti (savezne, državne i lokalne) preko integrisanih marketing tehnika komunikacije.

P2P trgovina odnosno tehnologija omogućava korisnicima razmenu fajlova i drugih resursa direktno bez posrednika ili servera putem interneta. Primeri su Gnutella, Napster, BitTorrent, eMule, BearShare, FireWire i mnogi drugi. Za korišćenje fajlova na nivou lokalne mreže koristi se IMB SMB protokol, koji je od 2000. uvršten na Majkrosoftovim operativnim sistemima.

Mobilna trgovina koristi bežičnu (wireless) tehnologiju za obavljanje transakcija na internetu preko bežičnih digitalnih uređaja kao što su mobilni telefoni, tablet ili Palm uređaji.

4 Skladištenje i analiza podataka

4.1 Skladištenje i analiza podataka

Svaka kompanija sakuplja podatke o svom poslovanju, kako bi ih kasnije obrađivala i analizirala. Svrha obrade i analize podataka je dobijanje informacija neophodnih za sticanje potrebnih znanja. Najčešće se analiziraju podaci iz baze podataka (Databases) koju možemo definisati kao skup povezanih informacija, koje se organizovano čuvaju tako da se mogu brzo izabrati i preuzeti za korišćenje. [7] Baze podataka su odlično razvijena tehnologija. Pouzdane su i efektivne, čime su povoljne za operativnu podršku odlučivanju. One su razvijane kako bi se upravljalo podacima. Osim baza podataka podaci mogu biti i tekstualnog tipa, podaci sa interneta odnosno web-a ili podaci koji su organizovani u određenim vremenskim serijama. SQL baze transformišu podatke u informacije.

U savremenim poslovnim sistemima postoji potreba za korišćenjem podataka iz različitih izvora. Čest je slučaj da se bitni podaci ne nalaze na jednom mestu ili u jednoj bazi podataka, pa čak i da nisu deo samo jednog informacionog sistema. Prebacivanje podataka sa sekundarnih sistema na glavni može da prouzrokuje niz problema oko nestabilnosti, nepouzdanosti pa čak i do pada sistema. Kako se to ne bi desilo pribegava se metodi skladištenja podataka. [12] Skladištenje podataka (Data Warehousing) je jedna od metoda poslovne inteligencije kojom se analizira i obrađuje velika količina podataka za potporu odlučivanju i upravljanju kompanije. Skladištenje podataka na ovaj način omogućava osnov za analizu, lak pristup podacima, mogućnost menjanja podataka i bezbednost podataka. Razvojem skladištenja podataka došlo se i do *Real Time Data Warehousing-a* koji osvežava podatke čim dođe do promena, čime se zadovoljavaju trenutni zahtevi korisnika.

Za stvaraoce skladištenja podataka smatra se Viliam Bil koji je 1991. godine napravio priručnik za upotrebu, i ako je koncept postojao još ranije. Karakteristike skladištenja podataka su organizacija, konzistentnost, vremenski podaci, multidimenzionalnost i web zasnovanost. Sa razvojem informatičkih tehnologija, razvijane su i tehničke mogućnosti za skladištenje, obradu i analizu velikog broja baza podataka. Koje su bivale svakim danom sve veće i prevazilazile čovekove mogućnosti da ih tradicionalno obrađuje i analizira. To je dovelo do potrebe za novim i drugačijim, automatskim sredstvima za analizu podataka. Uporedno u drugim oblastima istraživanja (bazama podataka, statistici,

mašinskom učenju, veštačkoj inteligenciji i drugima) razvijaju se procesi za automatizovane analize podataka u okviru složenih skupova. Proces u oblasti statistike u kojima se sve to dostiže naziva se "Data Mining" ili rudarenje podataka.

4.2 Rudarenje podataka - Data Mining

Data Mining označava automatizovani analitički proces podrške odlučivanju, otkrivajući obrasce i trendove u njihovim informacijama. Tehnike i alati koji otkrivaju potrebna znanja u bazama podataka nazivaju se Knowledge Discovery in Databases (KDD). KDD i Data Mining su termini koje ne treba mešati, jer KDD se bavim pripremom podataka, izborom obrazaca i procenjivanjem znanja u interaktivnom i iterativnom procesu otkrivanja korisnog znanja iz podataka, dok se Data Mining bavi aplikacijama algoritama za izvlačenje obrazaca iz podataka. [2] Data Mining aplikacije su od pomoći na najrazličitije vrste i opsege biznisa i imaju široku primenu u marketingu, bankarstvu, osiguranju, auto industriji, zdravstvenim ustanovama, radio difuziji, maloprodaji i prodaji, proizvodnji i produkciji, policiji, unutrašnjoj bezbednosti, ministarstvu odbrane i računarskim hardverima i softverima. [5] Data Mining-a služi za otkrivanje određenih šablona i veza, čime bi se pojasnile određene pojave koje kompanija može da iskoristi za unapređivanje svog poslovanja.

Ceo proces pronalaženja i tumačenja obrazaca koji su dobijeni iz podataka prolazi kroz više koraka:

- *razumevanje i prepoznavanje* problemske situacije, postavljanje ciljeva i razumevanje mogućih rešenja problema
- *selekcija i odabir* podataka koji su nam potrebni za analizu
- *čišćenje podataka* i predprocesiranje nepotrebnih podataka radi brže analize
- *izbor metoda Data Mining-a* koje će se koristiti za traženje najrelevantnijih obrazaca u podacima odnosno odabir samog modela i parametara
- *tumačenje i vrednovanje* obrazaca radi korišćenja za procese podrške odlučivanju
- *sređivanje* svih otkrivenih podataka i znanja iz predhodnih faza

4.2.1 Tehnike i algoritmi Data Mining-a: [2]

Tehnike Data Mining-a su uvođenje algoritama radi pronalaženja obrazaca u velikim bazama podataka. U Analysis Services SQL Server-u 2005 integrisani su modeli Data Mining-a. [1]

Decision Trees (stablo odlučivanja) je jedan od najpopularnijih metoda za predpostavke i klasifikaciju.

Association Rules (pravila asocijacije) je algoritam koji grupiše artikule i grane artikala po sličnosti i pronalazi relacije između različitih elemenata. Dobro radi i na ogromnom broju artikala.

Clustering (klasering) je tehnika koja služi za grupisanje podataka koji su slični u prethodnim događajima, a može i da grupiše klijente.

Classification (klasifikacija) služi za otkrivanje određene grupe klijenata koji su korisnici ili nisu korisnici određenih vrsta usluga. To se vrši posmatranjem ponašanja klijenata čime se smanjuju troškovi promocija.

Sequential Pattern Discovery (sekvencionalno otkrivanje obrazaca) je tehnika najbližnja tehnici asocijacije, sa tim da ona sekvencionalno vezuje događaje vremenski. Takođe određuje povezanost elemenata tokom određenog vremenskog perioda.

Naive Bayes je algoritam koji se zasniva na Bajesovoj teoremi o nezavisnosti promenljivih između različitih elemenata. Pomoću njega se izračunava verovatnoća između ulaza i predvidljivih kolona.

Neural Nets (neuronske mreže) kao deo mašinskog učenja (Machine Learning) imitiraju neurone u ljudskom mozgu i simuliraju učenje na osnovu datog iskustva.

Forecasting (predviđanje) je tehnika koje se obavlja iz dva dela. Prvi deo je regresiona analiza (regression analysis) koja koristi vrednost podataka iz prošlosti radi predviđanja budućih događaja i vrednosti. Dok druga analiza, slična prethodnoj je otkrivanje vremenske sekvence (time sequence discovery) i razlikuje se po tome što predviđanja vrši za jedan određeni vremenski period.

4.2.2 Text Mining i Web Mining

Text Mining pregledava neizdefinisane podatke tekstualnog tipa koje kompanije mogu da analiziraju. One mogu biti i komentari koje klijenti ostavljaju radi poboljšanja usluga i proizvoda.

Otkrivanje i izdvajanje određenog potrebnog sadržaja preko interneta postaje sve teži i duži posao zbog konstantnog rasta broja sajtova, dokumenata i fajlova drugog tipa. Zato se koriste metode Data Mining-a i KDD-a da bi se izdvojili potrebni podaci iz Web dokumenata.

4.2.3 Poslovna analitika

Poslovna analitika (Business analytics - BA) je skup znanja, veština, tehnologije i iskustva odnosno prakse u istraživanju iz prošlosti poslovanja. Ona se fokusira na razvoj novih saznanja i razumevanje poslovanja na osnovu podataka i statističkih metoda.

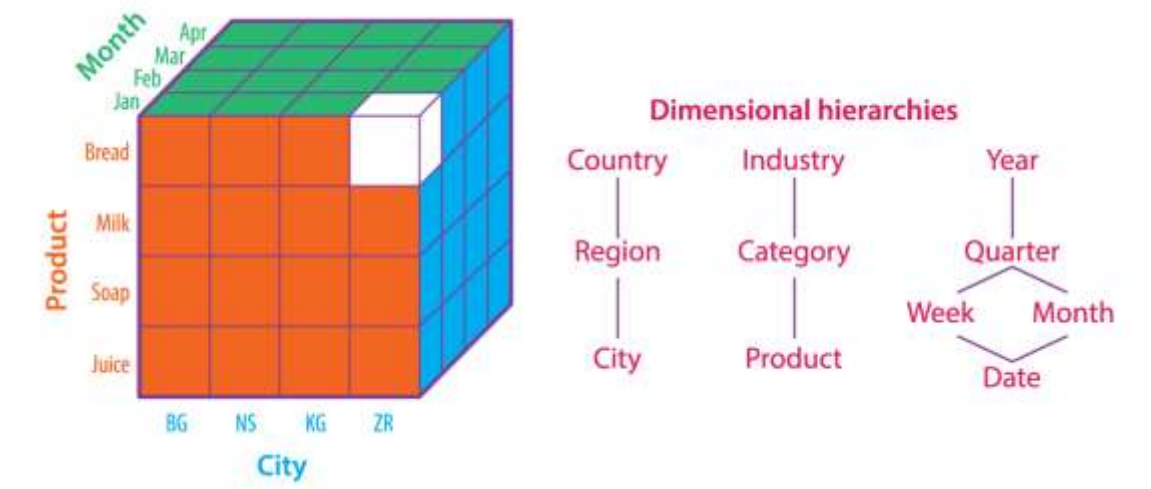
Poslovna analitika predviđa i optimizuje poslovne procese i trendove. [7]

5 OLAP sistemi

5.1 OLAP sistemi

OLAP (On Line Analytical Processing) sistemi su softverski sistemi za analitičko procesuiranje velikog i kompleksnog broja podataka za kratak vremenski period. Čime se postiže sažet i interaktivan uvid u obrađene informacije sa mogućnostima širokog spektra pogleda na njih, pa te informacije koriste krajni korisnici (analitičari i menadžeri). [8] Za razliku od OLTP (On Line Transaction Processing) sistema, koji služi za ažuriranje baza podataka i u procesima transakcija, OLAP sistemi su namenjeni analizama velikim brojem podataka. OLAP sistemi su osmišljeni kada OLTP sistemi više nisu mogli da daju odgovore u kratkim i odgovarajućim vremenskim intervalima. OLAP sistemi su znatno brži sistemi koji podržavaju kompleksne analize. Oni omogućavaju analizu podataka iz različitih perspektiva prema krajnim korisnicima, odnosno analitičarima. Osnovni elementi ovih sistema su OLAP serveri, baze podataka, korisnički interfejs i alati za administriranje. OLAP baza podataka sadrži numeričke mere, dimenzije

i kocke. [2] U OLAP dimenzije spadaju dimenzije, nivoi i članovi. OLAP kocke predstavljaju strukturu skladištenja OLAP baza podataka, kombinujući dimenzije i mere.



Slika 7 OLAP kocka

OLAP baze podataka omogućavaju fleksibilne modele pronalaženja podataka. Fleskibilnost omogućavaju dimenzije i hijerarhije. Kod dimenzija *Slice* i *Dice*, a kod hijerarhije su to *Drill Down* i *Drill Up*.

Slice ili kriške (podskup kocke) čine izbor člana iz dimenzije.

Dice ili domine su postavka više članova iz dimenzije na osu, kao i više članova iz druge dimenzije na drugu osu.

Drill Down omogućava fokusiranje samo na određen podatak ili na određen deo uz pomoć više pregleda nivoa hijerarhije po jednom članu.

Drill Up služi za sagledavanje opšte slike jer poznaje samo zbirne informacije članova.

OLAP pristup mora da ima poseban računar koje se ponaša kao server, kao i zaseban interfejs za komforan rad i izvođenje analitičkih operacija i pregleda. OLAP analiza poseduje definisanje izvora podataka, definisanje strukture kocke, podešavanje dimenzija i mera i izvršavanje upita nad kockom.

5.2 Arhitektura OLAP sistema

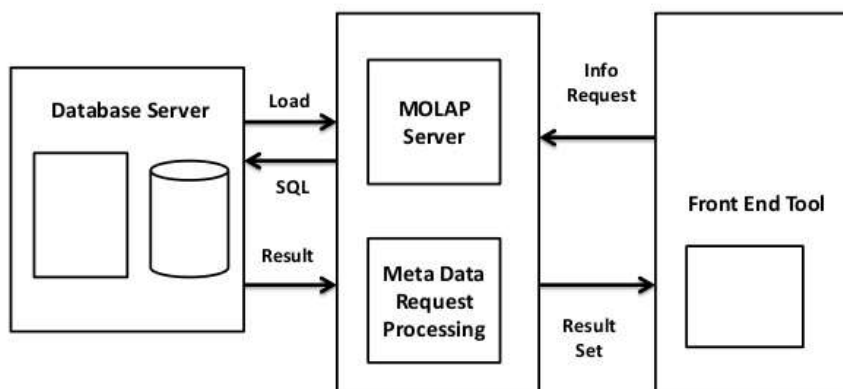
Osnove arhitekture OLAP-a su MOLAP (Multidimensional On Line Analytical Processing) i ROLAP (Relational On Line Analytical Processing). Akronim HOLAP (hybrid OLAP-a) je tehnologija koja kombinuje obe prethodne arhitekture. Najbitnije razlike su u načinu skladištenja podataka i pristupu podacima.

Svaka arhitektura OLAP-a ima svoje dobre i loše strane, što dovodi u razmišljanje koju treba koristiti u datim trenucima. To se rešava na osnovu kriterijuma kao što su performanske, veličina podataka, rad sa obimnim i teškim računima i agregacijama. ROLAP sistemi su prvenstveno namenjeni za pristupanje podacima, za razliku od MOLAP sistema koji zluže za prikupljanje podataka. Prvi proizvod koje je posedovao funkcije OLAP i pre pojave naziva OLAP je MDS's Express iz 1970. godine. Trenutno najpoznatiji proizvodi ove vrste su Business Object, Brio, IBM DB2 Olap Server, SAP BW, Cognos, Microsoft Analysis Services i MicroStrategy.

5.2.1 Multidimenzionalni OLAP (MOLAP)

Koristeći online analitičke obrade i multidimenzionalne baze podataka (MDD) direktno skladišti podatke u multidimenzionalne "kocke" što ga čini klasičnim sistemom OLAP-a. Zahteva prethodnu obradu podataka pre smeštanja u OLAP kocke. Obezbeđuje visoke performanse celog sistema i mogućnost obavljanja zahtevnih izračunavanja, što su najveće prednosti ovog sistema. Odgovori su brzi zbog unapred urađene konsolidacije kao i prikazivanje podataka, naročito prema upitima "Dicing" i "Data Slicing". U ROLAP-u je sve unapred izračunato kada su kocke kreirane, pa se brzo mogu povratiti. Glavni nedostaci ROLAP sistema su ograničena količina podataka, neomogućavanje detaljnog upita "Drill Down", obrada može dugo da traje u nekim koracima gde je veliki obim podataka.

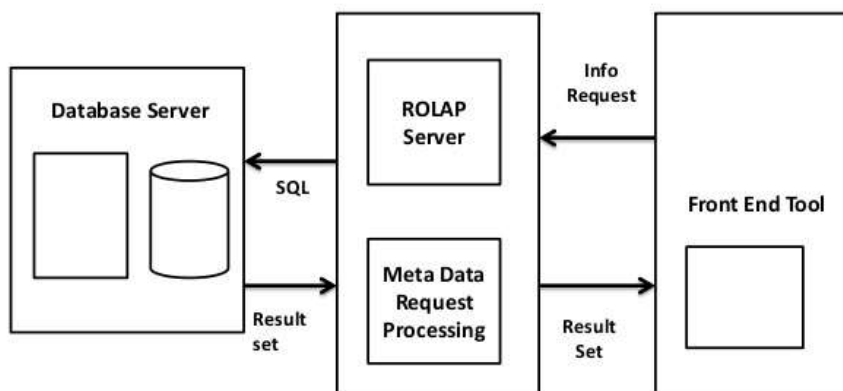
Proizvodi bazirani na MOLAP-u: Cognos Powerplay, Oracle Database OLAP Option, Microsoft Analysis Services, Essbase, TM1, Lilith Hicare, Daptech Keystone, Palo (open source MOLAP server).



Slika 8 MOLAP arhitektura

5.2.2 Relacioni OLAP (ROLAP)

ROLAP se koristi relacionim bazama podataka i radi sa velikom (neograničenom) količinom podataka. Takođe koristi SQL upite i hardverske mogućnosti kao što je paralelno procesuiranje. Kod ROLAP-a nema potrebe za prethodnu obradu podataka i dodatno memorisanje kao kod MOLAP-a. Sve to dovodi do dobrih performansi višekorisničkog rada i dobre zaštite. Postoji mogućnost da implementacija Data Warehouse-a bude završena prema standardnom ili prema proširenim sistemima za upravljanje bazom podataka. ROLAP se koristi u slučaju sa dosta atributa pa ih je zbog toga nemoguće staviti u kocke. Služe za širu sliku od MOLAP-a, jer sadrži tekstualne i numeričke podatke. Nedostaci su mu: sporija obrada u odnosu na ostale varijante, Najpoznatiji softveri bazirani na ROLAP-u su: Microsoft Analysis Services, MicroStrategy, Oracle Business Intelligence Suite Enterprise Edition (ranije Siebel Analytics), Tableau Software, Mondrian (besplatni ROLAP server).



Slika 9 ROLAP arhitektura

5.2.3 Hibridni OLAP (HOLAP)

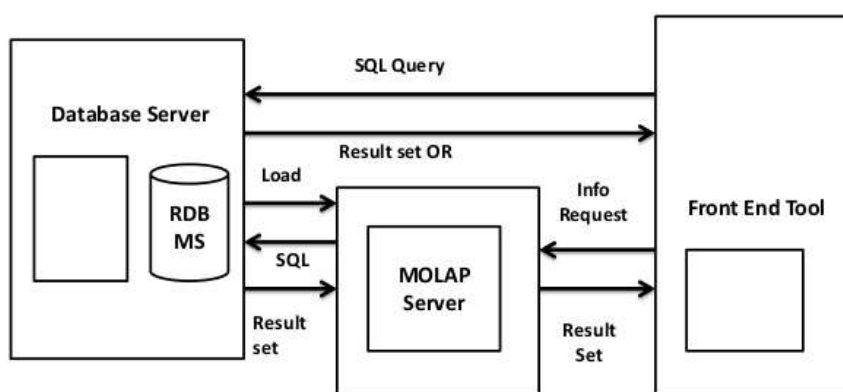
HOLAP kombinuje obe tehnike tako što uzima najbolje karakteristike iz njih, što ujedno predstavlja i najbolje rešenje. HOLAP koriste OLAP kocke, ali i "Drill Down" upite za detaljnije informacije. HOLAP tehnika uzima najbolje karakteristike iz obe OLAP tehnologije (bržu obradu i veću količinu podataka), a jedan deo podataka može da smešta u ROLAP, a drugi u MOLAP. Takođe podržava horizontalno i vertikalno particioniranje.

Vertikalno particioniranje sve agregacije memoriše u MOLAP kako bi poboljšao performanse, dok podaci ostaju u ROLAP-u zbog optimizacije.

Horizontalno particioniranje - U ovom modu HOLAP memoriše u MOLAP tehnologiji sloj podataka dobijen slicing operacijom, obično najnoviji (tj. dobijen operacijom slicing u funkciji vremena – podaci u datom trenutku) da bi upiti brže radili, a ostale podatke koristeći ROLAP, čime uvažava činjenicu da u velikoj OLAP kocki može biti "gustih" (sa mnogo podataka) i "retkih" regiona (oskudnih u podacima).

Hibridna online analitička obrada je u stvari ROLAP koji može da radi sa višedimenzionalnim bazama. Najpoznatiji HOLAP alati su Oracle Express, Holos proizvod kompanije Crystal Decisions (prvi proizvod koji je koristio HOLAP memoriju), Microsoft Analysis Services, MicroStrategy, SAP AG BI Accelerator, Microsoft SQL Server 7.0 OLAP Services (podržava hibridni OLAP server).

HOLAP nije jedini akronim, postoje još i deskop OLAP (DOLAP) i Web OLAP (WOLAP) koji se takođe koriste, ali nisu toliko popularni.



Slika 10 HOLAP arhitektura

6 Opis metoda i tehnika preporučivanja sadržaja na sajtu "Amazon"

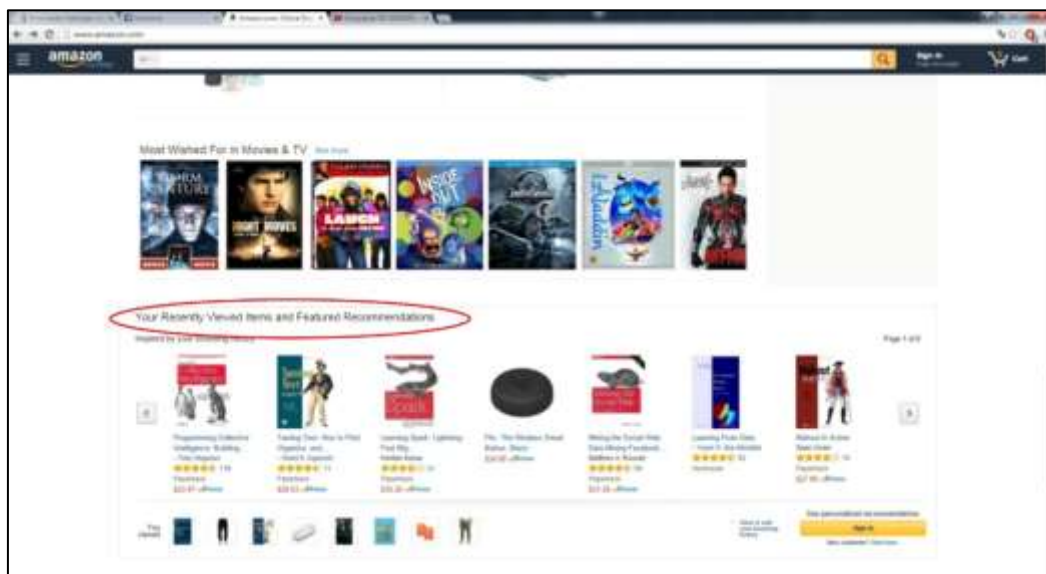
Algoritmi za preporučivanje proizvoda su najpoznatiji za korišćenje u elektronskoj trgovini. Koristeći ulaz o interesima potrošača za generisanje liste preporučenih predmeta. Mnoge aplikacije koriste samo stavke koji potrošači kupuju, ali se mogu koristiti i drugi atributi kao što su pregledani i omiljeni sadržaji, demografski podaci, interesovanje i drugi. Kada Amazon preporučuje proizvod na svom sajtu to nije nikakva slučajnost. [9] Njihov sistem preporučivanja zasniva se na velikom broju jednostavnih elemenata, od toga šta je potrošač kupio u prošlosti i koje je sadržaje pregledao i ocenio pa do stvari koje je video i kupio. Koristeći algoritme za preporučivanje proizvoda kojim personalizuje svakog potrošača posebno i menjajući preporučivanje na osnovu njegovih interesovanja, pokazujući programske knjige i literaturu softverskom inženjeru i bebi opremu novim majkama. [7] Postoje standardne tehnike za rešavanje problema u preporučivanju, a to su kolaborativno filtriranje (collaborative filtering), filtriranje po sadržaju (content based filtering), metode istraživanja znanja (Data Mining) i kontekstualne metode (context aware). [4]

Amazon koristi "item to item" algoritam za zajedničko filtriranje kojim preporučuje proizvode potrošačima koji se vraćaju na sajt. Upoređujući standardne tehnike sa "item to item" algoritmom, dolazimo do zaključka da Amazon algoritam nezavisno i on-line pravi preporuke u realnom vremenu, odabir odgovarajućih podataka iz baza podataka i generiše visok kvalitet preporuke. Posle uvođenja ovog algoritma Amazon ima porast prodaje od 30% do 35% samo od tehnika preporučivanja koje koristi. Pre počinjanja korišćenja "item to item" algoritma, Amazon je koristio BookMatcher algoritam.

6.1 Struktura sajta "Amazon"

Amazonov sistem preporučivanja proizvoda i praćenja ponašanja korisnika je kompleksan jer istovremeno koristi više metoda:

- *Your Recently Viewed and Featured Recommendations* - nedavno pregledani proizvodi i izdvojene preporuke



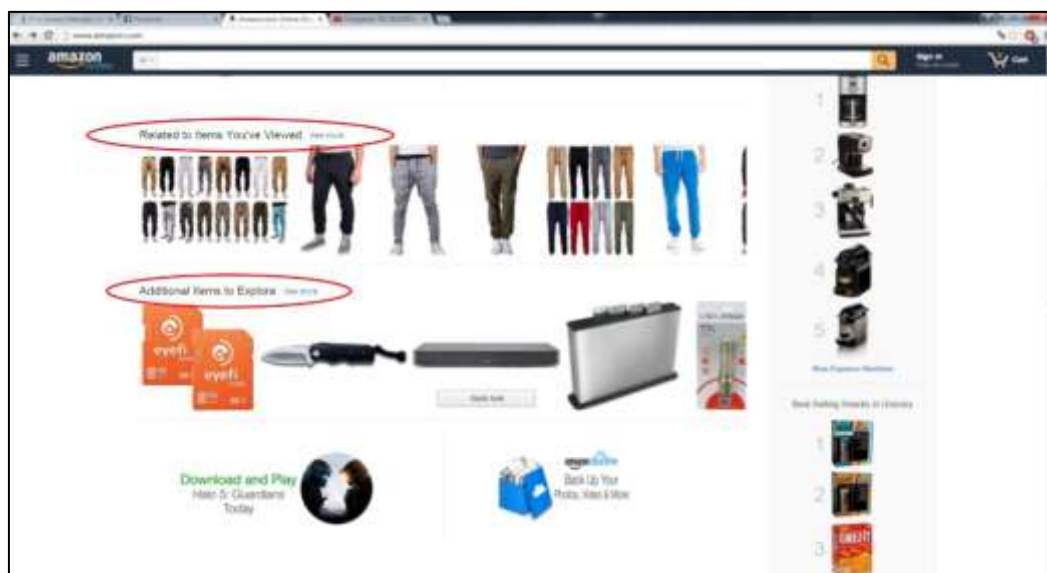
Slika 11 Amazon struktura sajta 1

- *More items to Consider* - pogledaj još proizvoda
- *Inspired by Your Browsing History* - preporučeni proizvodi prema istoriji pretraživača



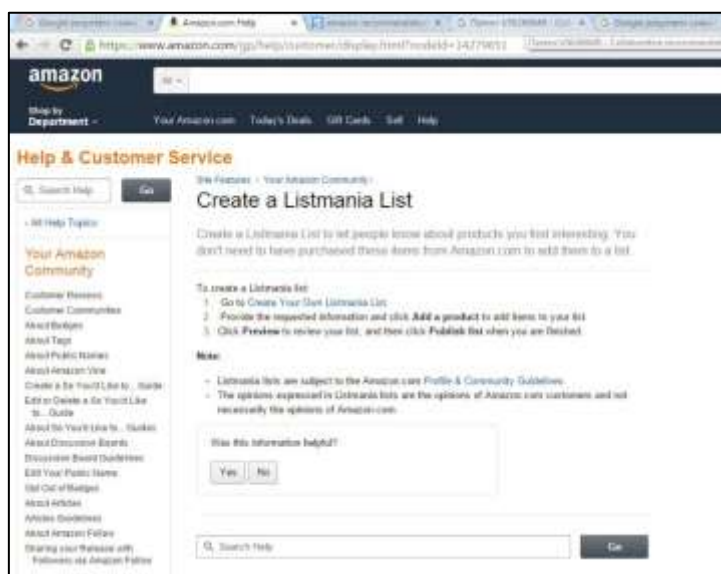
Slika 12 Amazon struktura sajta 2

- *Related To Items You've Viewed* - slični proizvodi od viđenih
- *Additional Items To Explore* - dodatni proizvodi za gledanje



Slika 13 Amazon struktura sajta 3

- *Listmania* - potrošači kreiraju svoje liste omiljenih proizvoda



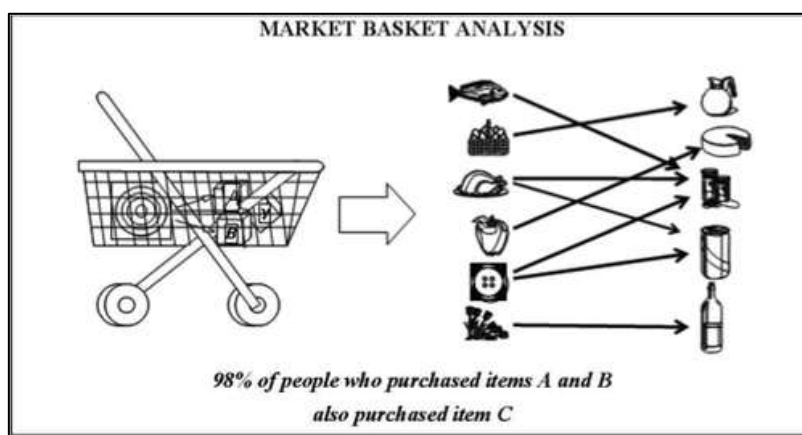
Slika 14 Amazon Listmania

- *People Who Bought This* (kupci koji su kupili ovaj proizvod) - međusobno poređenje kupaca
- *Alerts* (Upozorenja) - sistem šalje poruku (e-mail) kada se prodaje proizvod koji se dopada kupcu
- *Customer Reviews* (mišljenja posetilaca) - korisnici ocenjuju proizvode (od 1 do 5) i daju svoja mišljenja u vidu komentara.

7 Market Basket analiza

7.1 Uvod u Market Basket analizu

Market Basket analiza je tehnika modeliranja zasnovana na teoriji da u slučaju kupovine određene grupe ili pojedinačnog proizvoda od strane potrošača postoji manja ili veća verovatnoća da će potrošač kupiti još jednu grupu ili pojedinačan proizvod. Drugačije rečeno analiza potrošačke korpe pronalazi odnose između kupovine određenih proizvoda.[1]



Slika 15 Market Basket analiza

Namera ovakve analize i njenih metoda je da se identifikuju jaka pravila između raspoloživih podataka. U maloprodaji ova tehnika analize pomaže prodavcima da nađu najčešće kupljenje proizvode. Izvučene informacije se mogu koristiti kao osnova za donošenje poslovnih odluka, kako bi našao mogućnost za prodaju povezanih proizvoda, optimizaciju rasporeda proizvoda, upravljanje inventarom, marketinške aktivnosti (plasman i promocija proizvoda), unakrsnu prodaju (cross-selling) i dodatnu prodaju (up-selling). [2] Analiza potrošačke korpe je transakcija podataka koja sadrži: kupce, naloge i proizvode odnosno stavke. Sva tri nivoa su veoma važna za davanje odgovora na sledeća pitanja:

- Koji je prosečan broj naloga po potrošaču?
- Koji je prosečan broj određenih proizvoda po nalogu?
- Koji je prosečan broj proizvoda po nalogu?
- Koji je procenat klijenata koji su kupili određeni proizvod?
- Koji je prosečan broj naloga potrošača koji sadrži određeni proizvod?

Upotreba market basket analize je značajno porasla uvođenjem elektronskog poslovanja. Iako sam naziv analize asocira na usku primenu u supermarketima i trgovini, važno je da napomenemo da postoje mnoge druge oblasti u kojima se ovaj tip analize može primenjivati, kao što su: [7]

- pretraživanje korišćenja veba
- detekcije upada na mrežu
- neprekidna proizvodnja
- bankarstvo
- bioinformatika

Analiza potrošačke korpe se ne odnosi samo na jednu tehniku, to je skup tehnika koje sagledavaju poslovne procese i probleme u cilju razumevanja potrošača i njihovih transakcija. Postoji širok spektar algoritama za obavljanje ovog tipa analize koji je dostupan na raznim platformama. Mi ćemo koristiti jedan od najčešćih i najpoznatijih algoritama, a to je Apriori algoritam. [2]

7.2 Pravila asocijacije

Potrebe za korišćenje ovih pravila su otkrivanje potrošačkih navika kupaca, otkrivanje niza transakcija koje sugerišu na prevare u poslovanju, otkrivanje navika i tipova posetilaca web portala radi bolje personalizacije istih itd. Postoji mnogo načina da se vidi sličnost i povezanost između podataka, ova pravila su samo jedan od njih. Pravila asocijacije koriste if this / then that (ako ovo/onda to) pravila koja pomažu da se otkriju veze između naizgled nepovezanih podataka u relacionoj bazi podataka. Ova pravila se lako razumeju, ali nisu uvek korisna za analizu. Korisna pravila sadrže visok kvalitet adekvatnih informacija.

U terminologiji ovih pravila najčešće se srećemo sa: [8]

Elementom (item) kao sastavnim delom transakcija

Skupom elemenata (itemset) koji predstavlja više spojenih delova transakcije

Transakcijom (transaction) koje se sastoje od više skupova elemenata i razlikuju se po broju elemenata i rednim brojem transakcije. Više transakcija čine skup transakcija koje daju informacije za elemente koji se zajedno pojavljuju. Korišćenjem tablica koje su napravljene uz pomoć transakcija možemo utvrditi frekventnost pojavljivanja određenih parova elemenata.

Podrškom/značajem (support) kao merom frekventnosti parova (x, y), gde je N ukupan broj transakcija. Obeležava se sa: $supp(xy)=\%$ i odnosi se na elemente ili skup elemenata.

$$Support = \frac{freq(X,Y)}{N}$$

Pouzdanost/verovatnoća (confidence) se odnosi na pravila

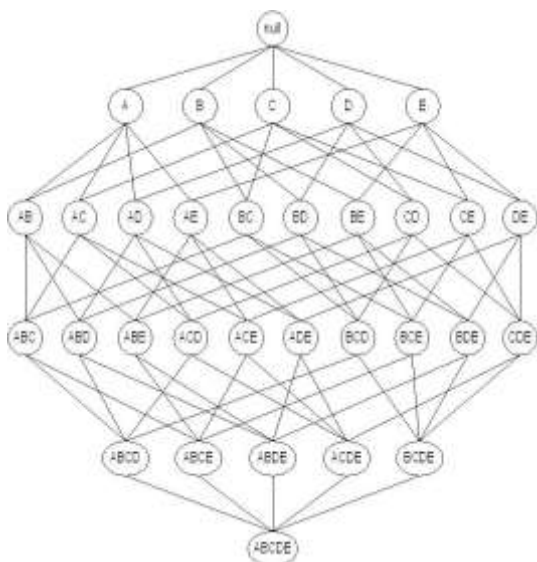
$$Confidence = \frac{freq(X,Y)}{freq(X)}$$

Poboljšanje (lift) služi za ocenjivanje kvaliteta određenih pravila, odnosno koliko smo sigurni da će potrošač kupiti proizvod koji čini desnu stranu pravila, znajući levu stranu pravila u odnosu na slučajno pogađanje.

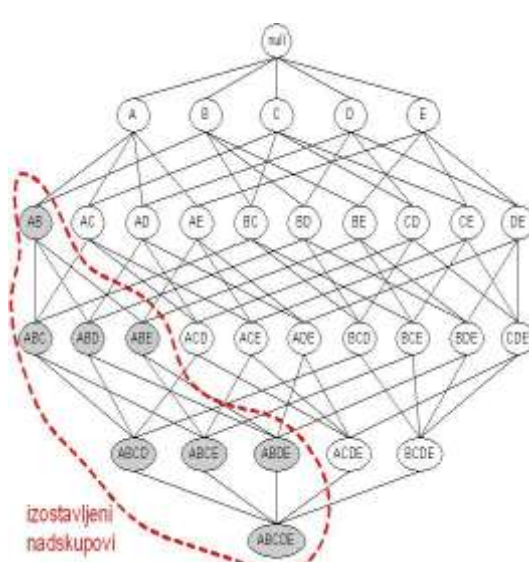
$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

7.3 Apriori algoritam

Apriori algoritam je jedan od najpoznatijih algoritama za pretraživanje pravila asocijacije, koji je dizajniran da radi u bazama podataka koji sadrže transakcije. Izvlači potrebne informacije iz seta podataka i transformiše ih u razumljivu strukturu radi lakšeg i daljeg korišćenja. Osnovna definicija ovog algoritma glasi: “Svaki podskup velikog skupa, takođe mora biti veliki skup”. Apriori koristi funkciju pretrage u širinu (BFS) i nalaženje kandidata sa osobinom zatvaranja na dole. Problem otkrivanja svih pravila možemo podeliti u više delova i time smanjiti vremensko trajanje analize: ograničenje broja elemenata i transakcija, otkrivanje velikih skupova (frequent itemsets) i uvođenje pravila. [4] Uvođenje pravila predstavlja traženje povezanosti među skupovima elemenata. Broj kombinacija skupova raste skoro eksponencijalno sa brojem elemenata. Upravo zbog ovog povećanja elemenata je vremenski neizvodljivo izvršiti analizu. Tada primenjujemo kategorizaciju elemenata odnosno taksonomiju. Drugi način je podela baze podataka na više particija koje ne moraju biti iste veličine. Zatim u svakoj particiji treba pronaći velike skupove elemenata. Algoritam primenjujemo sve dok ne izfiltriramo sve velike skupove i više ne možemo pronaći nijedan novi koji je veći od prethodnog ili su skupovi prazni.



Slika 16 Generisanje kandidata



Slika 17 Izostavljeni nadskupovi - pruning

8 Opis softvera

8.1 R studio

R studio je besplatno (open-source) softversko okruženje i programski jezik za računarsku statistiku i grafičku analizu. Glavne karakteristike R programa jeste činjenica da je besplatan i da ima dosta podrške na internetu. Sasvim je sličan drugim programskim paketima koji nisu besplatni, kao što je MatLab. Takođe je razumljiv programskim jezicima C++ i Fortran. [3] Njegova primena je široko rasprostranjena. Koriste ga istraživači i kompanije za statističke analize i rudarenje podataka. Prva verzija ovog programa je izašla 2009. godine. U nastavku su opisane neke od njegovih najosnovnijih opcija i mogućnosti. [5]

8.1.1 Radno okruženje

Radno okruženje R studija sadrži sledeće osnovne tabove:

Konzola (Console) - mesto za kucanje i pregled komandi

Ambijent / okruženje (Environment) - prikazuje sve aktivne objekte, vrednosti i funkcije koje su kreirane tokom sesije. Koji se mogu pregledati i na levoj strani klikom na presečeni kvadratić.

Istorija(History) - pokazuje spisak do sada urađenih komandi čime pomaže pri testiranju i pokretanju procesa. Postoji mogućnost čuvanja cele istorije kao zaseban fajl ili deo komandi koji želite mogu biti vraćene u konzolni tab.

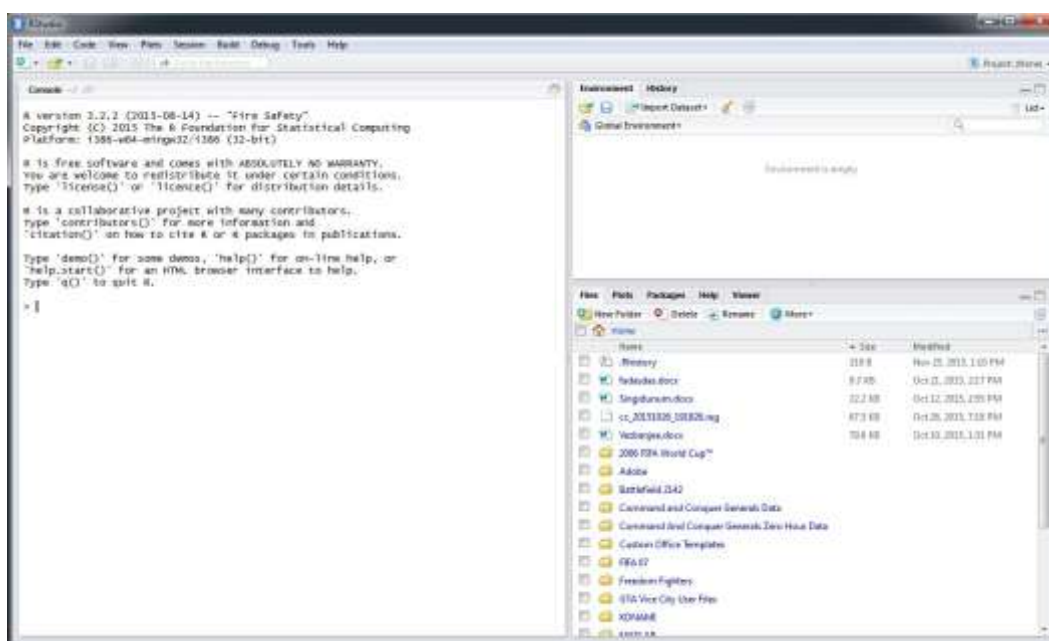
Fajlovi (Files) - prikaz svih fajlova i foldera u osnovnom radnom prostoru, koji se može podešavati odnosno menjati i birati po projektu (*slike 5,6,7,8*)

Plotovanje (Plots) - pokazuje sve tražene grafikone

Paketi (Package) - pokazuje sve pakete i dodatke koji su potrebni za određene procese. Dodatni paketi se mogu instalirati po potrebi klikom na ikonicu "Install"

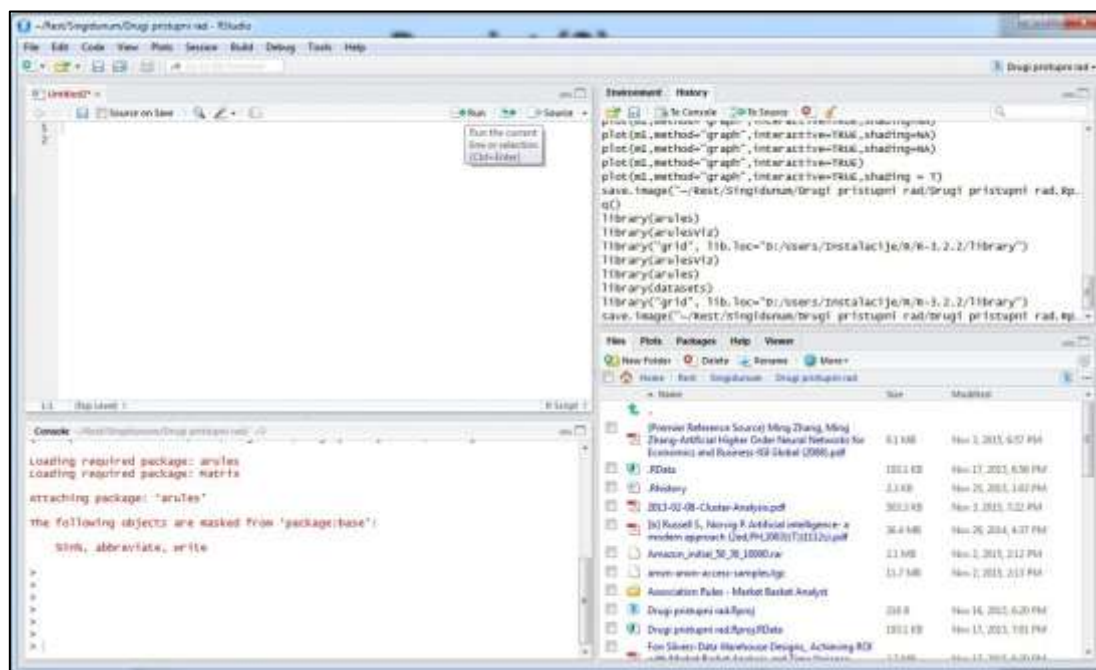
Pomoć (Help) - daje dodatne informacije i objašnjenja

Pregled (Viewer)

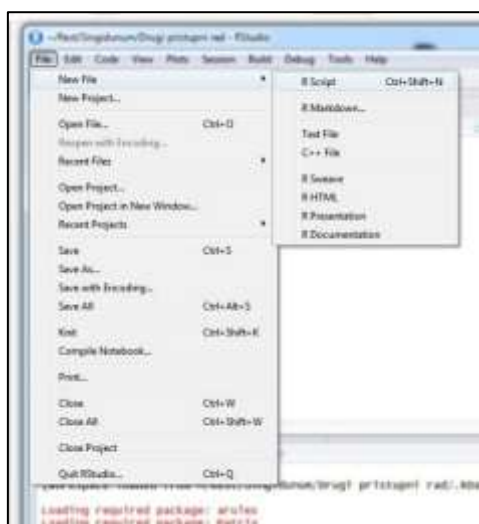


Slika 18 Radno okruženje R studija

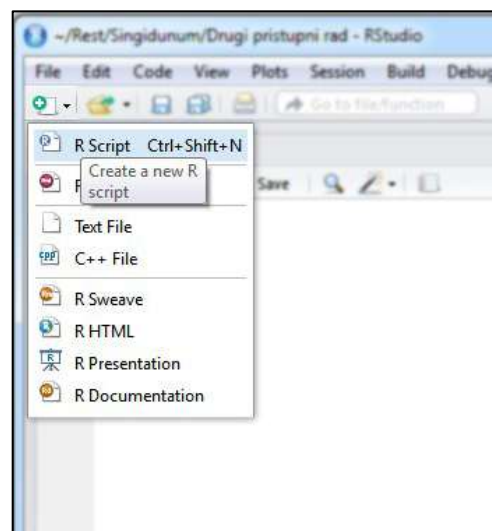
Pored svih nabrojanih komandi u radnom okruženju ovog softvera nalazi se i R skript(s), koji služi kao dnevnik evidencije o radu projekta. Za kreiranje nove skripte idite na *File/New File/ R Script* ili jednostavno putem komande *Ctrl+Shift+N* otvorite novu skriptu. Ovde možete kucati komande i klikom na ikonu “Run” ili prečicom *Ctrl+R* ubaciti komandu u konzolu. Naravno svaki Vaš skript možete i sačuvati.



Slika 19 Radno okruženje R-a sa R skriptom



Slika 20 Otvaranje R skripta način 1

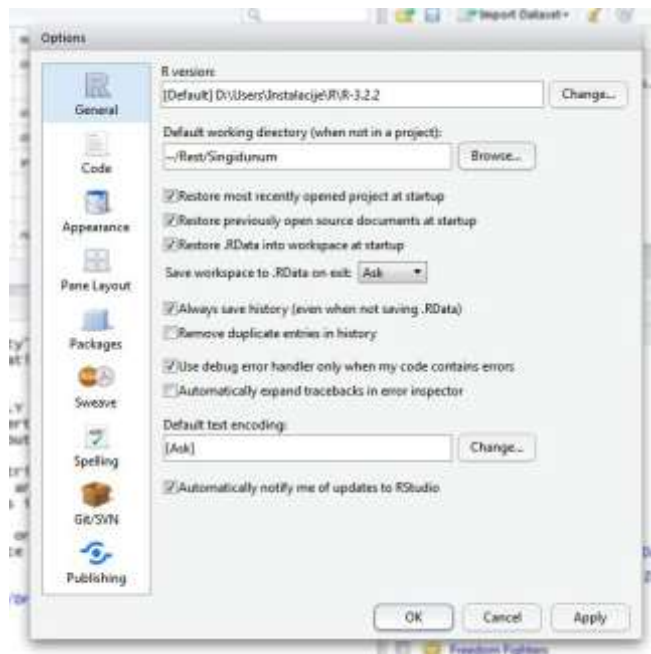
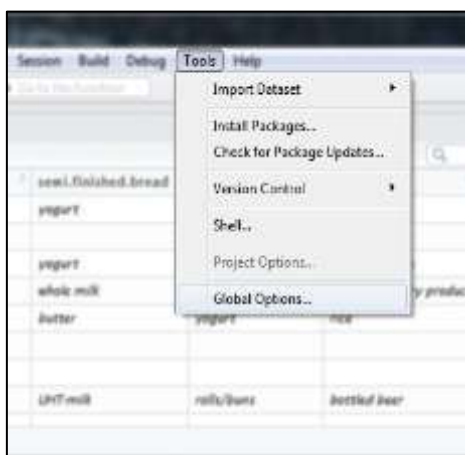


Slika 21 Otvaranje R skripta način 2

8.1.2 Podešavanje početnog radnog direktorijuma

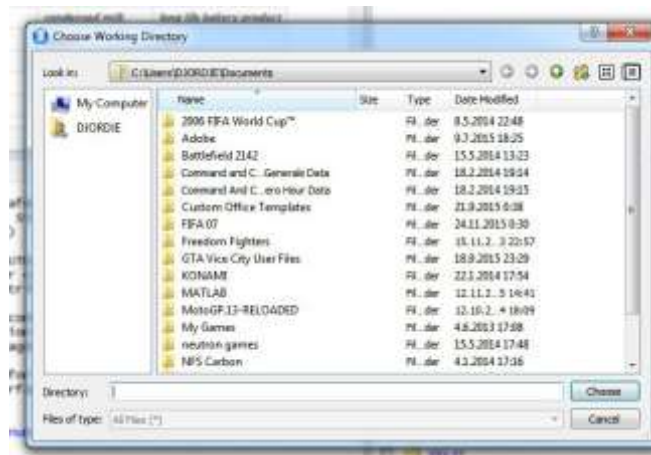
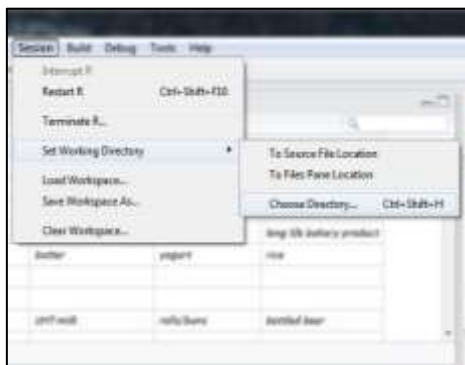
Početni radni direktorijum se menja ulaskom u meni *Tools/Global Options* (slika 8) i u odeljku *General* promeniti putanju *Default working directory* klikom na dugme *Browse* (slika 9), a u slučaju više različitih projekata možete promeniti privremeni radni direktorijum samo za određeni projekat (slika 10 i 11) ili kucati sledeće komande:

```
# Shows the working directory (wd)
getwd()
# Changes the wd setwd
"C:/myfolder/data")
```



Slika 22 Podešavanja početnog direktorijuma 1

Slika 23 Podešavanja početnog direktorijuma 2

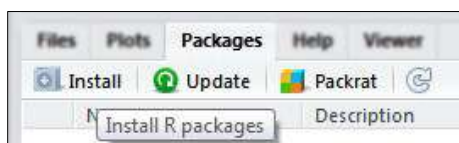


Slika 24 Privremeni radni direktorijum

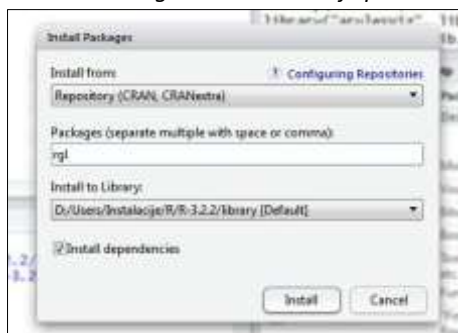
Slika 25 Odabir privremenog direktorijuma

8.1.3 Instaliranje paketa i dodataka

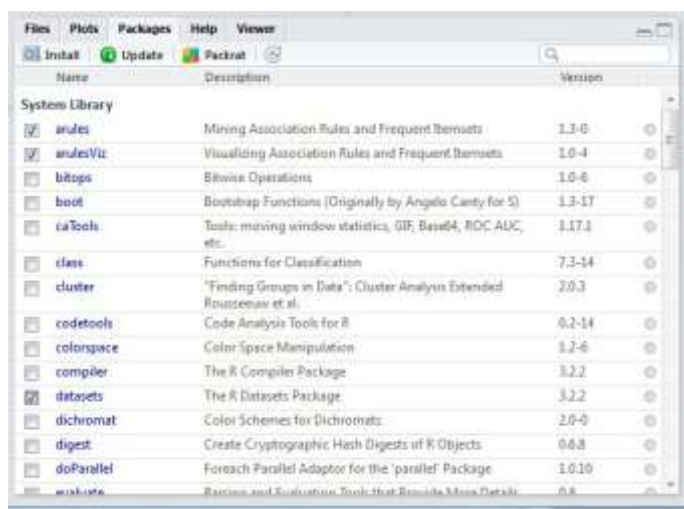
Package tab pokazuje spisak svih dodataka koji su uključeni u instalaciju R studija. Provereni paket se učitava, a u slučaju neučitavanja nijedna komanda vezana za taj paket neće raditi i tražiće da ga selektujete. Paketi se instaliraju klikom na ikonu “Install” u meniju Packages. Drugi način da instalirate paket je kucanje komande *library(željeni paket)* u konzoli, tada se paket automatski proverava i biće spreman za dalji rad. Pakete možete i ažurirati klikom na ikonu “Update”, ako u tom trenutku ima novijih verzija.



Slika 26 Dugme za instaliranje paketa



Slika 27 Instaliranje paketa rgl



Slika 28 Lista ponuđenih i instaliranih paketa

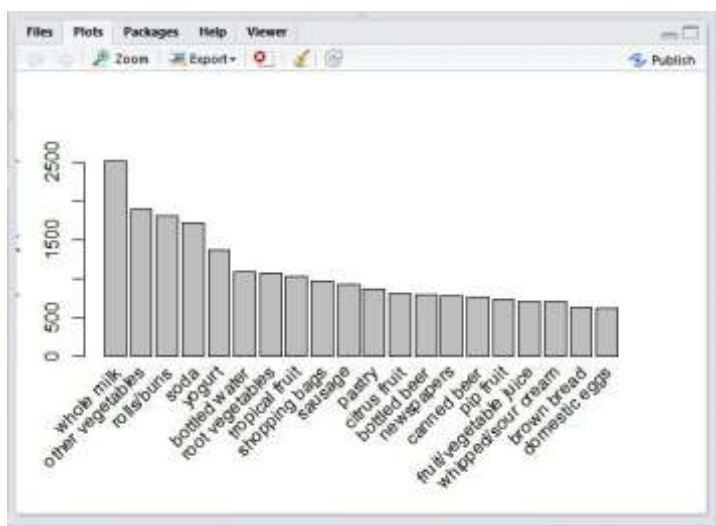
U ovom primeru instaliramo paket rgl (paket za plotovanje 3D grafikona preko OpenGL-a) koga nema u osnovnim paketima R studija. Klikom na ikonu “Install” u meniju Packages, otvara se iskaćući prozor gde ispisujemo ime željenog paketa i klikom na “Install” u donjem desnom uglu potvrđujemo i završavamo instalaciju. Zatim će se instalirani paket naći u listi ponuđenih paketa.

8.1.4 Plotovanje - vizualizacija rezultata

“Plots” meni prikazuje tražene grafikone koji se mogu listati pomoću strelica levo i desno u slučaju da postoji više grafikona u projektu i zumirati u samom prozoru R studija. Grafikoni se mogu eksportovati klikom na ikonu “Export” u jedan od formata za slike ili PDF format. U pod meniju ikone “Export” postoji i opcija “Copy to Clipboard” u kojoj možete birati veličinu exportovanja i direktno kopirati u druge programe. Postoje opcije za parcijalno i za celokupno brisanje urađenih grafikona, kao i njihovo osvežavanje. Sledeća opcija je dugme “Publish” koja služi za direktno publikovanje grafikona. Postoji i plotovanje 3D grafikona, ali se oni ne mogu izvesti odnosno sačuvati jer se otvaraju u posebnom prozoru.

```
> itemFrequencyPlot(Groceries,topN=20,type="absolute")  
>
```

Slika 29 Komanda za plotovanje najčešćih 20 proizvoda



Slika 30 Plotovanje najčešćih 20 proizvoda

9 Studija slučaja

9.1 Uvod u analizu

Set podataka “Namirnice” (Groceries) koji ćemo koristiti u analizi sadrži transakcije svih proizvoda koje su kupljeni u jednoj radnji od strane potrošača za određeni vremenski period. Elementi ovih transakcija su elementi skupa svih proizvoda. [3] Iz seta podataka želimo da generišemo određena pravila. Zbog uštede vremena za dobijanje traženih pravila iz velike baze podataka uvodimo pojedina ograničenja. [6] Analizom ovog seta podataka dolazimo do podataka koji su najčešći proizvodi kupljeni u toj radnji. Zatim ćemo dobiti i informaciju koja je verovatnoća da će potrošač sa određenim kupljenim proizvodima kupiti dodatni proizvod. Takođe ćemo videti koliko postoji pravila u setu podataka i sa koliko proizvoda ima najviše pravila. Sortiraćemo pravila po verovatnoći i sužavati i proširivati pretragu po potrebi. Na kraju analize biće urađena vuzalizacija dobijenih podataka.

9.2 Analiza seta podataka “Namirnice” [8]

9.2.1 Učitavanje radnih biblioteka i baze podataka:

- *Učitavanje biblioteka*

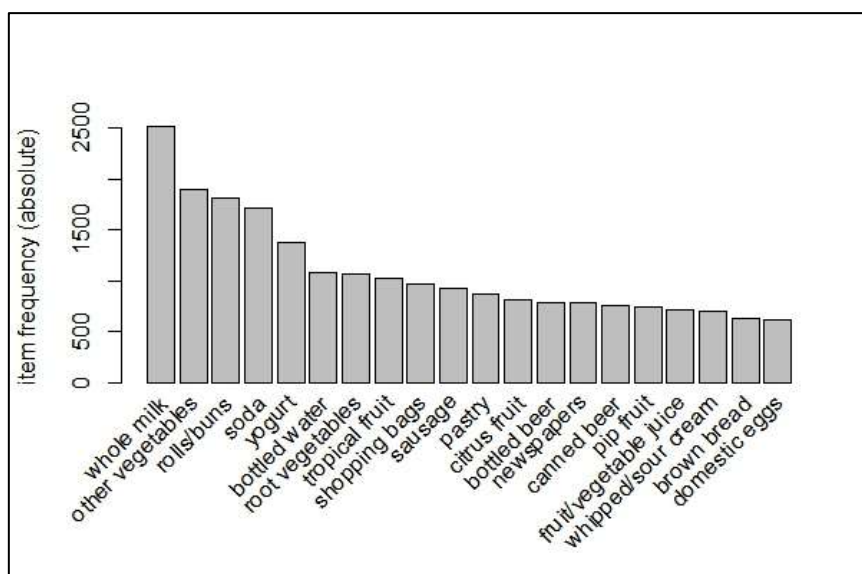
```
> library(arules)
> library(arulesviz)
> library(datasets)
```

- *Učitavanje baze podataka "Groceries"*

```
> data("Groceries")
```

9.2.2 Prikaz grafikona prvih (najčešćih) 20 proizvoda iz baze podataka

```
> itemFrequencyPlot(Groceries, topN=20, type="absolute")
```



Slika 31 Plotovanje najčešćih 20 proizvoda iz naše baze podataka

9.2.3 Uvođenje algoritma i prvih pravila

- Ubacujemo algoritam Apriori i stavljamo parametre *SUPPORT* i *CONFIDENCE* na minimum čime sužavamo pretragu i testiramo

```
> m1 <- apriori(Groceries, parameter = list(supp = 0.001, conf = 0.8))
```

Apriori

Parameter specification:

confidence	minval	smax	arem	aval	originalsupport	support	minlen	maxle
0.8	0.1	1	none	FALSE	TRUE	0.001	1	

10 rules
ext
FALSE

Algorithmic control:

filter	tree	heap	memopt	load	sort	verbose
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

Absolute minimum support count: 9

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [157 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 done [0.01s].
writing ... [410 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
```

9.2.4 Prikazujemo prvih pet pravila i postavljamo prikaz sa dve cifre.

- Ovim dobijamo da ako potrošač kupuje supu i flaširano pivo, postoji verovatnoća od 92% da će kupiti i mleko.

```
> options(digits=2)
> inspect(m1[1:5])
```

	lhs	rhs	support	confidence	lift
1	{liquor, red/blush wine}	=> {bottled beer}	0.0019	0.90	11.2
2	{curd, cereals}	=> {whole milk}	0.0010	0.91	3.6
3	{yogurt, cereals}	=> {whole milk}	0.0017	0.81	3.2
4	{butter, jam}	=> {whole milk}	0.0010	0.83	3.3
5	{soups, bottled beer}	=> {whole milk}	0.0011	0.92	3.6

9.2.5 Sumiramo set podataka i dobijamo sledeće informacije

Sumiramo čime vidimo moguće nepravilnosti i generalni opis podataka. Broj pravila je 410, najviše pravila (229) sadrži četiri proizvoda, zatim su tu minimum i maksimum proizvoda koji sadrže pravila, pregled pravila asocijacije (support, confidence, lift) i ukupni podaci sa minimumom.

```
> summary(m1)
```

set of 410 rules

rule length distribution (lhs + rhs): sizes

	3	4	5	6
	29	229	140	12

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	3.0	4.0	4.0	4.3	5.0	6.0

summary of quality measures:

	support	confidence	lift
Min.	:0.00102	Min. :0.80	Min. : 3.1
1st Qu.:	0.00102	1st Qu.:0.83	1st Qu.: 3.3
Median :	0.00122	Median :0.85	Median : 3.6
Mean :	0.00125	Mean :0.87	Mean : 4.0
3rd Qu.:	0.00132	3rd Qu.:0.91	3rd Qu.: 4.3
Max. :	0.00315	Max. :1.00	Max. :11.2

mining info:

	data	ntransactions	support	confidence
	Groceries	9835	0.001	0.8

9.2.6 Sortiranje parametara (proizvoda) po verovatnoći (CONFIDENCE)

```
> m1<-sort(m1, by="confidence", decreasing=TRUE)
> inspect(m1[1:5])
```

lhs	rhs	support	confidence	lift
1 {rice, sugar}	=> {whole milk}	0.0012	1.00	3.9
2 {canned fish, hygiene articles}	=> {whole milk}	0.0011	1.00	3.9
3 {whipped/sour cream, house keeping products}	=> {whole milk}	0.0012	0.92	3.6
4 {rice, bottled water}	=> {whole milk}	0.0012	0.92	3.6
5 {soups, bottled beer}	=> {whole milk}	0.0011	0.92	3.6

9.2.7 Sužavanje pretrage na tri proizvoda

- *Pravilo broj tri je veoma dugo, pa smanjujemo pretragu na maksimalno tri proizvoda dodavanjem parametra MAXLEN*

```
> m1 <- apriori(Groceries, parameter = list(supp = 0.001, conf = 0.8, maxlen=3))
```

lhs	rhs	support	confidence	lift
1 {liquor,red/blush wine}	=> {bottled beer}	0.0019	0.90	11.2
2 {curd,cereals}	=> {whole milk}	0.0010	0.91	3.6
3 {yogurt,cereals}	=> {whole milk}	0.0017	0.81	3.2
4 {butter,jam}	=> {whole milk}	0.0010	0.83	3.3
5 {soups,bottled beer}	=> {whole milk}	0.0011	0.92	3.6

9.2.8 Otpuštanje viškova

- *Ponekad će se pravila ponoviti. Na analitičarima je odabir da li će suziti pretragu ili eliminisati ponovljena pravila*

```
> subset.matrix <- is.subset(m1, m1)
> subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
> redundant <- colSums(subset.matrix, na.rm=T) >= 1
> m1.pruned <- m1[!redundant]
> m1<- m1.pruned
> subset.matrix <- is.subset(m1, m1)
```

9.2.8.1 Ciljanje proizvoda

- Znajući kako se generišu pravila i ograničava izlaz možemo postaviti da ciljani proizvodi stvaraju pravila. Šta kupci verovatno kupuju pre kupovine mleka?

```
> m1<-apriori(data=Groceries, parameter=list(supp=0.001,conf = 0.08),
appearance = list(default="lhs",rhs="whole milk"), control = list(verbose=F))
> m1<-sort(m1, decreasing=TRUE,by="confidence")
> inspect(m1[1:5])
```

lhs	rhs	support	confidence	lift
1 {rice, sugar}	=> {whole milk}	0.0012	1	3.9
2 {canned fish, hygiene articles}	=> {whole milk}	0.0011	1	3.9
3 {root vegetables, butter, rice}	=> {whole milk}	0.0010	1	3.9
4 {root vegetables, whipped/sour cream, flour}	=> {whole milk}	0.0017	1	3.9
5 {butter, soft cheese, domestic eggs}	=> {whole milk}	0.0010	1	3.9

9.2.9. Menjanje pravila

- *SUPPORT ostavljamo isti, a CONFIDENCE menjamo na 0,15 i ubacujemo MINLEN proizvoda da je 2. Odnosno stavljamo mleko sa leve strane.*

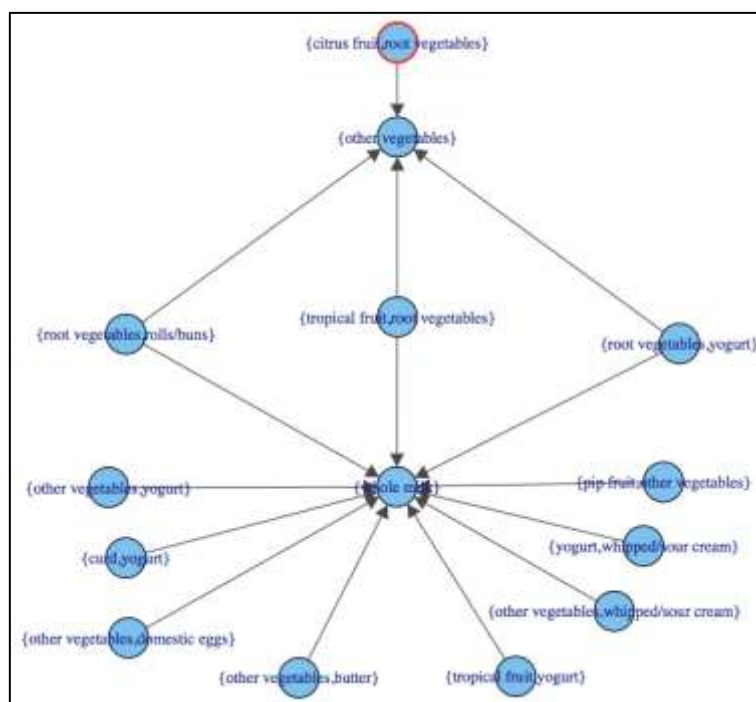
```
> m1<-apriori(data=Groceries, parameter=list(supp=0.001,conf = 0.15,minlen=2),
+           appearance = list(default="rhs",lhs="whole milk"),
+           control = list(verbose=F))
> m1<-sort(m1, decreasing=TRUE,by="confidence")
> inspect(m1[1:5])
```

lhs	rhs	support	confidence	lift
6 {whole milk}	=> {other vegetables}	0.075	0.29	1.5
5 {whole milk}	=> {rolls/buns}	0.057	0.22	1.2
4 {whole milk}	=> {yogurt}	0.056	0.22	1.6
2 {whole milk}	=> {root vegetables}	0.049	0.19	1.8
1 {whole milk}	=> {tropical fruit}	0.042	0.17	1.6

9.2.10. Vizualizacija

- Poslednji korak je vizualizacija. Uradićemo mapu pravila u grafikonu uz pomoć biblioteke ARULESVIZ koju smo prethodno instalirali.

```
> library(arulesviz)
> plot(rules,method="graph",interactive=TRUE,shading=NA)
```



Slika 32 Vizualizacija postojećih pravila

10 Zaključak

Oslanjanje na podatke i njihovo značenje u donošenju poslovnih odluka jedne kompanije neprestano raste. Gotovo svaki podatak zahteva analizu, radi izvlačenja njegovog značenja koji će se koristiti u daljim poslovnim odlukama. Uvođenje sistema poslovne inteligencije u kompanije je u porastu, što samim tim dovodi do većeg oslanjanja na njene procese. Takođe sve više postoji potreba za uvođenje analiza poslovne inteligencije u realnom vremenu (Real time) i za konstantnim svežim informacijama. Čemu teže najnovije tehnologije poslovne inteligencije. Napredak tehnologije, naročito memorije doveo je do pojave operativne inteligencije, koja ima sposobnost da analizira uživo, brzo menja podatke i odmah daje povratne informacije čime poslovnu inteligenciju stavlja na jedan viši nivo i pruža nove mogućnosti. Sve veće potraživanje ove oblasti stvara i sve više poslova za stručnjake kao što su ekonomisti, data i biznis analitičari, data IT inženjeri, IT inženjeri poslovne inteligencije i statističari, a istraživanja pokazuju da će se taj trend nastaviti i u budućnosti.

Transakcijski podaci sa POS terminala predstavljaju bogat i vredan izvor informacija o potrošačkim navikama kupaca, ali za pronalaženje svih korisnih informacija nisu dovoljne samo OLAP analize (On-Line Analytical Processing), već je potrebna primena novih tehnika kako bi se dobili odgovori na sva interesantna pitanja. Tehnike analize rudarenjem podataka (Data Mining) su relativno nove metode analize podataka koje omogućavaju izvlačenje informacija koje na prvi pogled nisu vidljive niti su dostupne klasičnim metodama.

Primer takve analize je analiza pravilima asocijacije. Korišćenje odgovarajućeg softvera i integrisanog programskog jezika omogućava analizu velikih količina transakcijskih elemenata i podataka. Softver R studio radi na proverenim tehnologijama čijom se primenom dobijaju dobre performanse i skalabilan sistem za dubinsku analizu podataka. Performanse se ocrtavaju u mogućnostima analize velikih količina podataka (preko 1GB), a skalabilnost u mogućnosti korišćenja aplikacije od strane više korisnika analize istih. Analizom dobijamo informacije o tome koji se proizvodi ili grupe proizvoda zajedno kupuju sa svim ostalim informacijama koje omogućavaju vrednovanje otkrivenih pravila asocijacije (podrška, pouzdanost, poboljšanje i interesantnost. Informacije dobijene na ovaj način mogu biti značajne za razumevanje kupovnih navika potrošača, što se može iskoristiti u CRM-u, marketingu, prodaji i organizaciji supermarketa.

Literatura

- [1] Nedo Balaban, Živan Ristić „Poslovna inteligencija“, Ekonomski fakultet Subotica, 2006.
- [2] Prof. dr Angelina Njeguš „Poslovni informacioni sistemi“, Univerzitet Singidunum, Beograd, 2015.
- [3] Prof. dr Milan Milosavljević, Doc. dr Vladislav Mišković „Elektronska trgovina“, Univerzitet Singidunum, Beograd, 2011.
- [4] Doc. dr Vladislav Mišković „Inteligentna podrška elektronskoj trgovini“, Univerzitet Singidunum, Beograd, 2013.
- [5] E. Turban, R. Sharda, J.F. Aronson, D. King „Business intelligence: A manager approach“, Prentice Hall, 2008.
- [6] S. Chaudhuri, U. Dayal, V. Narasayya „An Overview of Business Intelligence Technology“, ACM, 2011.
- [7] P. Bocij, A. Greasley, S. Hickie „Business Information Systems - Technology, Development and Management for the e-Business“, Prentice Hall, 2015.
- [8] <http://www.olap.com>
- [9] <http://www.targit.com>
- [10] J. Han, M. Kamber, J. Pei „Data Mining: concept and techniques“, Elsevier, Waltham USA, 3rd ed. 2012.
- [11] M. Berry, G. Linoff „Data Mining Technique: for marketing, sales and customer relationship management“, Wiley, Indianapolis USA, 2nd ed. 2004.
- [12] Prof. dr M. Veinović, Prof. dr G. Šimić, Prof. dr A. Jevremović, Prof. dr I. Franc „Baze podataka“, Univerzitet Singidunum, Beograd, 2015.
- [13] X. Wu, et al. „Top 10 algorithms in data mining“, Springer - Verlag, London, 2007.
- [14] Paul Torfs, Claudia Braver „A (very) short introduction to R“, Wageningen University, Netherland, 2014.
- [15] Mihaela Vranić „Dubinska analiza podataka: određeni postupci u području deskriptivne analize“, Fakultet elektrotehnike i računarstva, Zagreb, 2008.
- [16] dr Stipe Lovreta „Menadžment odnosa sa kupcima“, Ekonomski fakultet, Beograd, 2010.
- [17] http://www.albionresearch.com/data_mining/market_basket.php
- [18] <http://www.salemmarafi.com/code/market-basket-analysis-with-r/>
- [19] <http://snowplowanalytics.com/documentation/recipes/catalog-analytics/market-basket-analysis-identifying-products-that-sell-well-together.html>

Lista slika

Slika 1 Poslovna inteligencija.....	4
Slika 2 Tipična arhitektura poslovne inteligencije	5
Slika 3 CRM sistem.....	6
Slika 4 Poređenje besplatnih CRM sistema.....	7
Slika 5 ERP sistem	8
Slika 6 Elektronsko poslovanje	11
Slika 7 OLAP kocka	16
Slika 8 MOLAP arhitektura.....	17
Slika 9 ROLAP arhitektura	18
Slika 10 HOLAP arhitektura	18
Slika 11 Amazon struktura sajta 1	19
Slika 12 Amazon struktura sajta 2	20
Slika 13 Amazon struktura sajta 3.....	20
Slika 14 Amazon Listmania	21
Slika 15 Market basket analiza.....	22
Slika 16 Apriori - Generisanje kandidata	24
Slika 17 Apriori - izostavljeni nadskupovi (pruning).....	24
Slika 18 Radno okruženje R studija	25
Slika 19 Radno okruženje R studija sa R skriptom	26
Slika 20 Otvaranje R skripta (način 1)	26
Slika 21 Otvaranje R skripta (način 2)	26
Slika 22 Podešavanje početnog direktorijuma 1.....	27
Slika 23 Podešavanje početnog direktorijuma 2.....	27
Slika 24 Privremeni radni direktorijum	27
Slika 25 Odabir privremenog direktorijuma.....	27
Slika 26 Dugme za instaliranje paketa	28
Slika 27 Instaliranje paketa rgl	28
Slika 28 Lista ponuđenih i instaliranih paketa	28
Slika 29 Komanda za plotovanje najčešćih 20 proizvoda	29
Slika 30 Plotovanje najčešćih 20 proizvoda	29
Slika 31 Plotovanje najčešćih 20 proizvoda iz naše baze podataka	31
Slika 32 Vizualizacija postojećih pravila	35