

**Univerzitet u Beogradu
Matematički fakultet**

Seminarski rad

**Otkrivanje kupovnih navika i strategija prodaje u pekarskom
lancu**

Studenti: Emilija Stevanović 100/2019,
Iva Čitlučanin 143/2019

Mentor: Prof. Dr. Nenad Mitić

Decembar, 2023

Sadržaj

1	Uvod	3
1.1	Upoznavanje sa podacima	3
1.2	Opis baze podataka	3
2	Pretprocesiranje podataka	5
3	Pravila pridruživanja	5
3.1	Uvodni pojmovi	5
3.2	Mere interesantnosti pravila	6
3.3	Apriori algoritam	8
3.4	Algoritam FP rasta	10
4	Određivanje artikla koji se prodaju zajedno	11
4.1	Phyton	12
4.2	SPSS	14
4.3	Upoređivanje	17
5	Artikli na popustu	18
6	Najveći promet	20
7	Deo godine sa najvećom zaradom	21

8	Negativno korelisana pravila pridruživanja	25
9	Zaključak	27
10	Literatura	28

1 Uvod

Razumevanje kupovnih navika potrošača i analiza podataka o prodaji igraju ključnu ulogu u savremenom poslovanju. Praćenje ponašanja potrošača omogućava dublji uvid u njihove preferencije, identifikaciju obrasca kupovine i otkrivanje veza među proizvodima koji se često kupuju zajedno. Ova analiza pruža trgovcima dragocene informacije koje mogu koristiti za prilagođavanje marketinških strategija, unapređenje upravljanja zalihama, poboljšanje iskustva kupovine i, na kraju, optimizaciju poslovanja.

1.1 Upoznavanje sa podacima

U ovom radu su korišćeni podaci pod nazivom "EXTENDED BAKERY" i predstavljaju skup podataka razvijenih za kurs "CPE 466, Knowledge Discovery in Data" na Kalifornijskom politehničkom univerzitetu.

Skup podataka sadrži informacije o prodaji tokom jedne godine za nekoliko malih pekara.

U cilju olakšane obrade i analize podataka, skup podataka je prvobitno dat u formatu .sql. Da bi se podaci učinili pogodnijim za dalju obradu, prebacujemo ih u format .csv (Comma-Separated Values), koji je pogodan za široku upotrebu i prilagođen za analizu podataka u raznim alatima za obradu podataka.

1.2 Opis baze podataka

Baza podataka koja sadrži tabele "Zaposleni", "Roba", "Artikli", "Lokacije" i "Računi" i koristiti se za praćenje informacija o svim zaposlenima, proizvodima, transakcijama u prodavnicama.

Sledi opis svake tabele:

- Tabela "Zaposleni" (**Employees**):

- Polja: ime (**Name**), prezime(**Surname**), datum_zapošljavanja(**HireDate**), pozicija(**Position**), id_prodavnice(**StoreNum**), id_radnika(**EmpId**).
 - Ova tabela sadrži informacije o zaposlenima u prodavnicama, uključujući ime, prezime, datum zaposlenja, poziciju zaposlenog, identifikator prodavnice u kojoj radi i identifikator radnika.
- Tabela "Roba" (**Goods**):
 - Polja: id, ukus, hrana (**Food**), cena (**Price**), tip(**Type**).
 - Ova tabela čuva informacije o raznim proizvodima ili robi koje prodavnica prodaje, uključujući identifikator, ukus, vrstu hrane, cenu u dolarima i tip proizvoda.
- Tabela "Stavke" (**Items**):
 - Polja: id, broj_racuna(**ReceiptNum**) , količina(**Quantity**).
 - Tabela čuva informacije o kupljenim artiklima, uključujući identifikator, broj računa i količinu kupljenih proizvoda. Postoji 30000 različitih proizvoda.
- Tabela "Lokacije" (**Locations**):
 - Polja: drzava (**State**), ulica (**Steet**), broj_prodavnice(**StoreNum**).
 - Ova sadrži podatke o lokacijama prodavnica, uključujući državu, ulicu i broj prodavnice.
- Tabela "Računi" (**Receipts**):
 - Polja: broj_računa (**ReceiptNumber**), datum_prodaje, vikend, keš (**IsCash**), id_radnika (**EmpId**), broj_prodavnice(**StoreNum**).
 - Čuvaju se informacije o transakcijama, uključujući broj računa, datum prodaje, informacije da li je vikend, načinu plaćanja, identifikatoru radnika koji je izdao račun i broju prodavnice gde je obavljena kupovina. Baza sadrži podatke od 75000 računa.

2 Pretprocesiranje podataka

Uklanjanje **nedostajućih vrednosti** i **duplikata** podataka je važan korak u procesu pripreme podataka za analizu podataka o prodaji. Nedostajući ili nepotpuni podaci mogu značajno uticati na rezultate analize te stoga zahtevaju pažljivu obradu. Postupak uklanjanja nedostajućih podataka uključuje identifikaciju i obradu bilo kakvih nepotpunih ili neispravnih informacija unutar skupa podataka. Ovi podaci mogu biti nedostajući iz različitih razloga, kao što su greške u unosu, nedostajuće vrednosti, ili problemi s prikupljanjem. Jedan od pristupa uklanjanju nedostajućih podataka je zamena tih nedostajućih vrednosti nekim podacima ili metodama popunjavanja praznina, ili jednostavno izostavljanje redova ili atributa sa nedostajućim podacima iz daljnje analize.

Autlajeri predstavljaju vrednosti koje značajno odstupaju od ostalih podataka u skupu podataka i njihovo prisustvo može uticati na interpretaciju rezultata analize, stoga je obrada autlajera ključan korak u pripremi podataka.

Kraćom proverom se može ustanoviti da preuzeti podaci koji se koriste u ovom radu ne sadrže nedostajuće vrednosti, duplirane podatke niti autlajere čime je obezbeđena osnova za naredne korake njihove analize.

3 Pravila pridruživanja

3.1 Uvodni pojmovi

Neka su X i Y dva skupa stavki. Tada se ravilo u oznaci $X \rightarrow Y$ naziva pravilo pridruživanja sa minimalnom podrškom *minsup* i minimalnom pouzdanošću *minconf* ako važi:

- podrška skupa $X \cup Y$ je $\geq \text{minsup}$
- pouzdanost pravila $X \rightarrow Y \geq \text{minconf}$

Ako su A i B skupovi stavki tada se:

- podrška (eng. support) pravila pridruživanja $A \Rightarrow B$ računa kao količnik broja transakcija koje sadrže A i B i ukupnog broja istih.

$$sup(A \Rightarrow B) = \frac{\#(A \cup B)}{N} \quad (1)$$

- pouzdanost (poverenje, eng. confidence) pravila pridruživanja $A \rightarrow B$ računa kao količnik broja transakcija koje sadrže A i B i onih koje sadrže A.

$$conf(A \Rightarrow B) = \frac{\#(A \cup B)}{\#(A)} \quad (2)$$

Svaki skup je čest ako se javlja više od *minsup* puta, gde vrednost *minsup* označava vrednost minimalne podrške koja se zadaju unapred.

Svaki skup je pouzdan ako se javlja više od *minconf* puta, *minconf* se takođe zadaje unapred.

Za dati skup transakcija T cilj je pronaći sva pravila koja imaju podršku veću od minimalne i pouzdanost veću od minimalne.

U slučaju velikih baza *minconf* se obično stavlja visoko (npr. 80%), a *minsup* je obično značajno niži (npr. 5-10%).

Visoka podrška znači da se pravilo javlja često i da je manje verovatno da je slučajno. Dok visoka pouzdanost govori o tome da je pravilo od interesa i da je prisutna velika uzročnost između elemenata.

3.2 Mere interesantnosti pravila

Rezultat algoritama za pronalaženje pravila pridruživanja može biti veliki skup. Nisu sva pravila interesantna i potrebno je iz velikog broja pravila pridruživanja izdvojiti najinteresantnija. Ne postoji najbolja mera, svaka mera ima svoje prednosti i mane, a odabir prave mere zavisi od ciljeva i od prethodne analize.

Za ilustraciju mera se obično koristi tabela na Slici 1, preuzetoj sa prezentacije profesora Nenada Mitića.

	B	\bar{B}	
A	f_{11}	f_{10}	f_{1+}
\bar{A}	f_{01}	f_{00}	f_{0+}
	f_{+1}	f_{+0}	N

A - присутно у трансакцији
 \bar{A} - није присутно у трансакцији
 f_{ij} - бројач учесталости

f_{10} - број транс. које садрже само A
 f_{11} - број транс. које садрже и A и B
 f_{01} - број транс. које садрже само B
 f_{00} - број транс. које не садрже ни A ни B
 f_{1+} - бројач подршке за A
 f_{0+} - бројач подршке за \bar{A}
 f_{+1} - бројач подршке за B
 f_{+0} - бројач подршке за \bar{B}

Slika 1: Tabela kontigenata

U nastavku su nekoliko nedostataka podrške i pouzdanosti:

- Podrška:
 - Osetljivost na retkost: Ako imamo nebalansirane podatke, gde je jedna klasa zastupljena znatno više od druge, visoka podrška može biti rezultat često prisutne klase, dok manje prisutna klasa možda neće imati dovoljno podrške za donošenje odluka.
- Pouzdanost:
 - Neusaglašenost sa stvarnim vrednostima: Visok stepen pouzdanosti može biti pogrešan ako se pravilo oslanja na grešku ili netačnu informaciju u podacima. To može dovesti do nepreciznih rezultata.
 - Osetljivost na veličinu uzorka: Ako imamo mali uzorak podataka, pouzdanost može biti niska zbog nedostatka dovoljnog broja instanci za formiranje relevantnih pravila.
 - Zavisnost o prethodnim pretpostavkama: Ako pravila zasnivaju na pretpostavkama koje nisu tačne ili su pregrube, pouzdanost može biti kompromitovana.

Važno je napomenuti da se ove mere često koriste u kombinaciji s drugim evaluacionim metrikama kako bi se dobila potpunija slika o performansama pravila pridruži-

vanja. U nastavku su druge mere koje se koriste.

- *Lift* - jedna od mera koja nadoknađuje nedostatak pouzdanosti i uzima u obzir podršku desne strane pravila.

$$Lift(A \implies B) = \frac{conf(A \implies B)}{sup(B)} \quad (3)$$

Lift vrednosti veće od 1 označavaju da je posledična strana pravila mnogo češća u transakcijama koje sadrže levu stranu pravila negu u transakcijama koje je ne sadrže, dok vrednost manja od 1 označavaju pravila čija je pouzdanost manja od očekivane. U praksi je poželjno koristiti vrednosti lift mere koje su veće od 1.1.

- Koeficijent korelacije -

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (4)$$

$$\rho_{ij} = \frac{sup(i, j) - sup(i) \cdot sup(j)}{\sqrt{sup(i) \cdot sup(j) \cdot (1 - sup(i)) \cdot (1 - sup(j))}} \quad (5)$$

Gde su $sup(i)$, $sup(j)$ i $sup(i, j)$ relativne podrške.

- Deployability (mogućnost širenja) - procenat skupa podataka koji zadovoljava uslov na levoj strani, ali ne i posledični deo na desnoj strani. U kontekstu potošačke korpe ova mera bi označavala procenat broj kupaca koji su kupili artikle sa leve strane, ali nisu ove sa desne.

$$deployability = \frac{Antecedent\ Support\ in\ \# \text{ of Records} - Rule\ Support\ in\ \# \text{ of Records}}{Number\ of\ Records} * 100 \quad (6)$$

3.3 Apriori algoritam

Apriori princip: Ako je skup stavki čest tada su svi njegovi podskupovi česti

- Zatvorenje naniže (eng. downward closure) svaki podskup čestog skupa stavki je takođe čest.

Posledica anti-monotonosti¹ : ako neki skup artikla A nije podržan (nije čest), onda ni bilo koji skup B gde važi $A \subseteq B$ nije čest, tj. ne treba da bude razmatran. Apriori algoritam² je pokazao moguće efektivno dobijanje skupa pravila pridruživanja u slučaju velikih skupova podataka.

Pseudokod:

Apriori(Transakcije T , Podrška minsup):

begin

$k=1$;

F_1 =sve ceste 1-stavke;

 while F_k nije prazno do

 begin

 Generisati C_{k+1} spajanjem stavke iz F_k ;

 Odbaciti stavke iz C_{k+1} koje ne zadovoljavaju zatvorenje naniže;

 Odrediti F_{k+1} brojanjem podrške (C_{k+1} , T);

 Odbaciti one sa manjom podrškom od minsup;

$k=k+1$;

 end;

 return(unija svih F_i , $i=1:k$);

end

¹mera f poseduje anti-monotonost ako $\forall X, Y$ važi $X \subset Y \implies f(X) \leq f(Y)$

²R. Agraval i R. Srikant 1994. godine

3.4 Algoritam FP rasta

Algoritam FP-growth (Frequent Pattern growth) je efikasan algoritam za otkrivanje često ponavljajućih obrazaca u velikim skupovima podataka, često korišćen u otkrivanju pravila pridruživanja. Osnovni koraci algoritma FP-rasta:

- Ulaz u algoritam je skup podataka gde su transakcije predstavljene kao skupovi stavki (itemsets).
- Prvi prolaz kroz skup podataka identifikuje frekventne stavke (itemsets) i određuje redosled njihove frekvencije.
- FP-stablo (Frequent Pattern tree) se gradi na osnovu frekventnih stavki. To je efikasna struktura podataka koja olakšava reprezentaciju i analizu frekventnih obrazaca.
- Pravi se header tabela koja sadrži informacije o svakoj stavki i njenoj povezanosti u listi čvorova u FP-stablu.
- Transakcije se sortiraju prema smanjenju redosledu frekventnosti stavki, čime se povećava efikasnost izgradnje FP-stabla.
- Svaka transakcija se dodaje u FP-stablo, uzimanjem u obzir redosleda frekventnosti stavki i korišćenjem veza u stablu.
- Postupak se rekurzivno ponavlja na svakom podstablu, kako bi se pronašli svi frekventni obrasci.
- Na osnovu frekventnih stavki i informacija iz FP-stabla, generišu se pravila pridruživanja.

Ovaj algoritam zamenjuje tradicionalni pristup prebrojavanju (kao što je Apriori algoritam) sa strukturom FP-stabla, što može značajno ubrzati proces otkrivanja frekventnih obrazaca.

	Apriori	FP Growth
Generisanje frekventnih obrazaca	Koristi generisanje kandidata za jednostruk, dvostruk, i trostruk set.	Gradi FP-stablo za izradu frekventnih obrazaca.
Proces kandidata	Koristi generaciju kandidata i proširuje frekventne podskupove jednim elementom u svakom koraku.	Gradi uslovno FP-stablo za svaki predmet u podacima.
Efikasnost	Zahteva više prolaza kroz bazu podataka, što može biti vremenski zahtevno kod velikog broja predmeta.	Zahteva samo jedan prolazak kroz bazu podataka u početnim koracima, čime je efikasniji, posebno za velike skupove podataka.
Memorija	Sačuvana je konvertovana verzija baze podataka u memoriji tokom izvođenja algoritma.	Čuva set kondicionalnih FP-stabala za svaki predmet u memoriji.
Pregled	Koristi pretragu u širinu (breadth-first search).	Koristi pretragu u dubinu (depth-first search).

Tabela 1: Poređenje Apriori i FP rast algoritama.

4 Određivanje artikla koji se prodaju zajedno

U lancima pekara, praćenje prodaje proizvoda i razumevanje koje stavke često kupuju zajedno može biti od suštinskog značaja za optimizaciju asortimana proizvoda, marketing strategija i povećanje ukupne prodaje.

4.1 Phyton

Za analizu u Phyton programskom jeziku, koristimo unapred pripremljenu tabelu (Slika 2) koja sadrži informacije o broju računa i vrsti artikla sa njegovim identifikacionim brojem.

	ReceiptNumber	Item_y
0	1	Cookie_21
1	12	Cookie_21
2	37	Cookie_21
3	41	Cookie_21
4	71	Cookie_21
...
266204	74932	Croissant_30
266205	74966	Croissant_30
266206	74972	Croissant_30
266207	74979	Croissant_30
266208	74980	Croissant_30

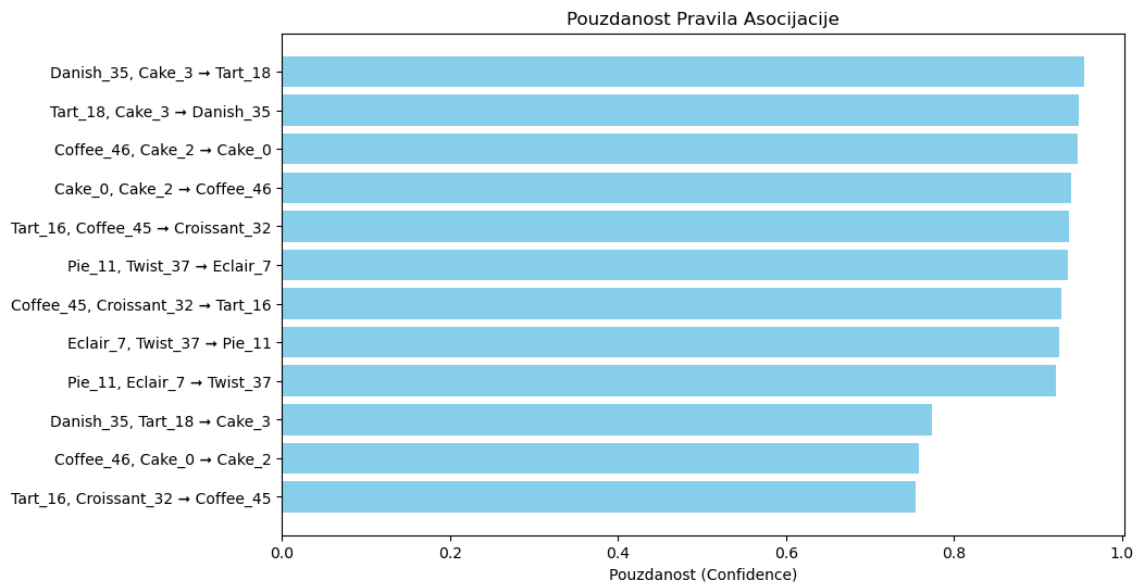
Slika 2: Prikaz dela podataka iz tabele

Apriori algoritam je primenjen na podatke kako bi se identifikovali česti skupovi artikala koji se često prodaju zajedno. Ovaj algoritam omogućava određivanje pravila asocijacije na osnovu podrške (support) i pouzdanosti (confidence) skupova proizvoda.

Za primenu Apriori algoritma nad ovim podacima, prvo je neophodno transformisati ih u format pogodan za analizu asocijativnih pravila. Grupisanjem podataka prema broju računa (Receipt Number), stvaramo pregled proizvoda kupljenih

na istom računu. Primena Apriori algoritma sa odabranim vrednostima minSup (minimalna podrška) i minConf (minimalna pouzdanost) na ovako organizovane grupe proizvoda je ključna. Vrednost minSup postavljena na 0,03 omogućava identifikaciju čestih kombinacija proizvoda, šireći analizu na različite moguće veze. Sa druge strane, vrednost minConf postavljena na 0,65 postavlja visok prag pouzdanosti, filtrirajući manje pouzdane veze i fokusirajući se na pravila sa visokom sigurnošću. Ove vrednosti su izabrane kako bi osigurale raznolikost identifikovanih veza među proizvodima u lancu pekara, istovremeno osiguravajući relevantnost i pouzdanost analiziranih pravila. Nakon primene Apriori algoritma na skupu podataka o kupovini, dobili smo pravila asocijacije koja pružaju uvid u povezanost između različitih proizvoda. Primer kako interpretiramo dobijene rezultate :

- $\{\text{Danish_35}, \text{Cake_3}\} \rightarrow \{\text{Tart_18}\}$ sa pouzdanošću od 0.95: Ovo pravilo ukazuje da se proizvod 'Tart_18' često kupuje (u 95% slučajeva) kada su već kupljeni proizvodi 'Cake_3' i 'Danish_35'.



Slika 3: Pravila dobijena primenom Apriori algoritma

Prethodno prikazana pravila asocijacije su identifikovana analizom računa iz svih prodavnica zajedno. Sada je cilj proveriti da li ova pravila važe za svaku prodavnicu pojedinačno. Ovo nam omogućava da razumemo da li postoje varijacije

u kupovnom ponašanju među različitim prodavnicama. Prvo, izdvajamo podatke o kupovini za svaku prodavnicu pojedinačno kako bismo stvorili skupove podataka koji odražavaju kupovne obrasce u svakoj prodavnici. Primena istog algoritma za identifikaciju pravila pridruživanja na podacima izdvojenim za svaku prodavnicu pojedinačno omogućila nam je da uporedimo dobijena pravila sa prethodno identifikovanim globalnim pravilima. Na osnovu ovog poređenja, zaključili smo da određena pravila pridruživanja važe u svim prodavnicama bez obzira na njihovu lokaciju. Pravila pridruživanja koja ostaju konzistentna u svim prodavnicama nezavisno od lokacije su sledeća:

- $\{\text{Danish_35}, \text{Cake_3}\} \rightarrow \{\text{Tart_18}\}$
- $\{\text{Tart_18}, \text{Cake_3}\} \rightarrow \{\text{Danish_35}\}$
- $\{\text{Pie_11}, \text{Eclair_7}\} \rightarrow \{\text{Twist_37}\}$

Takođe zaključujemo da u prodavnicama 2,3,4,5,6,7,8,10,15,16,18 i 20 važe sva globalna pravila dok u ostalim prodavnicama važi 9 od prethodno određenih 12 globalnih pravila.

4.2 SPSS

Prvi korak bio je određivanje pravila pridruživanja za sve prodavnice koristeći softver SPSS Modeler. Ovaj korak nam je omogućio da identifikujemo zajedničke karakteristike koje se primenjuju na sve prodavnice. Zatim su pronađena pravila pridruživanja za svaku prodavnicu pojedinačno, uzimajući u obzir specifičnosti svake lokacije. U nastavku je prikaz pravila pridruživanja za sve račune, ne uzimajući u obzir lokaciju, sortirana po pouzdanosti opadajuće (Tabela 2).

Consequent	Antecedent	Support %	Confidence %
Tart_18	Cake_3 and Danish_35	4.303	95.538
Danish_35	Cake_3 and Tart_18	4.337	94.774
Cake_0	Cake_2 and Coffee_46	3.524	94.741
Coffee_46	Cake_2 and Cake_0	3.553	93.959
Croissant_32	Tart_16 and Coffee_45	3.504	93.683
Eclair_7	Pie_11 and Twist_37	3.668	93.566
Tart_16	Croissant_32 and Coffee_45	3.537	92.801
Pie_11	Twist_37 and Eclair_7	3.712	92.457
Twist_37	Pie_11 and Eclair_7	3.727	92.093
Eclair_7	Twist_37 and Coffee_45	3.092	90.944
Pie_11	Twist_37 and Coffee_45	3.092	90.729
Eclair_7	Pie_11 and Coffee_45	3.103	90.546
Twist_37	Pie_11 and Coffee_45	3.103	90.417
Twist_37	Coffee_45 and Eclair_7	3.156	89.1
Pie_11	Coffee_45 and Eclair_7	3.156	89.016
Coffee_45	Pie_11 and Twist_37 and Eclair_7	3.432	81.352
Cake_3	Danish_35 and Tart_18	5.309	77.423
Coffee_45	Pie_11 and Twist_37	3.668	76.481
Cake_2	Coffee_46 and Cake_0	4.404	75.81
Coffee_45	Twist_37 and Eclair_7	3.712	75.754
Coffee_45	Tart_16 and Croissant_32	4.351	75.452
Coffee_45	Pie_11 and Eclair_7	3.727	75.385
Espresso_49	Tart_15 and Eclair_7	3.641	74.698
Cookie_29	Tart_17 and Frappuccino_47	3.596	74.416

Tabela 2: Pravila pridruživanja

Sledi primer pravila za specifičnu lokaciju.(Tabela 3)

Consequent	Antecedent	Support (%)	Confidence (%)
Tart_16	Croissant_32 and Coffee_45	3.343	96.85
Danish_35	Cake_3 and Tart_18	4.08	96.774
Tart_18	Cake_3 and Danish_35	4.133	95.541
Eclair_7	Twist_37 and Pie_11	3.053	93.966
Cake_0	Cake_2 and Coffee_46	3.264	93.548
Coffee_46	Cake_2 and Cake_0	3.317	92.063
Croissant_32	Tart_16 and Coffee_45	3.527	91.791
Pie_11	Twist_37 and Eclair_7	3.132	91.597
Twist_37	Pie_11 and Eclair_7	3.159	90.833
Tart_17	Cookie_29 and Frappuccino_47	3.053	89.655
Cookie_23	Cookie_24 and Lemonade_40	3.001	88.596
Lemonade_41	Cookie_24 and Lemonade_40	3.001	87.719
Coffee_45	Croissant_32 and Tart_16	4.08	79.355
Coffee_45	Twist_37 and Pie_11	3.053	76.724
Coffee_45	Twist_37 and Eclair_7	3.132	75.63

Tabela 3: Pravila za lokaciju 1

Pomoćnim programom napisanom u Python-u, provereno je da li pravila pridruživanja važe u svakoj prodavnici. Program je automatski analizirao podatke iz svake lokacije i uporedio ih sa zadatim pravilima, pružajući precizan uvid u usklađenost pravila sa stvarnim podacima.

Naredna pravila pridruživanja se izdvajaju bez obzira na lokaciju:

- {Danish_35, Cake_3} → {Tart_18}
- {Tart_18, Cake_3} → {Danish_35}
- {Twist_37, Eclair_7} → {Pie_11}
- {Pie_11, Eclair_7} → {Twist_37}

4.3 Upoređivanje

Kada uporedimo rezultate dobijene na dva načina, možemo primetiti da SPSS Modeler izdvaja više pravila pridruživanja, iako su pragovi za minimalnu podršku i pouzdanost isti. Razlika je u pravilima koja imaju vrednosti oko granice za ove dve mere. Pravila koja su izdvojena da važe bez obzira na lokaciju se gotovo i ne razlikuju.

5 Artikli na popustu

Kada se određeni proizvodi često kupuju zajedno, postoji potencijal za primenu strategija kao što je snižavanje cena jednog od proizvoda u tom paru ili grupi. Smanjenje cene može podstaći kupovinu jednog proizvoda i doprineti povećanju ukupne prodaje. Na osnovu prethodno određenih pravila pridruživanja koja vidimo na Slici 3 možemo odrediti artikle koji bi svojim popustom povećali ukupnu zaradu. U posmatranom lancu pekara, pravila pridruživanja su definisana parovima ili grupama proizvoda. Ova pravila se primenjuju kako bi se formirale kombinacije koje imaju potencijal za povećanje prodaje. Jedan od ključnih ciljeva pravila pridruživanja u ovom lancu pekara je identifikovanje najjeftinijeg proizvoda unutar datog para ili grupe. Kroz analizu cena svakog proizvoda unutar pravila pridruživanja, utvrđujemo proizvod koji će biti označen kao najjeftiniji. Ovaj proizvod postaje fokus marketinške strategije, jer će biti ponuđen sa dodatnim popustom kako bi se privukla pažnja potrošača. Ova taktika podstiče kupce da kupe ovaj proizvod, nadajući se da će istovremeno kupiti i druge proizvode iz istog para ili grupe.

Kroz Python programski jezik, pravila pridruživanja su analizirana i najjeftiniji proizvod za svako pravilo je identifikovan.

```
Najjeftiniji proizvod iz pravila 1 : Croissant_32 - Cena: 1.45
Najjeftiniji proizvod iz pravila 2 : Coffee_46 - Cena: 2.45
Najjeftiniji proizvod iz pravila 3 : Danish_35 - Cena: 1.15
Najjeftiniji proizvod iz pravila 4 : Twist_37 - Cena: 1.15
Najjeftiniji proizvod iz pravila 5 : Twist_37 - Cena: 1.15
Najjeftiniji proizvod iz pravila 6 : Croissant_32 - Cena: 1.45
Najjeftiniji proizvod iz pravila 7 : Twist_37 - Cena: 1.15
Najjeftiniji proizvod iz pravila 8 : Croissant_32 - Cena: 1.45
Najjeftiniji proizvod iz pravila 9 : Coffee_46 - Cena: 2.45
Najjeftiniji proizvod iz pravila 10 : Coffee_46 - Cena: 2.45
Najjeftiniji proizvod iz pravila 11 : Danish_35 - Cena: 1.15
Najjeftiniji proizvod iz pravila 12 : Danish_35 - Cena: 1.15
```

Slika 4: Izdvojeni najjeftiniji proizvodi iz pravila i njihova cena

Izdvojeni proizvodi prikazani na Slici 4 će se razmatrati kao potencijalni kandidati za primenu dodatnih popustia, s ciljem povećanja ukupne zarade i privlačenja pažnje kupaca.

U SPSS Modeleru, najpre su učitani podaci o pravilima pridruživanja, zatim su artikli odvojeni u posebnim kolonama i spojeni sa tabelom u kojoj se nalaze cene za svaki. Na taj način je moguće jednostavno odrediti minimalnu cenu u skupovima stavki koje predstavljaju pravilo, a rezultat je ime proizvoda za koji se može razmatrati popust. Slede rezultati 5.

	1_Item	1_Price	Item_2	2_Price	Item_3	Price	artikal_na_snizenju
1	Cake_0	8.950	Cake_2	15.950	Coffee_46	2.450	Coffee_46
2	Cake_2	15.950	Coffee_46	2.450	Cake_0	8.950	Coffee_46
3	Cake_3	15.950	Danish_35	1.150	Tart_18	3.250	Danish_35
4	Eclair_7	3.500	Pie_11	5.250	Twist_37	1.150	Twist_37
5	Eclair_7	3.500	Twist_37	1.150	Coffee_45	2.150	Twist_37
6	Eclair_7	3.500	Pie_11	5.250	Coffee_45	2.150	Coffee_45
7	Pie_11	5.250	Twist_37	1.150	Eclair_7	3.500	Twist_37
8	Pie_11	5.250	Twist_37	1.150	Coffee_45	2.150	Twist_37
9	Pie_11	5.250	Coffee_45	2.150	Eclair_7	3.500	Coffee_45
10	Tart_16	3.250	Croissant_32	1.450	Coffee_45	2.150	Croissant_32

Slika 5: SPSS rezultat

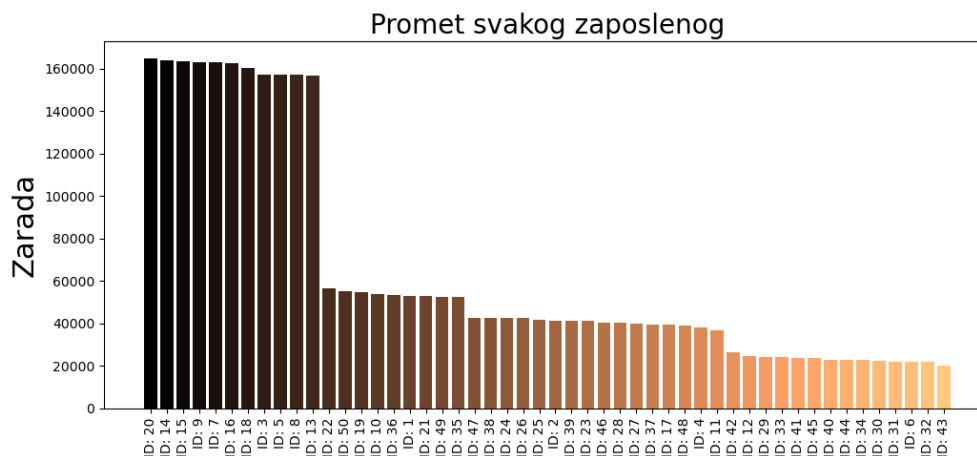
6 Najveći promet

Za dobijanje prodavca koji je imao najveći promet korišćeni su podaci iz tabela koje sadrže informacije o racunima, tabela koje sadrže informacije o artiklima i tabela koje sadrže informacije o prodavcima. Njihovim spajanjem po odgovarajućim atributima pravimo tabelu koja sadrži samo attribute **EmpId**, **Surname**, **Name**, **Food**, **Quantity**, **Price**.

U programskom jeziku **Python** računamo ukupnu zaradu svakog prodavca tako što prvo izračunamo zaradu pri svakoj transakciji množenjem cene proizvoda i kolicine tog proizvoda, a zatim sumiramo sve zasebne transakcije za tog prodavca.

Za dobijanje sortirane tabele prodavaca po njihovim prometima, u **SPSS-u**, pravimo tri čvora. Prvi za računanje zarade pri svakoj transakciji, drugi za njihovo sumiranje za svakog prodavca, a treći za sortiranje podataka po izračunatom atributu koji predstavlja zaradu.

Analizom podataka iz Python-a i SPSS alata, identifikovali smo prodavca broj 20, po imenu 'JERALD' i prezimenu 'DEBRECHT', kao pojedinca koji je ostvario najveći promet.



Slika 6: Grafički prikaz zarade po identifikacionom broju zaposlenog

7 Deo godine sa najvećom zaradom

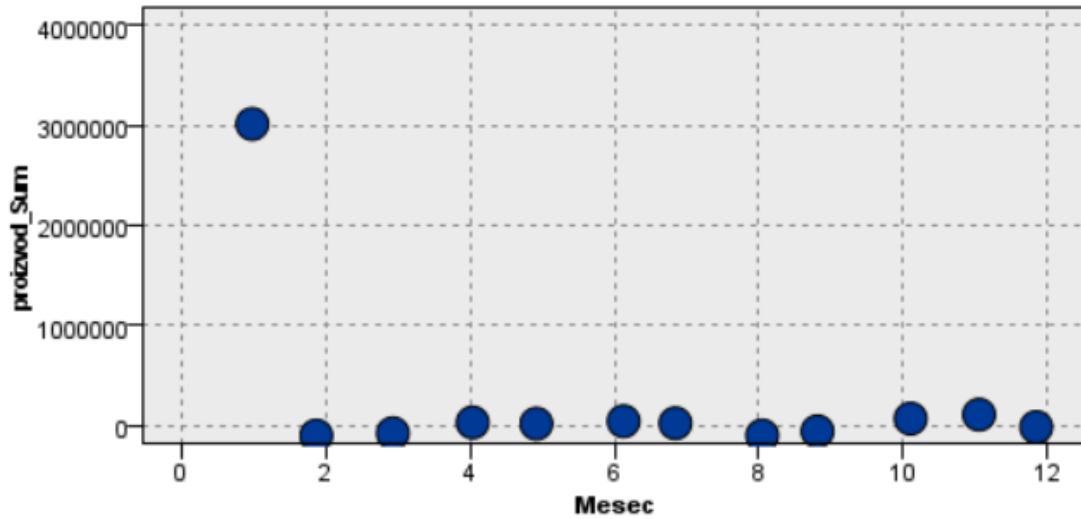
Identifikacija ključnih perioda sa najvećom zaradom predstavlja vitalnu komponentu strategijskog planiranja i donošenja poslovnih odluka. U ovom poglavlju, određujemo deo godine sa najvećom zaradom. Kroz analizu podataka, koristeći alat SPSS Modeler, a zatim i Python, otkrili dinamiku prihoda tokom vremena.

Unapred pripremljeni podaci u fajlu 'prodaja.csv' daju nam uvid o računu, šta je kupljeno, koliko je plaćeno i kada.

	ReceiptNumber	SaleDate	Id	Food	Price	Quantity
1	1	2000-01-12	21	Cookie	1.150	1
2	12	2000-01-06	21	Cookie	1.150	2
3	37	2000-01-16	21	Cookie	1.150	5
4	41	2000-11-02	21	Cookie	1.150	5
5	71	2000-01-06	21	Cookie	1.150	4
6	80	2000-01-28	21	Cookie	1.150	3
7	93	2000-01-25	21	Cookie	1.150	4
8	99	2000-01-24	21	Cookie	1.150	1
9	112	2000-01-23	21	Cookie	1.150	2
10	138	2000-01-28	21	Cookie	1.150	2

Slika 7: Zaglavlje tabele

Nakon transformacija pomoću čvora *Agregate* za svaki račun se čuva mesec u kojem je izvršena transakcija i iznos, a zatim se zarada agregira po mesecu (Slika 8).

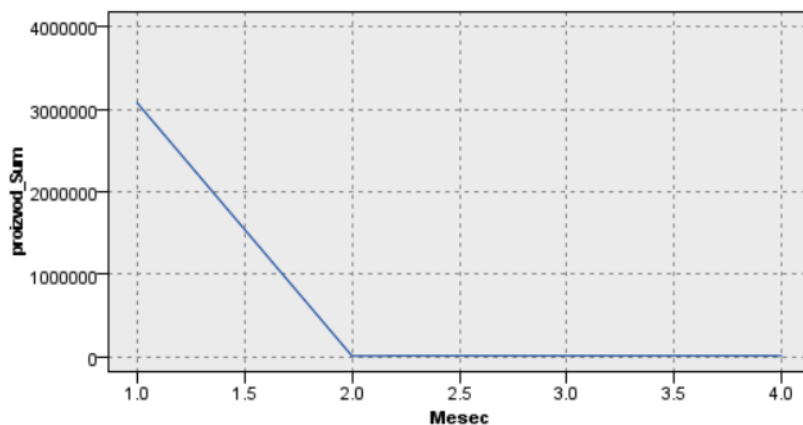


Slika 8: Promet tokom godine

Može se uočiti da zarada značajno opada nakon januara, a zatim malo varira.

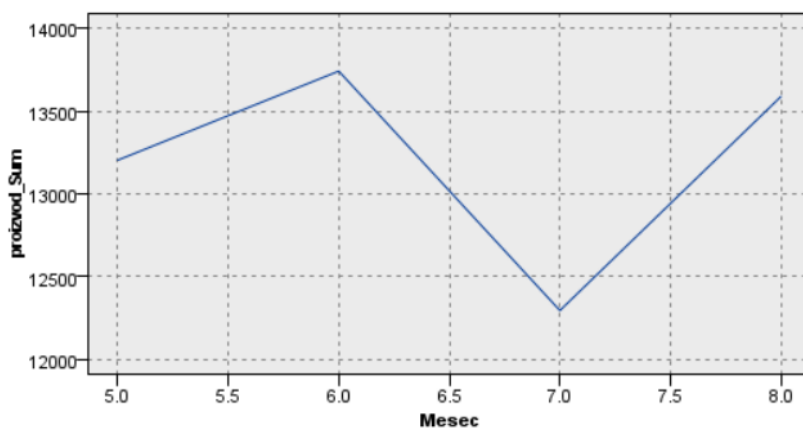
Sledi grafički prikaz zarade u zavisnosti od dela godine. Rezultati su grupisani po mesecima tako da januar,februar,mart i april predstavljaju početak godine, maj,jun,jul i avgust sredinu i preostali meseci u godini kraj godine.

Početak godine se dogodio nagli pad u zaradi i to se održalo do kraja ove trećine godine.



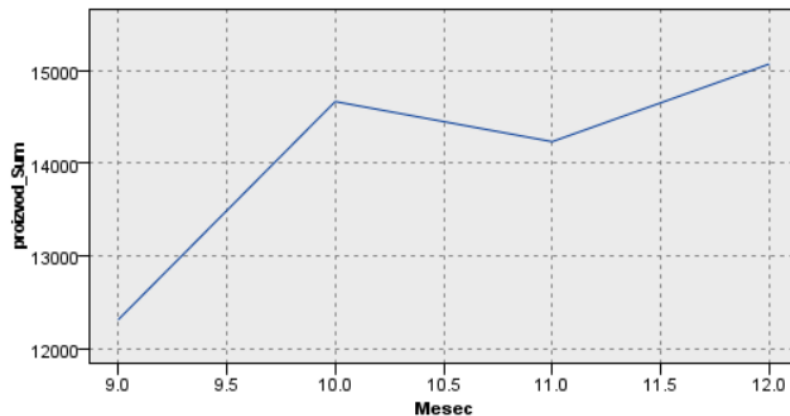
Slika 9: Promet početkom godine

Sredinom godine se izdvaja i povećanje i smanjenje prihoda.



Slika 10: Promet sredinom godine

Konačno, imamo blagi rast profita krajem godine.



Slika 11: Promet krajem godine

Želeli smo da vidimo kako bi nam apriori algoritam i pravila pridruživanja pomogla u ovom zaključku. U analizu uključujemo i informaciju da li je u dan kada je izdat račun bio vikend ili ne (Slika 12).

Opet grupišemo mesece po delu godine, dok sa druge strane odredimo opseg za visinu dnevne zarade. Deo godine i vikend predstavljaju input-e za *Association rules* algoritam, a kategorija zarade je target. Izlaz iz algoritma je pravilo koje nam govori da je u početku godine ostvaren najveći promet, što se i poklapa sa prethodnom analizom.

	mesec	dan	zarada_Sum	vikend
1	1	26	111142.260	False
2	1	15	110795.530	True
3	1	12	110741.450	False
4	1	4	110008.640	False
5	1	20	109107.080	False
6	1	18	108860.130	False
7	1	22	108394.120	True
8	1	21	108037.960	True
9	1	5	107981.160	False
10	1	25	107648.160	False

Slika 12: Zarada po svakom izdatom računu

8 Negativno korelisana pravila pridruživanja

Tipična pravila pridruživanja uzimaju u obzir samo stavke koje su nabrojane u transakcijama. Takva pravila se nazivaju pozitivnim pravilima pridruživanja. Negativno korelisana pravila takođe razmatraju iste stavke, ali pored toga uzimaju u obzir negirane stavke (tj. odsutne iz transakcija). Takva pravila su korisna u analizi prodaje da bi se identifikovali proizvodi koji se međusobno sukobljavaju ili proizvodi koji se međusobno dopunjuju. Primer: „mušterije koje kupuju koka-kolu ne kupuju pepsi” ili „mušterije koji kupuju sojino mleko ne kupuju šunku”. Otkrivanje takvih pravila zahteva ispitivanje eksponencijalno velikog prostora pretrage.

Pored pronalaženja pozitivnih pravila koja imaju jaku korelaciju, odredili smo negativno korelisana pravila pridruživanja sa jakom negativnom korelacijom između i posledice.

Prvi korak u ovom postupku je određivanje podrške za svaku stavku. U nastavku je tabela sa artiklima i njihovim podrškama, dobijena u Python-u. (Slika 12., Slika 13.)

Slika 13: Prvi deo proizvoda

Proizvod	Podrška
Meringue_26	0.04239
Eclair_8	0.04252
Eclair_6	0.04237
Cookie_29	0.06785
Lemonade_40	0.06824
Cake_0	0.08353
Cookie_23	0.06764
Tea_43	0.06247
Coffee_45	0.10267
Water_44	0.0752
Frappuccino_47	0.07747
Cookie_21	0.04324
Espresso_49	0.06797
Cake_1	0.08361
Cake_2	0.07501
Meringue_25	0.04193
Cake_5	0.08224
Croissant_34	0.04324
Cookie_24	0.06801
Coffee_46	0.08315
Lemonade_41	0.06775
Pie_11	0.07712
Tart_15	0.07587
Tart_16	0.08295
Tart_10	0.04204

Slika 14: Drugi deo proizvoda

Proizvod	Podrška
Cake_4	0.09264
Tart_12	0.06823
Cake_9	0.08275
Bear Claw_38	0.04244
Cookie_28	0.10075
Croissant_32	0.08399
Eclair_7	0.10924
Tart_19	0.0758
Twist_37	0.0772
Tart_17	0.07372
Danish_35	0.09257
Soda_48	0.06199
Danish_39	0.04409
Croissant_30	0.04273
Tart_20	0.04337
Danish_36	0.06769
Cake_3	0.08209
Tart_13	0.04236
Tart_14	0.08483
Cookie_22	0.09044
Croissant_33	0.08221
Croissant_31	0.06729
Tart_18	0.09316
Juice_42	0.09161
Cookie_27	0.08977

Drugi korak ovog postupka je određivanje pravila koja imaju jednu negativnu, a ostale pozitivne stavke, tj. čija je podrška ispod, odnosno, iznad zadatog praga. Ovaj uslov je primenjen nad pravilima koja su prethodno izdvojena u SPSS Modeleru. Slede negativno korelisana pravila pridruživanja:

- {Tart_15, Eclair_7} \rightarrow {Espresso_49}
- {Tart_17, Frappuccino_47} \rightarrow {Cookie_29}

9 Zaključak

U ovom radu analizirane su kupovne navike potrošača u lancima pekara, a rezultati pružaju uvid u obrasce ponašanja potrošača. Rezultati uključuju identifikaciju ključnih veza među proizvodima, mogućnosti unapređenja marketinških strategija, optimizaciju asortimana proizvoda i poboljšanje ukupnog korisničkog iskustva.

Dalji rezultati mogu ukazati za potrebu za prilagođavanje lokalnim potrebama, sezonske varijacije u kupovnim navikama ili dublje razumevanje uticaja društvenih faktora na ove obrasce. Ova analiza pravila pridruživanja i analiza dinamike prometa pruža temelj za efikasno upravljanje prodajom i prilagođavanje ponude prema preferencijama potrošača.

Osim sto su otkriveni proizvodi koji se prodaju zajedno, identifikovani su proizvodi čiju cenu treba sniziti, pronađena je sezona sa najvećim prometom, a takođe je utvrđeno koji radnik ostvaruje najveće prihode. Ova saznanja pružaju vredne smernice za strategije sniženja cena, planiranje zaliha i nagrađivanje zaposlenih, unapređujući ukupnu produktivnost i profitabilnost pekarskih lanaca.

U zaključku, istraživanje kupovnih navika igra ključnu ulogu u uspešnom poslovanju u maloprodaji. Rezultati ovakvih analiza bi trebalo poslužiti kao osnova za informisane odluke koje će unaprediti poslovanje i poboljšati zadovoljstvo njihovih kupaca.

10 Literatura

- [1] Slajdovi profesora Nenada Mitića
- [2] Jurnal of big data
- [3] Mining Positive and Negative Association Rules: An Approach for Confined Rules
Maria-Luiza Antonie and Osmar R. Zairane
- [4] IBM SPSS Modeler 18.3 Algorithms Guide
- [5] Python dokumentacija