

Goran Klepac

Metoda potrošačke košarice

odabrano poglavlje iz knjige

«Primjena inteligentnih računalnih metoda u menedžmentu»

Sadržaj:

1. Uvod	2
2. Definicija i opis metode	2
3. Mogućnosti analize podataka kod transakcija obavljenih kreditnim karticama	4
4. Osnovne strategije i metode razvoja aplikacija temeljene na metodi potrošačke košarice	5
5. Mogućnosti primjene metode potrošačke košarice	5
6. Mogućnosti šire primjene metode	14

1. Uvod

Vlasnike velikih maloprodajnih lanaca oduvijek je zanimala struktura potrošača koji kupuju njihovu robu. Često su se provodile klasične općeprihvaćene analize za istraživanje tržišta radi segmentiranja i otkrivanja pojedinih segmenata potrošača. Na te skupine najčešće se utjecalo reklamnim kampanjama ili nekim drugim metodama koje su trebale djelotvornije povećati prodaju robe.

Maloprodaja je donekle specifično područje iz perspektive marketinga jer uglavnom ima standardni set kupaca koji žive u blizini. Ciljevi povećanja prodaje robe trebali bi se usredotočiti na "preotimanje" kupaca ostalim maloprodajnim centrima, odnosno stvaranje potrebe za kupnjom novih proizvoda kod ustaljenih kupaca.

S pojavom kreditnih kartica (posebice u SAD-u i razvijenim europskim zemljama) u tom se segmentu poslovanja pružaju dotad nezamislive mogućnosti u obradi podataka u području rudarenja podataka.

Činjenica je da prilikom plaćanja gotovinom na POS terminalu kupac ostaje anoniman; jedini trag koji ostaje nakon obavljene transakcije jest kupljena roba, njezina količina i cijena te ukupan iznos transakcije.

Plaćajući kreditnom karticom kupac, uvjetno rečeno, "ostavlja" svoje podatke ako imamo pristup do baze podataka kompanije koja izdaje kreditne kartice. Informacije koje možemo dobiti o kupcu iz takve baze podataka mogu biti visina mjesečnih primanja, dob, bračni status, broj djece i niz drugih korisnih podataka.

Ne treba napominjati koliki su troškovi provođenja anketa na nekom segmentu tržišta. Pritom moramo uzeti u obzir i reprezentativni uzorak koji ponekad treba korigirati, što dodatno usložnjuje problem. Dakle, ovo samo po sebi može biti dobra polazna osnova za djelotvorno otkrivanje i segmentiranje kupaca. Redovito, osnovna ideja metode potrošačke košarice temelji se na transakcijama koje se bilježe na POS terminalima te prodana roba po transakcijama.

2. Definicija i opis metode

Ova metoda se najčešće rabi u maloprodajnim centrima, a podloga za obradu podataka su podaci s računa koji su izdani na POS terminalima.

Najčešće se tako spremljeni podaci (oni koji nam trebaju za analizu) nalaze u dvije relacijske tablice. Prva služi za registraciju broja računa, vremena, datuma i sata, dok je druga relacijska tablica povezana relacijskim ključem (šifra računa), a u njoj se nalaze podaci o šiframa proizvoda i količini kupljenih proizvoda po tom računu.

Ovo je opis relacijskih tablica koje su nam zanimljive iz perspektive spominjane metode. Premda u cijelom procesu u trenutku upisa računa sudjeluje još niz tablica s nazivima proizvoda, porezima, cijenama, one za razmatranje ovog problema trenutno nisu relevantne.

Analiza potrošačke košarice otkriva u nizu takvih transakcija skrivena pravila koja se tiču prodaje robe. Cilj metode je otkriti pravila koja nas upozoravaju na sljedeće: ako kupac kupuje proizvod X, tada će kupiti proizvod Y, naravno, uz određenu vjerojatnost temeljenu na povijesnim podacima transakcijske baze. U bazi podataka imamo informacije o izvršenim transakcijama, a svaki račun u bazi predstavlja jednog kupca. Maloprodajni račun također prikazuje set proizvoda koje je taj kupac kupio u određenom maloprodajnom centru.

To se može prikazati sljedećom tablicom.

Kupac	Artikli
1	sol, min. voda
2	mlijeko, sol, detergent
3	sol
4	sol, min. voda
5	detergent, min. voda

Zadatak ove metode je otkrivanje slučajeva kada su kupljeni određeni parovi proizvoda koji po nekoj zdravorazumskoj logici nemaju zajednički nazivnik. Prvi korak u metodi je određivanje matrice pojavnosti tako da za prethodni primjer imamo:

	Sol	Deterdžent	Mlijeko	Min. voda
Sol	4	1	1	1
Deterdžent	1	2	1	0
Mlijeko	1	1	1	0
Min. voda	2	1	0	1

Teoretski možemo vidjeti da u X postotaka slučajeva kupac koji kupuje detergent, kupuje i sol. Također, primjenom ove metode teoretski možemo doći do spoznaje da kupac koji kupuje mineralnu vodu ne kupuje mlijeko.

Ovakve podatke dobivamo ako se ograničavamo na područje proizvoda. Budući da ovdje govorimo o bazama podataka s informacijama o dobavljačima proizvoda i proizvođačima, možemo doći do informacija upravo o proizvođačima ili dobavljačima ili kombinaciji proizvoda s nekom od ovdje spomenutih kategorija. Zanimljivost ove metode je u mogućnosti određenog stupnja slobode u određivanju formiranja pravila s obzirom na vjerojatnosti koje se pojavljuju u transakcijama, a koje su povezane s pojavljivanjima parova proizvoda u transakcijama. Bilo bi neopravdano očekivati da je vjerojatnost kupnje određenih parova proizvoda uvijek 1, što znači da dizajner jednoga takvog sustava određuje donju granicu vjerojatnosti koja utječe na formiranje pravila.

Važno je napomenuti da se ta metoda ne rabi samo za parove proizvoda koji se pojavljuju u transakciji, već i za više od dva proizvoda. Tada proizvode promatramo u tri dimenzije, s obzirom na kombinacije koje se pojavljuju prilikom kupnje. Obilježje je ove metoda da daje rezultate svoje obrade u obliku:

AKO a, ONDA b uz postotak vjerojatnosti

U slučaju višedimenzionalnih analiza oblik generiranog znanja može se svesti na tvrdnje tipa:

AKO b i c, ONDA a (vjerojatnost) ili
AKO b i NE c, ONDA a (vjerojatnost) ili
AKO NE b i NE c, ONDA f (vjerojatnost) i slično.

Oblik generiranog znanja ovisi o dizajnu informacijskog sustava koji obrađuje podatke, odnosno o dubinama analize koje se žele postići.

Prilikom uporabe ove metode dolazimo u situaciju da je ponekad potrebno kategorizirati skupine varijabli, odnosno skupine proizvoda zbog potreba analiza. Neki autori, npr. Gordon i Linoff, taj proces nazivaju dodjelom prividnih varijabli. Primjer uvođenja prividne varijable može biti dodjeljivanje prividne varijable svim onim artiklima koji su proizvedeni od mlijeka, te im dodjeljujemo prividnu varijablu «mliječni proizvodi». Cilj analize može biti otkrivanje koji se proizvodi pojavljuju u kombinaciji kupnje s «mliječnim proizvodima». U tom slučaju rezultat analize mogao bi glasiti:

AKO mliječni proizvodi, ONDA x (vjerojatnost).

Prividne varijable mogu se rabiti u kombinaciji s klasičnim varijablama, a jednako se tako mogu rabiti u kombinaciji s ostalim prividnim varijablama, pa tako rezultat obrade nastao kao proces obrade prividnih varijabli može glasiti:

AKO mliječni proizvodi, ONDA proizvodi proizvođača X (vjerojatnost).

Kod upotrebe prividnih varijabli treba biti oprezan jer njihovom neadekvatnom upotrebom možemo doći do trivijalnih rezultata, koji opterećuju resurse sustava, a ne daju nikakvu korisnu informaciju. Karakteristika metode potrošačke košarice je sklonost ka kombinatornoj eksploziji, koja može uz neoptimizirane algoritme dovesti do zagušenja sustava. Jedna od općeprihvaćenih metoda, koja u osnovi ima zadatak redukcije

kombinatornih vrijednosti, jest tehnika rezanja¹ (*minimum support pruning*) koja omogućava redukciju broja slogova koji ulaze u obradu.

Osnovna ideja te tehnike je sljedeća: ako imamo 1 000 000 transakcija, a odredimo "minimalnu podršku" od 1%, to znači da svaki od elemenata koji analiziramo (prividna varijabla ili proizvod) mora biti sadržan u najmanje 10 000 transakcija, inače ga ne uzimamo u obzir.

Tako, na primjer, u pravilu:

AKO x i y , ONDA z

x mora biti sadržano u najmanje 10 000 transakcija,
 y mora biti sadržano u najmanje 10 000 transakcija i
 z mora biti sadržano u najmanje 10 000 transakcija.

Tehnika rezanja olakšava posao jer u samom začetku smanjuje broj transakcija koje nisu relevantne za obradu pa ne gubimo vrijeme na obradu podataka. Ta metoda je sklona kombinatornoj eksploziji, što ponekad, ako ne definiramo dobro ciljeve, može rezultirati zagušenjem stroja ili generiranjem mnoštva tvrdnji od kojih ne vidimo suštinu.

Algoritam funkcionira tako da u prvom prolazu izračunamo broj pojavljivanja svih proizvoda koji sudjeluju u analizi i omjer između broja pojavljivanja proizvoda u transakcijama te ukupnog broja transakcija. Zatim to isto radimo za kombinacije proizvoda, u dvije ili više dimenzija.

Osnovni cilj metode nije maksimalno smanjiti osnovni uzorak obrade već sve ovisi o problemu koji namjeravamo riješiti. Zato ove metode nastoje optimizirati veličinu uzorka, odnosno izbaciti segmente koji ne zadovoljavaju najnužniji kriterij za ulazak u konačnu obradu podataka.

Postoje dva relevantna pokazatelja u ovoj metodi - stupanj pouzdanosti i podrška. Stupanj pouzdanosti računamo kao

$pouzdanost (A \rightarrow B) = (\text{transakcije koje sadrže } A \text{ i } B) / (\text{transakcije koje sadrže } A)$

Stupanj podrške računamo kao:

$podrška (A \rightarrow B) = (\text{transakcije koje sadrže } A \text{ i } B) / \text{ukupan broj transakcija}$

3. Mogućnosti analize podataka kod transakcija obavljenih kreditnim karticama

Ta je metoda uglavnom orijentirana na transakcije, odnosno prodaju robe nepoznatim kupcima. Ako uzmemo u obzir korisnike kreditnih kartica, te da su nam oni dominantni korisnici u prodavaonicama navedenog tipa, možemo mnogo više saznati o strukturi i potrebama naših kupaca.

Kompanije koje izdaju kreditne kartice imaju informacije o visini primanja nosioca kartice, broju djece, imovini, kreditnoj sposobnosti, mjestu stanovanja i niz drugih korisnih informacija. Kada kupac kupuje robu karticom, ostavlja svoj identifikacijski broj koji ga jednoznačno identificira u bazi podataka korisnika kartica. Ako imamo pristup do te baze podataka, možemo doći do zanimljivih informacija o strukturi kupaca. Tako, na primjer, možemo doznati da kupci u dobi između 20 i 30 godina, s dvoje ili više djece, imaju sklonost kupnji proizvoda tipa X.

Ta je informacija korisna jer možemo slati kataloge s proizvodima tipa X po sniženim cijenama, odnosno donosiocu toga kataloga prodavati po sniženoj cijeni. Rizik prevelikog troška i nezahvaćenosti ciljne skupine je manji. Očekuje se veliki odaziv zbog čimbenika sigurnosti koji je generiran iz transakcijske baze.

¹ Berry, J. A. Michael ; Linoff, Gordon: DATA MINING TECHNIQUES FOR MARKETING SALES AND CUSTOMER SUPPORT, John Wiley & Sons Inc, 1997., str 147.

Interes možemo usmjeriti na kupca i promatrati ga iz perspektive vremena i sklonosti kupnji određenih tipova proizvoda te tako dobiti sekvencijalne vremenske serije. One nam mogu dati odgovore na pitanja nakon što izdvojimo ciljne skupine, segmentiramo prihode od prodaje po svakom određenom kupcu tijekom vremena, odredimo sezonske oscilacije kupnji određenih proizvoda po ciljnim skupinama. Iz baze podataka možemo filtrirati sve one korisnike kartica koji ne posjećuju prodavaonicu te im odrediti zajednički nazivnik. Na osnovi podataka iz baze podataka možemo otkriti zajedničke karakteristike tih kupaca radi razvoja strategije koja bi ih ponovno privukla.

Ovo područje donekle graniči s područjem rudarenja podataka jer ne postoji univerzalni algoritam koji bi razriješio te probleme. Prilikom razrade strategije takvog sustava moramo se dobrim dijelom osloniti na svoju intuiciju.

4. Osnovne strategije i metode razvoja aplikacija temeljene na metodi potrošačke košarice

Kod čiste metode potrošačke košarice postoji jasno koncipiran algoritam koji je sastavni dio nekih aplikacija koje služe za analizu podataka. Kada samostalno razvijamo aplikaciju, trebamo imati na umu zakonitosti koje proizlaze iz algoritma te ga ugraditi u aplikaciju. Tako izgrađenoj aplikaciji prosljeđujemo filtrirane uzorke radi daljnje obrade.

Ta metoda svoje rezultate daje u obliku pravila pa je moguće kanalizirati rezultate obrade u format pravila ekspertnog sustava i uklopiti ga s postojećim pravilima radi pružanja što kvalitetnijih informacija. Ovo je posebice zanimljivo kod sustava za podršku odlučivanju koji zahvaća velik broj organizacijskih jedinica poduzeća, uključujući i ekspertni sustav.

Kada se orijentiramo na korisnike kartica kao promatranu skupinu, situacija je složenija jer ne postoji univerzalni algoritam koji bi u jednom prolazu dao odgovore na sva naša pitanja.

Jednim se dijelom možemo osloniti na klasičan algoritam metode potrošačke košarice, no gotovo se nikada nećemo zaustaviti samo na tome. Pomoćni alat tada mogu biti SQL upiti u traženju odgovarajućih informacija iz baze. Tako, na primjer, pokušavamo saznati koju kombinaciju proizvoda kupuju kupci između 20 i 30 godina koji žive na određenoj lokaciji.

Da bismo to saznali, moramo filtrirati podatke SQL upitima, odnosno izolirati sve one slogove u transakcijskoj bazi koji su nastali kao rezultat kupnje kreditnim karticama i koji zadovoljavaju postavljene kriterije. Nakon što smo obavili taj dio posla, u idućem koraku možemo primijeniti tehniku rezanja radi optimizacije osnovnog skupa slogova koje namjeravamo analizirati. Na tako obrađenim podacima primjenjujemo metodu potrošačke košarice.

U drugom slučaju više dolaze do izražaja tehnike rudarenja podataka.

Sam pristup metodi potrošačke košarice možemo promatrati s dva aspekta. Ako mu pristupamo radi testiranja određene pretpostavke, jasno možemo izdvojiti segment testnih uzoraka te metoda rezanja nije neophodna. Ako je cilj otkrivanje novih zakonitosti, koristimo se tehnikom rezanja, pri čemu ne možemo izdvojiti samo jedan segment uzoraka podataka jer je istraživanje usmjereno na cjelokupni segment baze podataka koji čuva podatke o provedenim transakcijama.

5. Mogućnosti primjene metode potrošačke košarice

Metoda potrošačke košarice je po svojoj interpretaciji u obliku pravila iznimno jasna. Rabimo pojmove vjerojatnosti pomoću kojih možemo predviđati ponašanje potrošača kod kupnje proizvoda. Primjenom te metode možemo dovesti u vezu dvije kategorije ili niz kategorija, odnosno proizvoda koji se mogu kupiti u paru i za koje postoji velika vjerojatnost da će kupac koji kupi artikl X, kupiti i artikl Y.

Tu spoznaju možemo iskoristiti u formiranju odjela u trgovačkim centrima te robu na policama poredati prema rezultatima analize. Kod obrade kupaca s karticama možemo formirati posebne odjele pokušavajući privući njihovu pažnju. Ako na osnovi podataka otkrijemo da su osobe u dobi između 20 i 30 godina sklone kupovati dijetalne proizvode, a oni čine dobar dio prometa u prodavaonici (ovdje je riječ o dvije prividne varijable koje sudjeluju u procesu obrade), možemo oformiti odjel zdrave hrane i slično.

Ovom metodom također možemo smanjiti troškove držanja zaliha robe. Ako prodajemo robu uz koju se najčešće kupuju proizvodi koji ju nadopunjuju, kao što je na primjer kupovanje određenog tipa smrznute hrane s priložima, možemo dobiti informacije o učestalosti kupnje određenih tipova priloga uz osnovnu namirnicu. Na taj način smanjujemo zalihe, odnosno alociramo novčane resurse u priloge za koje postoji veća vjerojatnost kupnje.

Ne smije se zanemariti niti demografski aspekt koji utječe na ponašanje potrošača. Znači da se primjenom te metode mogu uspoređivati trgovine locirane na različitim područjima koje pripadaju istoj kompaniji. Na primjer, uvođenjem virtualnih varijabli teoretski možemo saznati da srijedom većina transakcija sadrži pivo, što naravno možemo povezati s demografskom strukturom potrošača i mogućim prijenosima športskih priredbi srijedom.

Kod korisnika kartica posao nam je bitno jednostavniji ako imamo pristup njihovoj bazi podataka. Tada možemo segmentirati skupine potrošača po određenim kriterijima i jednako tako pratiti pojedince ili skupine kroz vrijeme. Ovakve spoznaje omogućuju nam razradu strategije prema određenoj ciljnoj skupini radi povećanja prodaje.

Moramo također uzeti u obzir da su posjetitelji trgovina na malo uglavnom stalni kupci koji gravitiraju tim trgovinama te prvenstveno moramo djelovati na njih. Želimo prepoznati njihove modele ponašanja, a ta metoda teži upravo tome.

Ta se metoda usredotočuje na prodaju robe. Prilikom analize podataka prvi je korak izdvajanje segmenta baze podataka u kojemu se nalaze podaci o prodaji robe. Metoda je prvenstveno namijenjena obradi maloprodajnih transakcija, iako se može primijeniti i u veleprodajnom poslovanju.

Postoje dvije strategije primjene ove metode:

- pronalaženje skrivenih pravila na razini svih transakcija
- pronalaženje skrivenih pravila na razini dijela transakcija.

Razlika između ta dva pristupa temelji se na cilju koji pokušavamo postići.

U prvom slučaju pokušavamo na kompletnom uzorku podataka prodaje generirati skrivena znanja koja bi nam ukazala na neke pravilnosti, ali se pri tome ne usredotočujemo na određenu skupinu potrošača ili proizvoda. Zbog opsežnosti ove metode i kombinatorne eksplozije koju ona izaziva u procesu obrade podataka, ovaj se pristup rijetko primjenjuje. Prethode mu određene predradnje u obliku metoda eliminacije koje ovise o svrsi provođenja te metode.

U drugom se slučaju usredotočujemo na određeni dio uzorka podataka prodaje radi pronalaženja određenih zakonitosti. Na primjer, ako na osnovi podataka uvidimo da raste prodaja proizvoda X, ovom metodom pokušavamo otkriti postoji li proizvod Y koji se prodaje u paru s proizvodom X i s kolikom vjerojatnošću. Radi povećanja prihoda možemo, na primjer, smanjiti cijenu proizvodu Y za određeni postotak.

U bazi podataka koju ovdje uzimamo kao referentni primjer segment baze podataka u koji se smještaju podaci o maloprodajnim transakcijama sastoji se od tri relacijske tablice koje su, naravno, povezane s ostalim relacijskim tablicama, ali za potrebe ove analize nisu relevantne.

Te tablice jesu: tablica zaglavlja maloprodajnih računa, tablica stavki maloprodajnih računa i tablica poreza. Za potrebe ove analize tablica koja sadrži informacije o porezima također nam nije relevantna.

Strukturu tablica zaglavlja i stavaka računa dajemo u nastavku.

Tablica: RACUN

Ime polja	Tip	Dužina
SIFRA_RAC	Character	15
DATUM	Date	8
VRIJEME	Character	10
SIFRA_BLAG	Character	15
SIFRA_USER	Character	15
STORNIRAN	Character	1
NACIN_PLA	Character	15
SIFRA_KOR	Character	15
SIF_SKL	Character	15
TIP_SKL	Character	2
TIP_KALK	Character	2
STATUS	Character	15
STOL	Character	15
SIFRA_IZV	Character	15
STOL_STAT	Character	1
RABAT_P	Numeric	9,4
SIFRA_KOM	Character	15
SMJENA	Character	1
SIF_VALUTE	Character	3
TIP_NAZIV	Character	30
TIP_IZNOS	Numeric	17,2
GOTOVINSKI	Character	2
ID	Character	15

Tablica: PROMET

Ime polja	Tip	Dužina
SIFRA_RAC	Character	15
SIFRA_ART	Character	15
KOLICINA	Numeric	17,4
NORMATIV	Numeric	17,4
PRODAJNA	Numeric	17,2
MARZA_P	Numeric	9,4
RABAT_P	Numeric	9,4
NABAVNA	Numeric	17,2
NOVA_NAR	Character	1
IZNOS	Numeric	17,2
FLAG	Character	1
PCBEZPOR	Numeric	17,2

Ove dvije relacijske tablice povezane su ključem šifra računa (SIFRA_RAC) po načelu "jedan naprama više", što znači da jedan račun može sadržavati jednu ili više stavki. Ovdje obrađujemo realne podatke poduzeća koje ima dvanaest maloprodajnih organizacijskih jedinica, od kojih ćemo jednu analizirati. Cilj analize je izdvajanje najprodavanijih proizvoda i pronalaženje zakonitosti u njihovoj prodaji u parovima, odnosno osnovni model ove metode jest:

AKO X, ONDA Y s određenim stupnjem pouzdanosti.

Kao alat za obradu podataka poslužit će program koji smo za tu namjenu razvili u programskom jeziku Visual Fox pro.

Program omogućava određivanje stupnja minimalne podrške, a prolazi kroz kompletnu bazu prodaje proizvoda u maloprodajnom centru radi pronalaženja učestalosti pojavljivanja proizvoda unutar transakcija.

Kao temeljna struktura podataka transakcijske baze poslužit će spomenute strukture relacijskih tablica u kojima su podaci o prodaji robe u posljednjih osam mjeseci 1999. godine. Program u prvoj etapi selektira sve one proizvode u zadanome razdoblju koji zadovoljavaju kriterij minimalne podrške. Njega pak zadaje operater u trenutku obrade podataka. Najprije eliminiramo sve one proizvode koji ne zadovoljavaju osnovni kriterij minimalne podrške. Program je izračunao frekvencije pojavnosti pojedinih proizvoda u transakcijama.

U sljedećoj etapi program obavlja proračune o pojavljivanjima robe u parovima i prati zakonitosti algoritma metode potrošačke košarice. Program koji rješava taj problem izgleda ovako:

```
WAIT WIND " Obrada u toku ..." NOWAIT
m.gdje=1

SELECT * FROM ARTIKLI INTO TABLE C:DspomrDARTIKLI1
SELECT RACUN
GO TOP
m.ukupno= RECCOUNT()
SELECT ARTIKLI1
GO TOP
m.auk= RECCOUNT()
GO TOP
DO WHILE NOT EOF()
    IF LASTKEY()=27
        close databases
        QUIT
    ENDIF

    SELECT COUNT(sifra_art) AS transakcij FROM PROMET WHERE ;
ALLTRIM(promet.sifra_art)= ALLTRIM(artikli1.sifra_art) INTO TABLE
C:DspomrDizracunaj

SELECT izracunaj
GO TOP
m.k=(izracunaj.transakcij/m.ukupno)*100

SELECT artikli1
WAIT WIND "Koeficijent = "+ STR(m.k,7,4)+ " za: "+ artikli1.naziv+" Rb = "+
STR(m.gdje,9)+" od "+STR(m.auk,9) NOWAIT
REPLACE artikli1.kalo_p WITH m.k

    IF m.k < thisform.pageframe1.page1.koeficijent.value
        REPLACE artikli1.jm with "****"
    ENDIF
    SKIP 1

    m.gdje = m.gdje + 1
ENDDO
SELECT ARTIKLI1
GO TOP
DELETE ALL FOR artikli1.jm = "****"
PACK

SELECT promet

SELECT * from promet ORDER BY sifra_rac into table C:DspomrDpromet1

SELECT * from artikli1 into table C:DspomrDartikli2
m.h=0

SELECT artikli2
GO TOP
DO WHILE NOT EOF()
```



```
m.prvi=0
m.drugi=0
m.brojac=0
m.racun=""
m.sifra_art=artikli2.sifra_art

SELECT artiklil
GO TOP

DO WHILE NOT EOF()
    IF LASTKEY()=27
        close databases
        QUIT
    ENDIF

*   IF artiklil.sifra_art # m.sifra_art
    SELECT promet1
    GO TOP
    m.racun=promet1.sifra_rac
    m.brojac=0

    DO WHILE NOT EOF()
        WAIT WIND artiklil.sifra_art+" "+m.sifra_art+STR(m.h,9) NOWAIT
        IF LASTKEY()=27
            close databases
            QUIT
        ENDIF

        IF promet1.sifra_art=m.sifra_art
            m.prvi=1
        ENDIF
        IF promet1.sifra_art=artikli1.sifra_art
            m.drugi=1
        ENDIF

        SELECT promet1
        SKIP 1
        IF promet1.sifra_rac # m.racun
            IF (m.prvi=1) AND (m.drugi=1)
                m.brojac=m.brojac+1
            WAIT WIND "+1" NOWAIT
            ENDIF

            m.racun=promet1.sifra_rac
            m.prvi=0
            m.drugi=0
        ENDIF

    ENDDO

SELECT mbar
APPEND BLANK
REPLACE mbar.art2 WITH m.sifra_art
REPLACE mbar.art1 WITH artiklil.sifra_art
REPLACE mbar.brtrans WITH m.brojac
REPLACE mbar.posto WITH m.brojac/m.ukupno*100
*WAIT WIND "POSTO"+m.sifra_art+STR(m.brojac/m.ukupno*100,9,5) NOWAIT
* upisivanje vrijednosti art1,art2, broj

* ENDIF ** Za jednake sifre
IF LASTKEY()=27
```

```

close databases
QUIT
ENDIF

SELECT artikli1
SKIP 1
ENDDO

SELECT artikli2
SKIP 1
m.h=m.h+1
ENDDO

WAIT WIND "Gotovo ..." NOWAIT
thisform.refresh

```

S obzirom na prirodu problema i cilj koji si zadajemo možemo postojeći algoritam nadograditi tako da traži zakonitosti “nepojavljivanja” proizvoda u paru. Također možemo svrstati skupine proizvoda kao prividne varijable po nekom ključu ili dodati u sam algoritam neku “treću dimenziju”.

Nakon obrade na 123121 stavku koje su integralni dio 42632 maloprodajna računa i 2732 proizvoda, dobivamo vjerojatnosti pojavljivanja pojedinih artikala u bazi. Od 2732 proizvoda, nakon obrade prikazanim programom, sa stupnjem minimalne podrške dobivamo uzorak od 89 proizvoda.

Taj uzorak je prevelik za daljnje obrade pa povećanjem stupnja minimalne podrške dobivamo uzorak od 14 artikala. Tablica navedenih uzoraka izgleda ovako:

R.br,	Artikl	Pojavljivanje (%)
1.	PIVO OŽUJSKO 0.5	85,9450
2.	MLIJEKO SVJEŽE 3.2 1L	81,7460
3.	JAMNICA 1/1	60,8230
4.	KRUH BIJELI 0.70 KLAS	58,6880
5.	KRUH BIJELI BRD.0.65	56,0850
6.	KRUH BIJELI DUB.0.70	45,9040
7.	KRUH SEN.POLUBIJELI BRD.0.70	41,7530
8.	VEČERNJI LIST	39,3740
9.	KRUH CRNI 0.70 KLAS	32,1350
10.	CIG.ROHILL LIGHT 100	30,8920
11.	CIG.ROHILL LIGHTS	30,8920
12.	ŠEĆER 1/1	27,8660
13.	KRUH SEN.POLUBIJELI DUB.0.70	27,7730
14.	MLIJEKO TRAJ.1.6 1L	26,9750
15.	MLIJEKO SVJ.3.2 VINDIJA	26,7640

Slijedeći korak, koji se vidi iz programskog koda, odnosi se na računanje vjerojatnosti pojavljivanja parova proizvoda. Zbog opsega baze podataka, zadržat ćemo se na analizi četiri proizvoda: svježe mlijeko, trajno mlijeko, ožujsko pivo i šećer, te ćemo ih analizirati s proizvodima iz prethodne tablice. Rezultate analize pokazuje iduća tablica.

Prvi artikl	Drugi artikl	Podrška %
VEČERNJI LIST	MLIJEKO SVJEŽE 3.2 1L	16,10
JAMNICA 1/1	PIVO OŽUJSKO 0.5	18,10
KRUH SEN.POLUBIJELI DUB.0.70	MLIJEKO SVJEŽE 3.2 1L	6,20
KRUH BIJELI DUB.0.70	MLIJEKO SVJEŽE 3.2 1L	9,40
KRUH SEN.POLUBIJELI BRD.0.70	MLIJEKO SVJEŽE 3.2 1L	8,40
KRUH BIJELI 0.70 KLAS	MLIJEKO SVJEŽE 3.2 1L	10,60
VEČERNJI LIST	PIVO OŽUJSKO 0.5	7,10
KRUH CRNI 0.70 KLAS	MLIJEKO SVJEŽE 3.2 1L	5,40
VEČERNJI LIST	MLIJEKO TRAJ.1.6 1L	5,70
KRUH BIJELI BRD.0.65	MLIJEKO SVJEŽE 3.2 1L	7,80
KRUH BIJELI BRD.0.65	PIVO OŽUJSKO 0.5	6,70
VEČERNJI LIST	ŠEĆER 1/1	3,60
KRUH BIJELI DUB.0.70	MLIJEKO TRAJ.1.6 1L	3,40
KRUH SEN.POLUBIJELI DUB.0.70	PIVO OŽUJSKO 0.5	2,00
KRUH CRNI 0.70 KLAS	MLIJEKO TRAJ.1.6 1L	2,30
KRUH BIJELI 0.70 KLAS	PIVO OŽUJSKO 0.5	4,00
KRUH BIJELI DUB.0.70	PIVO OŽUJSKO 0.5	3,00
CIG.RONHILL LIGHTS	PIVO OŽUJSKO 0.5	2,00
JAMNICA 1/1	MLIJEKO SVJEŽE 3.2 1L	3,80
KRUH SEN.POLUBIJELI DUB.0.70	ŠEĆER 1/1	1,70
MLIJEKO SVJ.3.2 VINDIJA	PIVO OŽUJSKO 0.5	1,60
MLIJEKO SVJ.3.2 VINDIJA	ŠEĆER 1/1	1,40
CIG.RONHILL LIGHTS	MLIJEKO SVJEŽE 3.2 1L	1,50
KRUH BIJELI 0.70 KLAS	MLIJEKO TRAJ.1.6 1L	2,80
KRUH SEN.POLUBIJELI BRD.0.70	ŠEĆER 1/1	1,90
KRUH BIJELI 0.70 KLAS	ŠEĆER 1/1	2,60
JAMNICA 1/1	ŠEĆER 1/1	2,60
KRUH BIJELI DUB.0.70	ŠEĆER 1/1	1,90
MLIJEKO SVIJ.3.2 VINDIJA	MLIJEKO SVJEŽE 3.2 1L	1,10
KRUH SEN.POLUBIJELI BRD.0.7	MLIJEKO TRAJ.1.6 1L	1,60
KRUH SEN.POLUBIJELI DUB.0.70	MLIJEKO TRAJ.1.6 1L	1,00
KRUH CRNI 0.70 KLAS	ŠEĆER 1/1	1,10
KRUH BIJELI BRD.0.65	MLIJEKO TRAJ.1.6 1L	1,90
KRUH BIJELI BRD.0.65	ŠEĆER 1/1	1,90
JAMNICA 1/1	MLIJEKO TRAJ.1.6 1L	1,90
KRUH SEN.POLUBIJELI BRD.0.70	PIVO OŽUJSKO 0.5	1,30
KRUH CRNI 0.70 KLAS	PIVO OŽUJSKO 0.5	0,80
CIG.RONHILL	MLIJEKO TRAJ.1.6 1L	0,70
MLIJEKO SVIJ.3.2 VINDIJA	MLIJEKO TRAJ.1.6 1L	0,40
CIG.RONHILL LIGHTS	ŠEĆER 1/1	0,40
CIG.RONHILL LIGHT 100	PIVO OŽUJSKO 0.5	0,10
CIG.RONHILL LIGHT 100	MLIJEKO SVJEŽE 3.2 1L	0,00
CIG.RONHILL LIGHT 100	MLIJEKO TRAJ.1.6 1L	0,00
CIG.RONHILL LIGHT 100	ŠEĆER 1/1	0,00

Potpora u prethodnoj tablici, iskazana u postocima, izražava vjerojatnost pojavljivanja oba navedena proizvoda u retku. Dobivene su vjerojatnosti pojavljivanja selektiranih proizvoda pojedinačno i u parovima, što je osnova za daljnju analizu. Da bismo došli do krajnjih vrijednosti, morali smo spojiti tablice i proračune, a sam postupak vidi se iz idućega koda:

```
SELECT mbar
REPLACE mbar.naz1 WITH artiklil.naziv
REPLACE mbar.p1 WITH artiklil.posto

SELECT artiklil
GO TOP
LOCATE FOR ALLTRIM(ARTIKLI1.SIFRA_ART)=ALLTRIM(mbar.art2)

SELECT mbar
REPLACE mbar.naz2 WITH artiklil.naziv
REPLACE mbar.p2 WITH artiklil.posto

REPLACE mbar.conf WITH (mbar.posto/mbar.p1)
REPLACE mbar.imp WITH (mbar.posto/mbar.p1*mbar.p1)

SKIP 1
ENDDO

WAIT WIND "Gotovo ..." NOWAIT

select naz1,naz2,posto,p1,p2,conf,imp from mbar order by conf desc into table
c:DspomrDkraj
```

Iz ovoga se koda vidi da su prvo obrađene tablice kao pretpriprema za konačni izračun stupnja pouzdanosti i stupnja poboljšanja, a krajnja operacija programa unutar petlje koja prelazi kroz cijelu osnovnu tablicu podataka jest upravo izračun tih elemenata.

Nakon izvršavanja programa dobivamo rezultate za svježe mlijeko, trajno mlijeko, ožujsko pivo i šećer za koje tražimo vjerojatnosti pojavljivanja u paru s proizvodima izdvojenim u prethodnoj tablici. Kao što vidimo iz posljednjeg SQL upita iz prethodnoga segmenta koda, vrijednosti su poredane opadajućim redoslijedom po stupnju pouzdanosti. U idućoj tablici nalaze se krajnji rezultati analize.

Prvi artikl	Drugi artikl	Pouzdan	Koef. P.
VEČERNJI LIST	MLIJEKO SVJEŽE 3.2 1L	0,4089	16,1000
JAMNICA 1/1	PIVO OŽUJSKO 0.5	0,2976	18,1000
KRUH SEN.POLUBIJELI DUB.0.70	MLIJEKO SVJEŽE 3.2 1L	0,2232	6,2000
KRUH BIJELI DUB.0.70	MLIJEKO SVJEŽE 3.2 1L	0,2048	9,4000
KRUH SEN.POLUBIJELI BRD.0.70	MLIJEKO SVJEŽE 3.2 1L	0,2012	8,4000
KRUH BIJELI 0.70 KLAS	MLIJEKO SVJEŽE 3.2 1L	0,1806	10,6000
VEČERNJI LIST	PIVO OŽUJSKO 0.5	0,1803	7,1000
KRUH CRNI 0.70 KLAS	MLIJEKO SVJEŽE 3.2 1L	0,1680	5,4000
VEČERNJI LIST	MLIJEKO TRAJ.1.6 1L	0,1448	5,7000
KRUH BIJELI BRD.0.65	MLIJEKO SVJEŽE 3.2 1L	0,1391	7,8000
KRUH BIJELI BRD.0.65	PIVO OŽUJSKO 0.5	0,1195	6,7000
VEČERNJI LIST	ŠEĆER 1/1	0,0914	3,6000
KRUH BIJELI DUB.0.70	MLIJEKO TRAJ.1.6 1L	0,0741	3,4000
KRUH SEN.POLUBIJELI DUB.0.70	PIVO OŽUJSKO 0.5	0,0720	2,0000
KRUH CRNI 0.70 KLAS	MLIJEKO TRAJ.1.6 1L	0,0716	2,3000
KRUH BIJELI 0.70 KLAS	PIVO OŽUJSKO 0.5	0,0682	4,0000
KRUH BIJELI DUB.0.70	PIVO OŽUJSKO 0.5	0,0654	3,0000
CIG.RONHILL LIGHTS	PIVO OŽUJSKO 0.5	0,0647	2,0000
JAMNICA 1/1	MLIJEKO SVJEŽE 3.2 1L	0,0625	3,8000
KRUH SEN.POLUBIJELI DUB.0.70	ŠEĆER 1/1	0,0612	1,7000
MLIJEKO SVJ.3.2 VINDIJA	PIVO OŽUJSKO 0.5	0,0598	1,6000
MLIJEKO SVJ.3.2 VINDIJA	ŠEĆER 1/1	0,0523	1,4000
CIG.RONHILL LIGHTS	MLIJEKO SVJEŽE 3.2 1L	0,0486	1,5000
KRUH BIJELI 0.70 KLAS	MLIJEKO TRAJ.1.6 1L	0,0477	2,8000
KRUH SEN.POLUBIJELI BRD.0.70	ŠEĆER 1/1	0,0455	1,9000
KRUH BIJELI 0.70 KLAS	ŠEĆER 1/1	0,0443	2,6000
JAMNICA 1/1	ŠEĆER 1/1	0,0427	2,6000
KRUH BIJELI DUB.0.70	ŠEĆER 1/1	0,0414	1,9000
MLIJEKO SVJ.3.2 VINDIJA	MLIJEKO SVJEŽE 3.2 1L	0,0411	1,1000
KRUH SEN.POLUBIJELI BRD.0.7	MLIJEKO TRAJ.1.6 1L	0,0383	1,6000
KRUH SEN.POLUBIJELI DUB.0.70	MLIJEKO TRAJ.1.6 1L	0,0360	1,0000
KRUH CRNI 0.70 KLAS	ŠEĆER 1/1	0,0342	1,1000
KRUH BIJELI BRD.0.65	MLIJEKO TRAJ.1.6 1L	0,0339	1,9000
KRUH BIJELI BRD.0.65	ŠEĆER 1/1	0,0339	1,9000
JAMNICA 1/1	MLIJEKO TRAJ.1.6 1L	0,0312	1,9000
KRUH SEN.POLUBIJELI BRD.0.70	PIVO OŽUJSKO 0.5	0,0311	1,3000
KRUH CRNI 0.70 KLAS	PIVO OŽUJSKO 0.5	0,0249	0,8000
CIG.RONHILL	MLIJEKO TRAJ.1.6 1L	0,0227	0,7000
MLIJEKO SVJ.3.2 VINDIJA	MLIJEKO TRAJ.1.6 1L	0,0149	0,4000
CIG.RONHILL LIGHTS	ŠEĆER 1/1	0,0129	0,4000
CIG.RONHILL LIGHT 100	PIVO OŽUJSKO 0.5	0,0032	0,1000
CIG.RONHILL LIGHT 100	MLIJEKO SVJEŽE 3.2 1L	0,0000	0,0000
CIG.RONHILL LIGHT 100	MLIJEKO TRAJ.1.6 1L	0,0000	0,0000
CIG.RONHILL LIGHT 100	ŠEĆER 1/1	0,0000	0,0000

Prije oblikovanja krajnjih rezultata, tablica je očišćena od redundantnih elemenata. Sada na osnovi podataka možemo donositi određene zaključke:

- Kupac koji kupuje cigarete Ronhill light 100, sigurno neće kupiti svježe mlijeko u pakiranju od 1l s 3,2 % masnoće ili trajno mlijeko u pakiranju od 1l s 1,6 % masnoće.
- Kupac koji kupuje mineralnu vodu Jamnica, s pouzdanošću 0,4089 kupit će i svježe mlijeko u pakiranju od 1l s 3,2 % masnoće.

Što je koeficijent poboljšanja veći od 1, to je veći i stupanj pouzdanosti. Tako očitavamo podatke iz tablice te nam oni mogu biti polazna osnova za daljnje analize. Na primjer, igra li "marka" ulogu kod proizvoda za koje otkrijemo zakonitosti kupnje u parovima.

Kod ispitivanja osjetljivosti na marku proizvoda promatramo prodaju u parovima na razini prividnih varijabli. Važno je spomenuti da prihvaćanje ili odbacivanje tvrdnje u ovoj metodi ovisi o pragu tolerancije koji filtrira pravila. O pragu tolerancije ovise i daljnji koraci analitičara jer metoda potrošačke košarice katkad samo prethodi nekoj drugoj metodi. Na primjer, ako s određenim stupnjem pouzdanosti uvidimo da kupci koji kupuju proizvod X kupuju i proizvod Y, a jedan od ta dva proizvoda donosi glavninu prihodovnu trgovačkom poduzeću, rezultati koje možemo dobiti nakon provođenja metode potrošačke košarice i uporabom SQL upita mogu biti kako pokazuje sljedeća tablica. (Pritom, poduzeće glavninu transakciju naplaćuje kreditnim karticama i može doći do baze podataka korisnika kartica).

Dobna skupina (godine)	Mjesečni prihodi u kunama	Udio u kupnji (%)
20 - 30	2000 - 3000	15
20 - 30	3000 - 4000	45
30 - 40	2000 - 3000	10
30 - 40	3000 - 4000	20
40 - 50	2000 - 3000	2
40 - 50	3000 - 4000	8

Na osnovi ovih podataka koji su rezultat obrade temeljnih podataka iz transakcijske baze s eksternom bazom podataka korisnika kreditnih kartica, možemo doći do dodatnih informacija koje služe u razradi strategije prodaje ili identifikacije ciljne skupine.

Na primjer, želimo povećati prihode od prodaje u poduzeću u kojemu proizvod X donosi glavninu prihodovnu, a spoznali smo da se prodaje s određenim stupnjem pouzdanosti s proizvodom Y. Kupci toga proizvoda su uglavnom osobe između 20 i 30 godina, s mjesečnim prihodima između 3000 i 4000 kuna. Možemo procijeniti isplati li se smanjiti maržu nekom od tih proizvoda i ići agresivnijom reklamnom kampanjom prema navedenoj ciljnoj skupini. Također možemo slati kataloge pripadnicima definiranog segmenta koji ne kupuju u navedenoj trgovini, žive u blizini i korisnici su kreditnih kartica (pronađemo ih iz baze podataka korisnika kreditnih kartica). Omogućimo, na primjer, dobivanje određenog popusta donosiocu kupona u katalogu za proizvod X ili Y. Tako možemo pridobiti nove kupce te povećati prihode od prodaje robe na dulji rok.

6. Mogućnosti šire primjene metode

Metoda potrošačke košarice može biti snažan alat za otkrivanje zakonitosti ponašanja potrošača u maloprodajnim centrima. Na ponašanje potrošača utječe niz čimbenika specifičnih za pojedine lokacije. Izdvajamo socijalne, demografske i kulturne. Ta metoda objedinjuje sve te čimbenike i usredotočuje se izravno na izvršene transakcije radi generiranja pravila koja ocrtavaju model ponašanja kupaca.

Navedena metoda dobiva na snazi kada se kombinira s drugim metodama ili se rezultati njezine obrade prezentiraju nekima od modela prikaza znanja. Tako, na primjer, integrirajući otkrivene zakonitosti u bazu činjenica ekspertnoga sustava, dobivamo fleksibilan sustav koji objedinjuje "tvrda" pravila (proizišla iz znanja eksperata) i "meka" pravila (proizišla kao rezultat obrade ovom metodom). U slučaju primjene takva modela, s obzirom na karakteristike tržišnog poslovanja, potrebno je osvježavati "meka" pravila periodičnim obradama podataka. Nadalje, tu metodu možemo kombinirati s drugima te rezultate tih metoda možemo preusmjeravati u ovu metodu i obratno. Svaka dodatna mogućnost obrade proizišla iz dodatnih izvora informacija (npr. kreditne kartice) daje novu dimenziju pri pronalaženju zakonitosti unutar transakcijskih baza podataka.

Informacije dobivene djelovanjem te metode menadžeri mogu iskoristiti kao informacije od strateške važnosti jer metoda predviđa ponašanje potrošača. Odluke koje mogu proizići kao rezultat djelovanja te metode mogu se bazirati na agresivnijim reklamnim kampanjama određenih proizvoda usmjerenih točno ciljanom segmentu kupaca, razmještaju polica i zajedničkom izlaganju robe na policama koje su otkrivene kao «komparativne». Jednako tako ta metoda može utjecati na politiku rabata uvjetovanu kupnjom određenog tipa robe u paru s generiranom robom. Jedna od mogućnosti je i izravno obraćanje kupcima koji su identificirani kao ciljna skupina sklona kupnji određenoga tipa robe, s određenim popustima, u svrhu maksimizacije profita. Sa stajališta menadžera postoji čitava paleta mogućnosti koju nam pruža ta metoda. Bitno je napomenuti da ona nema samo usku primjenu u maloprodajnim centrima, nego da ju možemo primijeniti i za veleprodajna poduzeća, banke i ostale institucije, iako je njezina prvenstvena primjena vezana uz maloprodajne centre, i obradu podataka proizišlu kao rezultat transakcija vezanih prvenstveno uz krajnje potrošače na malo.