

# OTKRIVANJE ZNANJA IZ VREMENSKIH SERIJA

## 1. SAŽETAK

Rad izlaže novi pristup u analizi vremenskih serija, koji se oslanja na jedinstveni model transformacije vremenske serije. Prednost ovakvog pristupa očituje se u mogućnosti integracije čitavog niza analitičkih postupaka nad vremenskom serijom, mogućnost ulančavanja postupaka analize nad vremenskom serijom, te integraciju metoda data mininga unutar vremenskih serija, što dosadašnjom poznatom metodologijom nije bio slučaj, upravo iz razloga što se zanemaruje važnost modela transformacije.

Dosadašnji znanstveni radovi upućuju na taj problem i uglavnom rješavanje ovog problema uvjetuju razvojem moćnije hardverske podrške.

REFII model daje cjeloviti rješenje za analizu vremenskih serija, te je posebno pogodan za rješavanja problema iz domene tržišnih analiza koje zahtijevaju integraciju i ulančavanje čitavog niza postupaka i metoda za analizu vremenskih serija.

## 2. UVOD

Jedinstveni model transformacije vremenske serije (REFII) omogućuje povezivanje različitih konceptualnih modela analize vremenske serije, što je unapređenje u odnosu na tradicionalan način korištenja niza nepovezanih metoda prilikom analize.

REFII model također omogućuje analize vremenskih serija primjenom tradicionalnih metoda rudarenja podataka (stabla odlučivanja, klasteriranja, metode potrošačke košarice,...)

Povezivanjem različitih konceptualnih modela analize u okviru REFII modela možemo riješiti složene analitičke zadatke na temelju vremenskih serija, poput segmentacije tržišta te otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata).

Primjena REFII modela nudi unapređenje u otkrivanju tržišnih zakonitosti iz vremenskih serija prilikom "ad hoc" analize tržišnih problema za razliku od tradicionalnoga pristupa, prilikom kojega koristimo niz nepovezanih i nekompatibilnih metoda, koje nam ponekad ne dopuštaju mogućnost dublje analize i modeliranje rješenja za nestandardne probleme povezivanjem niza različitih metodoloških postupaka analize.

Unapređenje pojmovno znači povezivanje različitih metodoloških koncepcija analize vremenskih serija, primjenu tradicionalnih metoda rudarenja podataka na vremensku seriju, ulančavanje metoda za analizu vremenskih serija te nadogradnja temeljnog REFII modela s izvornim algoritmima s ciljem izgradnje modela za segmentaciju tržišta na osnovi vremenskih serija i procjene ponašanja tržišnih segmenata i subjekata.

Dublja analiza pojmovno znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije, koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka.

Povezivanje se ostvaruje posredstvom jedinstvenoga modela transformacije vremenske serije, koji je temelj analitičkim postupcima, metodama te razvoju novih metoda s obzirom na problemski prostor.

Dosadašnje metode, koje se bave analizom vremenskih serija, fokusirane su uglavnom na određeni problem te se ponašaju poput "crnih kutija". Na osnovi ulaznih parametara (vremenska serija) one daju informacije poput postojanja sezonskih oscilacija ili ponavljajućih uzoraka i slično. Za sve njih karakteristično je da imaju vlastite modele transformacije vremenske serije koji su uglavnom prilagođeni konkretnoj problematici kojom se bave te, osim što su međusobno nekompatibilni, ne može ih se ekstrahirati kao izlazne vrijednosti iz modela.

REFII model nudi jedan do sada sasvim novi pristup u analizi.

Novi pristup, koji donosi REFII model, očituje se u modelu transformacije kao osnovi analize, na kojem se temelje sve daljnje analitičke metode.

Analitičke metode su algoritamski postupci kojima se djeluje na transformirane vrijednosti, a ti algoritamski postupci mogu biti izvorni algoritamski postupci koji rješavaju određene tipove problema, poput tržišne segmentacije na temelju vremenskih serija ili pak tradicionalni algoritmi za rudarenje podataka, kao što je to, primjerice, stablo odlučivanja, odnosno novootkrivene metode za analizu vremenskih serija prilagođene REFII sintaksi.

Ovakva orijentiranost na model transformacije kao polazišne točke analize, omogućuje i temeljitiju analizu vremenskih zakonitosti ulančavanjem metoda koje djeluju na vremensku seriju te modeliranje rješenja za nestandardne probleme, proizašle iz otkrivanja tržišnih zakonitosti i segmentacije tržišta.

Pojam temeljitosti u ovom slučaju znači mogućnost provedbe različitih tipova analiza na reduciranom setu podataka iz vremenske serije koji je proizašao kao rezultat obrade prije primijenjene metode ili analitičkoga postupka, s ciljem rješavanja kompleksnih zadataka kao što je to npr. segmentacija tržišta na temelju vremenskih serija.

### 3. KARAKTERISTIKE I CILJEVI REFII MODELA

Osnovna karakteristika ovog modela je jednoznačnost opisa vremenske serije pomoću parametara modela. Matematička jednoznačnost implicira mogućnost provođenja temeljnih matematičkih operacija nad vremenskim odsječcima poput jednakosti, različitosti i sličnosti.

U uvjetima kada krivulju, ili neki njen segment možemo komparirati sa drugom krivuljom odnosno nekim njenim segmentom matematički precizno, tada dolazimo do snažnog instrumentarija na kome se može temeljiti čitav sustav analize vremenskih serija. Upravo na ovim osnovama počiva koncepcija REFII modela.

Ova je koncepcija nastojala zadovoljiti još jedan kriterij, a to je kriterij povezanosti sa algoritmima koji se primjenjuju u data miningu. Do sada poznate metode za analizu vremenskih serija davale su određene pokazatelje koji se kasnije nisu mogli procesuirati posredstvom nekog od poznatih algoritama u cilju ekstrahiranja dodatnog znanja. REFII model teži ka otvorenosti, odnosno uz svoj matematički instrumentarij koji služi kako za opis, tako i za generiranje znanja koje se krije u vremenskoj seriji, daje modalitete rješenja povezivanja sa ostalim data mining algoritmima.

Na taj način možemo iskoristiti snagu provjerenih algoritama na području vremenskih serija, u sklopu standardnih softverskih rješenja.

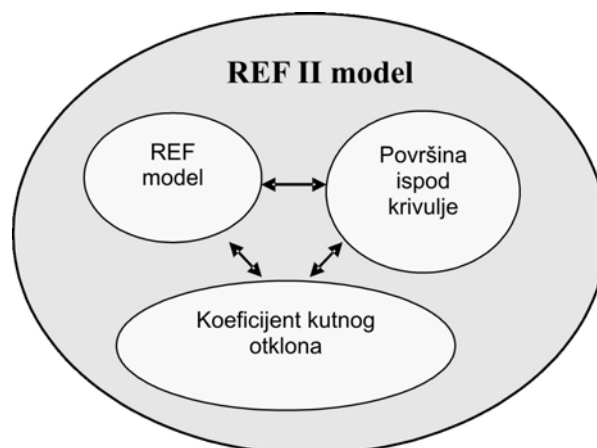
U skupinu klasičnih data mining algoritama ubrajamo neuralne mreže, klasteriranje, stabla odlučivanja, analizu potrošačke košarice, link analizu, i slično, kao i sve mutacije i izvedenice ovih algoritama.

Ovom koncepcijom nam se otvara čitavo jedno novo područje koje nudi detaljniji i precizniji analitički instrumentarij u domeni vremenskih serija.

REFII model se koncentrira se na tri osnovna segmenta kojima se jednoznačno može opisati krivulja, to su :

- oblik krivulje (opis izgleda vremenske serije)
- površina ispod krivulje (kvantifikacija vremenske serije)
- koeficijent kutnog nagiba pravca unutar vremenskog odsječka ("jačina" trenda)

Odnosi ovih triju elemenata koji čine REFII model prikazani su na slici 1.



Slika 1. Struktura REFII modela

#### 4. ALGORITAMSKA INTERPRETACIJA REFII MODELA

Krajnji algoritam mora objediniti sve tri spomenute cjeline i stvoriti temelj za primjenu analitičkih postupaka.

Algoritam za transformaciju vremenske serije u REFII model odvija se u nekoliko koraka

Vremensku seriju možemo deklarirati kao niz vrijednosti  $Vs=(X1,...,Xn)$

##### 1. Korak- Vremenska interpolacija

Formiranje samostalnog vremenskog niza  $V_i$  na intervalu  $<1..n>$  (Dani, tjedni, mjeseci, kvartali, godine) sa vrijednostima 0. Na temelju tako formiranog niza potrebno je provesti interpolaciju nedostajućih vrijednosti u  $Vs$ -u sa 0 na temelju formiranog niza  $V_i$ . Rezultat ovakve obrade je niz  $Vs$  sa interpoliranim vrijednostima niza  $V_i$ .

##### 2. Korak - Vremenska granulacija

U ovom koraku definiramo stupanj sažimanja vremenske serije  $Vs$  koja se nalazi u elementarnoj vremenskoj jedinici (dan, tjedan, mjesec...). U drugom koraku elemente postojeće vremenske serije sažimamo korištenjem statističkih funkcija poput  $AVG()$ ,  $SUM()$ ,  $MOD()$  na razini granularanog odsječka. Na taj način vremensku seriju možemo svesti na veći stupanj granulacije (dani u tjedne, tjedni u mjesec ...), te dobivamo vremensku seriju  $Vg$  sa većim stupnjem granulacije.

Na ovaj korak možemo se vraćati tijekom procesa analize s obzirom na ciljeve analize, što podrazumijeva obavezno ponovno provođenje procesa opisanim u narednim koracima

##### 3. Korak - Normiranje

Postupak normiranja podrazumijeva transformaciju vremenske serije  $Vg$  u  $Ns$  pri čemu je svaki element niza podvrgnut postupku min-max normizacije na intervalu  $<0,1>$  i to

a)  $Ns (Y1,..., Yn)=((Xi- \min(Vs))/(\max(Vs)-\min(Vs)))$

b) mjerilo vremenske kompleksnosti odsječka na X osi  $d(Y_i, Y_{i+1})=0.1$

##### 4. Korak - transformacija u REF notaciju

Prema formuli  $Tr=Y_{i+1}- Y_i$   $Tr > 0 \Rightarrow R$ ;  $Tr < 0 \Rightarrow F$ ;  $Tr=0 \Rightarrow E$  , gdje su  $Y_i$  elementi niza  $Ns$

##### 5. Korak – Proračun nagiba pravca na osnovu kuta

Koeficijent ktnog otklona=>

$Tr > 0 (R)$	Koeficijent $=y2-y1$
$Tr < 0 (F)$	Koeficijent $=y1-y2$
$Tr = 0 (E)$	Koeficijent $= 0$

gdje su  $Y_i$  elementi niza  $Ns$

##### 6. Korak – Proračun Površine ispod krivulje

Numerička integracija metodom pravokutnika

$$p= ((y1*0.1)+(y2*0.1))/2$$

gdje su  $Y_i$  elementi niza  $Ns$

##### 7. Korak - Kreiranje vremenskih indeksa

Građenje hijerarhijskog stabla indeksa ovisno o karakteru analize, gdje element strukturiranog indeksa može biti i atribut poput šifra klijenta

##### 8. Korak - Kreiranje razreda

Kreiranje izvedenih vrijednosti atributa na temelju površine ispod krivulje i otklona kutova

## 9. Korak - Povezivanje tablice transformacije REFII modela sa relacijskim tablicama koje sadrže attribute koji nemaju vremensku dimenziju

Ovih devet osnovnih koraka temelj su algoritmiziranog postupka na kojem se temelji REFII model čiji je krajnji rezultat formiranje matrice transformacije. Matrica transformacije je temelj za provođenje daljnjih analitičkih postupaka s ciljem analize vremenske serije.

## 5. KAKO ANALIZIRATI VREMENSKE SERIJE POSREDSTVOM REFII MODELA

REFII model u osnovi je koncepcija sastavljena iz tri opisane podcjeline, kojoj je prvenstveni cilj transformacija vremenske serije u niz pokazatelja koji jednoznačno definiraju vremensku seriju, REF, površina ispod krivulje, te koeficijent kutnog otklona su pokazatelji koji jednoznačno opisuju odsječak vremenske serije, a niz takvih odsječaka zajedno čine transformiranu vremensku seriju. Ovakav niz pokazatelja poredani su redoslijedom pojavnosti u jedinici vremena i objedinjeni u zajedničku koncepcijsku strukturu transformirane vremenske serije koju nazivamo matricom transformacije.

Takva struktura gledano sa perspektive dinamičke memorije može biti matrica reda  $4 \times tn-1$ , odnosno gledano sa perspektive trajnog zapisa na disku datoteka sa 4 atributa i dužine  $tn-1$ .

Grafički prikaz ove strukture dan je u narednoj tablici:

Tablica I Matrica transformacije

Indeks vremenskog odsječka	I1	I2	I3	In
REF oznaka	REF(I1)	REF(I2)	REF(I3)	REF(I n)
Koeficijent kutnog otklona	Koeficijent kutnog otklona(I1)	Koeficijent kutnog otklona (I2)	Koeficijent kutnog otklona (I3)	Koeficijent kutnog otklona (In)
Površina vremenskog odsječka	P(I1)	P(I2)	P(I3)	P(In)

Pokazatelji vremenskog odsječka izračunati su na temelju koordinata dvije susjedne vrijednosti u vremenskoj seriji. Tako je primjerice odsječak sa indeksom 1 formiran na osnovu vrijednosti koordinata vremenske serije  $t_0$  i  $t_1$ . Indeks vremenskog odsječka služi za jednoznačnu identifikaciju vremenskog odsječka s ciljem njegove analize.

Indeksi mogu biti složeno strukturirani te mogu u sebi sadržavati hijerarhijske elemente, ako i elemente pripadnosti, te vezne elemente prema ostalim izvorima podataka.

Elementi obuhvaćeni u prethodnoj tablici temeljni su elementi REFII modela, sa kojima je moguće jednoznačno opisati krivulju i izvršiti sve analize zbog kojih je i ovaj model i razvijen. Osim opisanih pokazatelja moguće je obuhvatiti i izvedene pokazatelje prikazane kroz razvoj modela, ali to je opcionalni pristup koji ovisi o karakteru analize. Pokazatelji navedeni u prethodnoj tablici temelj su za sve analize zbog kojih je ovaj model razvijen.

Nakon transformacije vremenske serije posredstvom REFII modela dobijemo transformiranu vremensku seriju u obliku prikazanu u tablici.

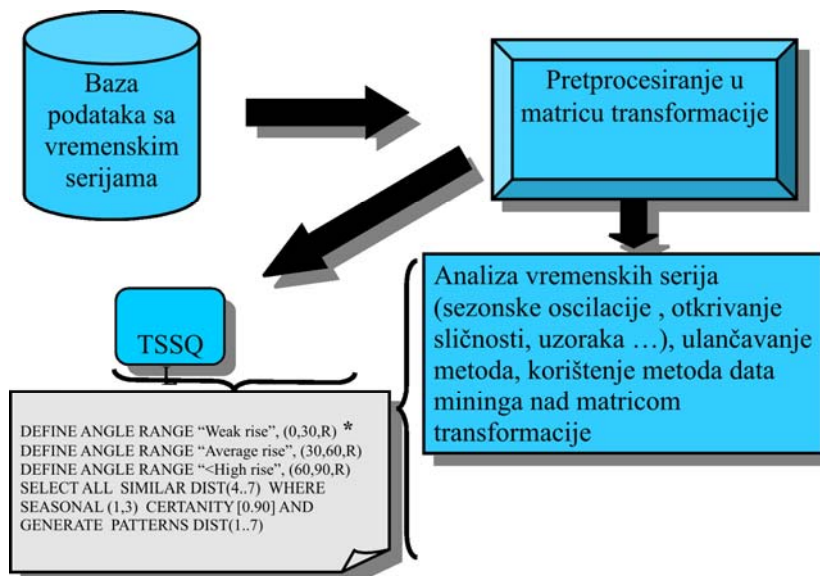
OVAKO transformirani podaci procesuiraju se algoritmiziranim metodama s ciljem rješavanja konkretnih problema iz domene vremenskih serija. Algoritmizirane metode i postupci koje služe za rješavanje konkretnih problema spadaju u širi pojam REFII modela

O samom karakteru analize ovisi kako će se tako transformiranoj vremenskoj seriji pristupiti po pitanju "lomljenja" ovako formirane strukture u manje logičke cjeline. Ako na primjer promatramo tjedne, a vremenska serija sadrži podatke za svaki dan u tjednu kroz cijelu godinu, tada logički razlomimo seriju na tjedne i analiziramo vremenske odsječke posredstvom konkretnog algoritma. Ovaj postupak "lomljenja" vremenske serije na manje analitički usporedive logičke odsječke vezemo uz pojam vremenske kompleksnosti. Ova mjera određuje točku logičkog loma vremenske serije s obzirom na cilj analize, a dio je algoritmiziranog postupka analize. Tako primjerice ovaj koeficijent za tjedan može imati vrijednost 7, ili 5 (radni dani u tjednu), pri čemu treba voditi računa o nepostojanju vrijednosti u određenoj vremenskoj točki, sa čim se ova mjera također mora nositi.

Osnovni zaključak je da je REFII model u užem smislu, model transformacije vremenske serije, dok je on u širem smislu skup algoritmiziranih postupaka nad tako transformiranim podacima koji primjerice omogućuju:

- Otkrivanje sezonskih oscilacija
- Otkrivanje relevantnosti atributa
- Otkrivanje korelatornih odnosa varijabli 1:1 i 1:n
- Predviđanje trenda vremenskih serija
- Otkrivanje epizoda u vremenskim serijama
- Grupiranje vremenskih odsječaka
- Procjene "sličnosti" vremenskih odsječaka i vremenskih serija
- Otkrivanje inverznih oscilacija
- Otkrivanje modela ponašanja klijenata
- Implementacija "teorije spavaća" kroz vremenske serije
- Temelj upitnog jezika za analizu vremenskih serija

Postoji čitav niz mogućih primjena. Ovaj transformacijski model, univerzalna je polazišna točka za sve moguće vrste analiza koje se provode nad vremenskim serijama, a koje ova koncepcija uspješno rješava. Kao što će biti vidljivo kroz daljnji tekst elementi prikazani u tablici proceduralno procesuirani kroz različite vrste algoritama mogu otkriti različite vrste znanja iz vremenskih serija, i upravo u tome leži snaga REFII koncepcije. Ovaj koncept može biti uspješno primijenjen za razvoj upitnog jezika za vremenske serije čija je struktura prikana na slici 2.



Slika 2. Struktura upitnog jezika za vremenske serije baziranog na REFII modelu

## 6. PRIMJER- IZRAVNO OTKRIVANJE PRAVILA IZ VREMENSKIH SERIJA

Jedna od karakteristika REFII modela je mogućnost direktnog otkrivanja pravila iz vremenskih serija, pri čemu elementi koji ulaze u analizu mogu, ali i ne moraju nužno imati temporalni karakter [Ohsaki, 2003].

Analitičara može interesirati odnos međuzavisnosti među pojavama koje su temporalnog i netemporalnog karaktera. Isto tako, za analizu nam može biti interesantno da li postoji pravilnost između pojave  $x$  koja se desila u vremenu  $t$  i pojave  $y$  koja se je desila u vremenu  $t+1$ , odnosno ( $t+$  procijenjeni prag tolerancije) .

Predmet istraživanja ne moraju nužno biti vremenski odsječci, već to mogu biti i vremenski uzorci, epizode, te nevremenske komponente.

Kada se govori o nevremenskim komponentama kao elementima koji ulaze u proces analize, tada u tu skupinu ulaze atributi poput regije, dobi, spola, pripadnosti određenoj grupi i slično.

Na taj način rezultati analize mogu dati informaciju da ako postoji rastući trend u nekom vremenskom periodu, da je najveći utjecaj na taj trend imala grupa atributa čija je vrijednost atributa dobi u nekom određenom razdoblju.

REFII model ne nudi svoje vlastito rješenje za direktno otkrivanje pravila iz vremenskih serija, već je jedno od mogućih rješenja u ostvarivanju tog cilja korištenje asocijativnih algoritama nad vremenskom serijom/serijama koje su transformirane u REFII notaciju.

Primjenom asocijativnih algoritama nad jednom transformiranom vremenskom serijom, koja u sebi sadrži vremenske indekse možemo otkrivati sezonske oscilacije. Ovo je alternativni način traženja sezonskih oscilacija unutar vremenskih serija transformiranih u REFII notaciju.

Ako u vremensku seriju uvedemo atribut koji nije zavisen u vremenu, poput regije, pripadnosti nekoj interesnoj skupini i slično, tada ti atributi postaju ravnopravni vremenskim atributima i sudjeluju u procesu analize.

Ovaj primjer ilustrira način kako izravno otkrivati pravila povezivanjem REFII modela i asocijativnih algoritama kao ilustraciju mogućnosti direktnog povezivanja elemenata vremenske serije preko REFII modela sa algoritmima data mininga.

Kao predmet istraživanja uzeti su podaci sa :

[http://www.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www.stat.duke.edu/~mw/ts_data_sets.html)

koji prikazuju indikatore proizvodnje na mjesečnoj razini u periodu od 1947-1993.

Izvor: Federal Reserve Statistical Release G.17

Struktura podataka : vremenske serije za period od 1947-1993

YR = Godina  
MN = Mjesec  
IP = Indeks industrijske proizvodnje  
MFG = Prerađivačka industrija  
MFGD = Trajna dobra  
MFGN = Netrajna dobra  
MIN = Rudarstvo  
UTIL = Javne usluge  
P = Proizvodi ukupno  
MAT = Materijali

Svaka vremenska serija ima 564 elementa.

Cilj analize je otkriti pravilnosti u kakvoj su međuzavisnosti indeks industrijske proizvodnje, javne usluge, rudarstvo, i proizvodnja materijala.

Prilikom transformacije originalnih vrijednosti vremenske serije korištena je naredna tablica klasifikacije otklona kutova vremenskih odsječaka:

Tablica II Klasifikacije otklona kutova

Razred	Donja granica	Gornja granica	REF
Nizak rast	0.000000000	0.300000000	R
Srednji rast	0.300000000	0.700000000	R
Visok rast	0.700000000	1.000000000	R
Nizak pad	0.000000000	0.300000000	F
Srednji pad	0.300000000	0.700000000	F
Oštar pad	0.700000000	1.000000000	F
Bez promjene	0.000000000	0.000000000	E

Nakon provedene transformacije četiri vremenske serije svaka duljine 562 elementa, vremenske serije su spojene primjenom SQL upita na temelju vremenskih indeksa, te je rezultat obrade eksportiran u tablicu formata :

Tablica III Tablica transformacije u REFII model za potrebe izravnog otkrivanja pravila

ciklus	indeks	ip	ut	min	mat
discrete	discrete	discrete	discrete	discrete	discrete
47	2	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	3	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	4	Nizak pad	Nizak rast	Nizak pad	Nizak pad
47	5	Nizak rast	Nizak rast	Nizak rast	Nizak rast
47	6	Bez promjene	Bez promjene	Nizak pad	Nizak pad
47	7	Nizak pad	Nizak rast	Nizak rast	Nizak pad
...	...	...	...	...	...

Ovaj format tablice je standardni format podataka u dana mining modulu, open source- Orange, <http://magix.fri.uni-lj.si/orange/default.asp> za programski jezik Python [www.python.org](http://www.python.org) razvijenom na Ljubljanskem univerzitetu, katedra za umjetnu inteligenciju.

Na temelju tog modula kreiran je program za otkrivanje asocijativnih pravila u iz REFII modela.

U nastavku je prikazan izvorni kod pisan u programskom jeziku Python sa korištenjem Orange modula za otkrivanje asocijativnih pravila na temelju transformiranih vremenskih serija:

```
import orange, orngAssoc
data = orange.ExampleTable("serije.tab")
minSupport = 0.3
rules = orngAssoc.build(data, minSupport)
print "%i pravila sa support koeficijentom vecim ili jednakim %5.3f .\n" % (len(rules), minSupport)
subset = rules[0:100]
subset.printMeasures(['support', 'confidence'])
print
del subset[0:2]
subset.printMeasures(['support', 'confidence'])
```

Uz zahtijev da support koeficijent bude veći ili jednak od 0.3 dobiveno je 17 pravila :

supp.	conf.	Pravilo
0.521	0.891	mat=Nizak rast -> ip=Nizak rast
0.521	0.880	ip=Nizak rast -> mat=Nizak rast
0.425	0.718	ip=Nizak rast -> ut=Nizak rast
0.425	0.664	ut=Nizak rast -> ip=Nizak rast
0.415	0.708	mat=Nizak rast -> ut=Nizak rast
0.415	0.647	ut=Nizak rast -> mat=Nizak rast
0.377	0.910	ut=Nizak rast mat=Nizak rast -> ip=Nizak rast
0.377	0.887	ut=Nizak rast ip=Nizak rast -> mat=Nizak rast
0.377	0.724	mat=Nizak rast ip=Nizak rast -> ut=Nizak rast
0.377	0.644	mat=Nizak rast -> ut=Nizak rast ip=Nizak rast
0.377	0.637	ip=Nizak rast -> ut=Nizak rast mat=Nizak rast
0.377	0.589	ut=Nizak rast -> mat=Nizak rast ip=Nizak rast
0.319	0.746	min=Nizak rast -> mat=Nizak rast
0.319	0.544	mat=Nizak rast -> min=Nizak rast
0.313	0.733	min=Nizak rast -> ip=Nizak rast
0.313	0.529	ip=Nizak rast -> min=Nizak rast
0.302	0.708	min=Nizak rast -> ut=Nizak rast
0.425	0.718	ip=Nizak rast -> ut=Nizak rast
0.425	0.664	ut=Nizak rast -> ip=Nizak rast
0.415	0.708	mat=Nizak rast -> ut=Nizak rast
0.415	0.647	ut=Nizak rast -> mat=Nizak rast
0.377	0.910	ut=Nizak rast mat=Nizak rast -> ip=Nizak rast

0.377	0.887	ut=Nizak rast ip=Nizak rast -> mat=Nizak rast
0.377	0.724	mat=Nizak rast ip=Nizak rast -> ut=Nizak rast
0.377	0.644	mat=Nizak rast -> ut=Nizak rast ip=Nizak rast
0.377	0.637	ip=Nizak rast -> ut=Nizak rast mat=Nizak rast
0.377	0.589	ut=Nizak rast -> mat=Nizak rast ip=Nizak rast
0.319	0.746	min=Nizak rast -> mat=Nizak rast
0.319	0.544	mat=Nizak rast -> min=Nizak rast
0.313	0.733	min=Nizak rast -> ip=Nizak rast
0.313	0.529	ip=Nizak rast -> min=Nizak rast
0.302	0.708	min=Nizak rast -> ut=Nizak rast

Primjerice pravilo  $ip=Nizak\ rast \rightarrow mat=Nizak\ rast$ , koje govori da ako je ukupan indeks industrijske proizvodnje u niskom rastu, tada je i proizvodnja materijala u niskom rastu ima vrijednost support koeficijenta 0.521 i visoku pouzdanost od 0.891.

Iz ovog je primjera vidljiv način na koji REFII model otvara mogućnost primjene data mining algoritma nad elementima vremenske serije.

Na spomenuti set podataka moguće je djelovati i ostalim algoritmima poput stabla odlučivanja, ili pak procjenjivati relevantnosti atributa temeljem Gini indeksa s obzirom na ciljnu varijablu primjerice varijablu "godina". Asocijativni algoritmi su samo jedan od primjera kako iz vremenskih nakon transformacije u REFII notacije možemo otkrivati pravila primjenom data mining algoritama.

U fokusu analize ne moraju biti samo odsječci sa istim vremenskim indeksom. Ako želimo pratiti analizu utjecaja, tada možemo upariti pojavom na vremenskoj razini  $t$ , sa pojavom na vremenskoj razini  $t+n$  za koju sumnjamo da bi mogle imati međuzavisni utjecaj sa vremenskim pomakom, te tada na njih djelujemo sa asocijativnim algoritmima.

Prednost ovakvog pristupa očituje se i u mogućnosti detaljnijih analiza nad izdvojenim skupom podataka, gdje obuhvaćamo vrijednosti atributa otklona sa zadovoljavajućim vrijednostima koeficijenata, te im pridružujemo i dodatne attribute s ciljem dodatne analize.

Primjer za to može biti izdvajanje dijela vremenske serije koji zadovoljava kriterije da je  $ip="Nizak\ rast"$  i  $mat="Nizak\ rast"$ , pri čemu se izdvajaju i ostale vrijednosti za  $ut$ ,  $min$ , ciklus i indeks i pokušavaju naći zakonitosti nad tim podskupom podataka. Kada je vremenska serija pretprocesirana na spomenuti način, tada je izdvajanje spomenutog podskupa moguće realizirati i primjenom jednostavnih SQL upita na vremensku seriju.

## 7. ZAKLJUČAK

REFII model konstruiran je s ciljem objedinjavanja različitih koncepcija analize vremenskih serija, tradicionalnih metoda rudarenja podataka te s ciljem konstrukcije novih algoritamskih postupaka iz oblasti analize tržišta, te pronalaženju rješenja problema automatskog pretprocesiranja vremenskih serija.

Važnost problema razjedinjenosti metodoloških postupaka analiza vremenskih serija uočio je Graham Williams u svom radu [Williams, 2002] koji daje pregled data mining metoda u domeni vremenskih serija. U istom radu naveo je izazove budućih istraživanja (eng. challenge questions) gdje bi trebalo riješiti problem generalne teorije analize vremenskih serija koja bi objedinila sadašnja i buduća istraživanja na području data mining analize vremenskih serija.

Osnovni nedostatak spomenutog Williamsovog modela proizlazi iz činjenice što je premala pažnja usmjerena na model transformacije, te je on stavljen u drugi plan prilikom definicije rješenja.

Kao implikacija ovakvog zapostavljanja modela transformacije pojavljuju se problemi:

Svi problemi proizlaze zbog nedovoljno dorađenog i nefleksibilnog modela transformacije vremenske serije koji se u hodu dorađuje s obzirom na analitičke potrebe [Williams, 2000], [Williams, 2001], [Williams, 2001], [Williams, 2001a], [Williams, 2002a], [Williams, 2003], [Williams, 2003a]

Zanemarivanje kontingencijskog pristupa analizi

Nije predviđena metodologija povezivanje temporalnih atributa sa netemporalnim

Kao aktivnosti budućih istraživanja napominje se prilagođavanje i dorada postojećih data mining algoritama, umjesto da se putem modela transformacije izgrade mostovi za spajanje, što naravno implicira i generalni problem sa primjenom fuzzy logike unutar modela



REFII model ima odgovore na probleme koji su navedeni kao rezultati budućih istraživanja u Williamsovom radu, te nudi puno fleksibilniji pristup analizi vremenskih serija od one koju je zamislio spomenuti autor.

Jedinstveni model transformacije vremenske serije (REFII) omogućuje povezivanje različitih konceptualnih modela analize vremenske serije, što je unapređenje u odnosu na tradicionalan način korištenja niza nepovezanih metoda prilikom analize.

REFII model također omogućuje analize vremenskih serija primjenom tradicionalnih metoda rudarenja podataka (stabla odlučivanja, klasteriranja, metode potrošačke košarice,...)

Povezivanjem različitih konceptualnih modela analize u okviru REFII modela možemo riješiti složene analitičke zadatke na temelju vremenskih serija, poput segmentacije tržišta te otkrivanja tržišnih zakonitosti (modeli ponašanja tržišnih segmenata i subjekata) .

Primjena REFII modela nudi unapređenje u otkrivanju tržišnih zakonitosti iz vremenskih serija prilikom "ad hoc" analize tržišnih problema za razliku od tradicionalnoga pristupa, prilikom kojega koristimo niz nepovezanih i nekompatibilnih metoda, koje nam ponekad ne dopuštaju mogućnost dublje analize i modeliranje rješenja za nestandardne probleme povezivanjem niza različitih metodoloških postupaka analize.

Ove mogućnosti su do sada tradicionalnim pristupima analizi vremenskih serija zbog primjene niza nepovezanih metoda (nepostojanja generalne strategije povezivanja) bile nemoguće ili vrlo teško izvedive, a pojavljuju se kao nužnost u analizi tržišnih zakonitosti putem vremenskih serija.

**mr.sc. Goran Klepac**

**v.d. direktor BI direkcije**

**Raiffeisen consulting d.o.o.**

**e-mail: goran.klepac@rba-zagreb.raiffeisen.at**

## **POPIS LITERATURE**

[Debregeas, 1998] Debregeas, A. & Hebrail, G. (1998). Interactive interpretation of kohonen maps applied to curves. In proceedings of the 4th Int'l Conference of Knowledge Discovery and Data Mining. New York, NY, Aug 27-31. pp 179-183.

[Klepac, 2000] Klepac Goran, Otkrivanje zakonitosti iz perspektive poduzeća primjenom metoda umjetne inteligencije, Ekonomski fakultet Zagreb, magistarski rad, 2000

[Klepac, 2001] Klepac Goran, Primjena inteligentnih računalnih metoda u menadžmentu, Sinergija, 2001

[Klepac, 2003] Klepac Goran, Panian Željko, Poslovna inteligencija, Masmedia, 2003

[Ohsaki, 2003] Ohsaki, M., Sato, Y., Yokoi, H. & Yamaguchi, T. (2003). A Rule Discovery Support System for Sequential Medical Data In the Case Study of a Chronic Hepatitis Dataset. ECML 2003.

[Williams, 2000] Williams, J. Graham, Weiqiang Lin, Mehmet Orgun, Temporal Data Mining using Multi-Level Local Polynomial Models, Proceedings of the 2nd International Conference on Intelligent Data Engineering and Automated Learning (IDEAL00)Lecture Notes in Computer Science, Volume 1983, Springer Hong Kong, December 2000

[Williams, 2001] Williams, J. Graham, Weiqian Lin, Mehmet Orgun, Temporal Data Mining Using Hidden Markov-Local Polynomial Models Edited by David Cheung, Graham Williams, Qing Li, Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001 Proceedings of the 5th Pacific Asia Conference on Knowledge Discovery and Data Mining(PAKDD01)Hong Kong, China, April 2001

[Williams, 2001a ] Williams, J. Graham, Rohan Baxter, Graham Williams, Hongxing He, Feature Selection for Temporal Health Records, Advances in Knowledge Discovery and Data Mining Edited by

David Cheung, Graham Williams, Qing Li Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001 Proceedings of the 5th Pacific Asia conference on Knowledge Discovery and Data Mining (PAKDD01) Hong Kong, China, April 2001

[Williams, 2002] Williams, J. Graham , Weiqiang Lin, Mehmet A Orgun, An Overview of Temporal Data Mining, Proceedings of the 1st Australian Data Mining Workshop (ADM02) Canberra, Australia, December 2002 Edited by Simeon J. Simoff, Graham J. Williams, Markus Hegland Published by University of Technology, Sydney, Pages 83-90, ISBN 0-9750075-0-5

[Williams, 2002a ] Williams, J. Graham, Weiqiang Lin, Mehmet Orgun, Mining Temporal Patterns from Health Care Data , Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery (DaWaK02) Lecture Notes in Computer Science, Vol 2454, Springer, 2002 Pages 221-231, ISBN 3-540-44123-9

[Williams, 2003] Williams, J. Graham, Mining the Data Stream, Invited Plenary, International Conference on Hybrid Intelligent Systems Melbourne, Australia, December 2003.

[Williams, 2003a] Williams, J. Graham, Chris Kelman, Rohan Baxter, Lifang.Gu, Simon Hawkins, Hongxing He, Chris Rainsford, Deanne Vickers, Temporal Event Mining of Linked Medical Claims Data , Proceedings of the PAKDD03 Workshop on Data Mining for Actionable Knowledge DMAK-2003 Seoul, Korea, April