



INFORMACIJSKA PODPORA NAPOVEDOVANJU ZNAČILNOSTI PROMETA NA PODLAGI PODATKOV ČASOVNIH SERIJ

Andrej Sevčnikar

15.4.2014



VSEBINA PREDSTAVITVE

- MOTIVACIJA, DOSEDANJE REŠITVE, TEZA
- PRIPRAVA PODATKOV
- DOLOČITEV ČASOVNIH OKVIRJEV
- NAPOVEDOVANJE PODATKOV in REZULTATI
- ZAKLJUČEK



VSEBINA PREDSTAVITVE

- **MOTIVACIJA, DOSEDANJE REŠITVE, TEZA**
- PRIPRAVA PODATKOV
- DOLOČITEV ČASOVNIH OKVIRJEV
- NAPOVEDOVANJE PODATKOV in REZULTATI
- ZAKLJUČEK



MOTIVACIJA

- Promet in logistika vpeta v vsakdanje življenje
- Vpliv na vsakdanji urnik in časovne zamude
- Odvisnost od prometa
- Osrednja tematika → napovedovanje prometa z uporabo tehnologij podatkovnega rudarjenja



DOSEDANJE REŠITVE

- Primerjava algoritmov
- Izboljšava algoritmov
- Hibridne rešitve

- Manj poudarka na pripravi podatkov



PRISTOP K REŠITVI

- PRIPRAVA PODATKOVNE STRUKTURE in INTEGRACIJA PODATKOV
 - Podatkovno skladišče
 - Proces ETL
- GRUČENJE in NAPOVEDOVANJE PODATKOV
 - Podatkovno rudarjenje
 - Obstoječe metode
 - Razvoj novih metod

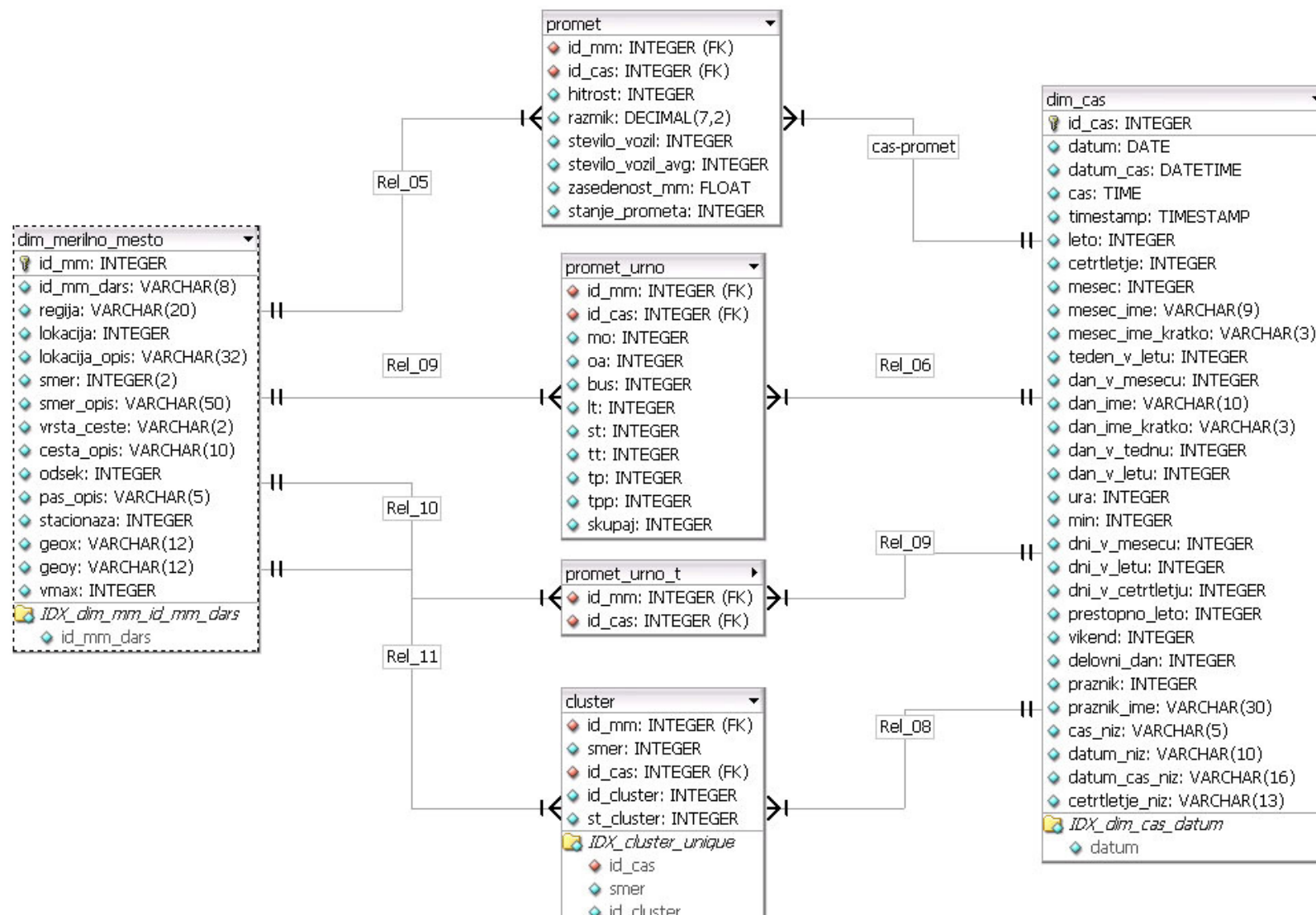


VSEBINA PREDSTAVITVE

- MOTIVACIJA, DOSEDANJE REŠITVE, TEZA
- **PRIPRAVA PODATKOV**
- DOLOČITEV ČASOVNIH OKVIRJEV
- NAPOVEDOVANJE PODATKOV in REZULTATI
- ZAKLJUČEK



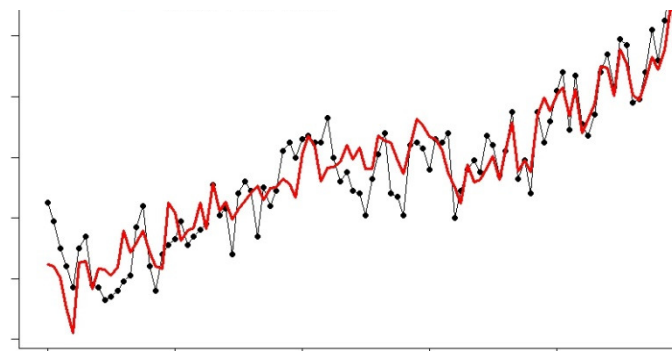
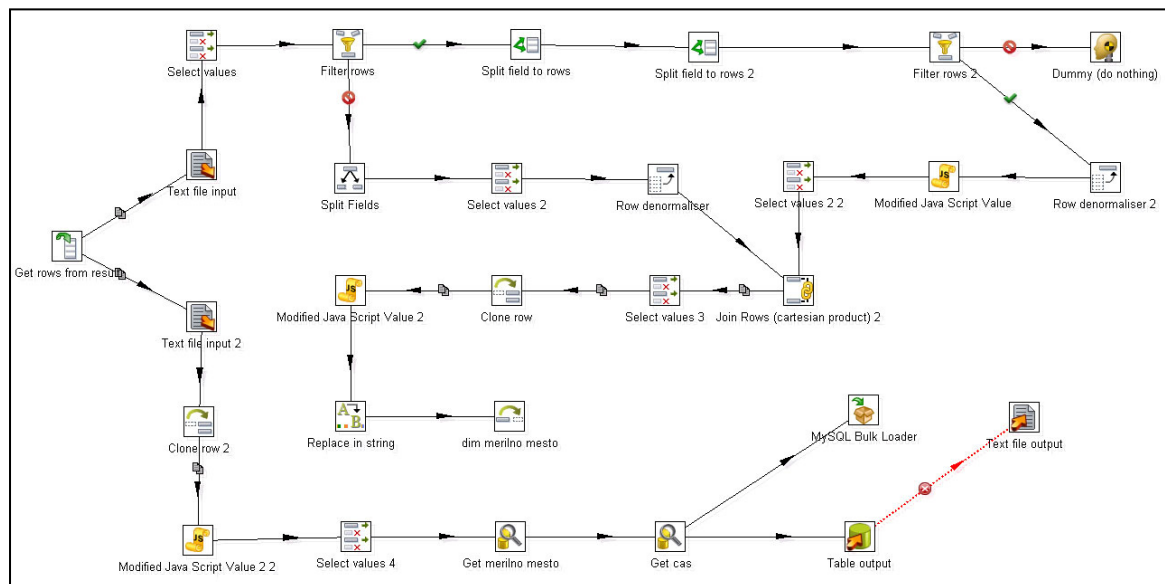
PRIPRAVA PODATKOV - PODATKOVNO SKLADIŠČE





PRIPRAVA PODATKOV - PROCESI ETL

- Orodje Pentaho Data Integration
- Trije glavni procesi ETL (Extract, Transform, Load)
- Čiščenje podatkov zaradi manjkajočih vrednosti – eksponentno glajenje





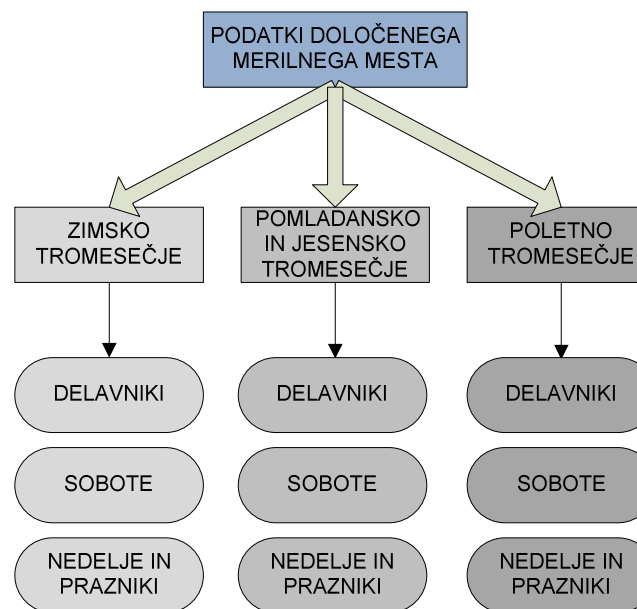
VSEBINA PREDSTAVITVE

- MOTIVACIJA, DOSEDANJE REŠITVE, TEZA
- PRIPRAVA PODATKOV
- **DOLOČITEV ČASOVNIH OKVIRJEV**
- NAPOVEDOVANJE PODATKOV in REZULTATI
- ZAKLJUČEK



GRUČENJE PODATKOV

- Problem določitve časovnih okvirjev
- Logična porazdelitev časovnih okvirjev ni optimalna

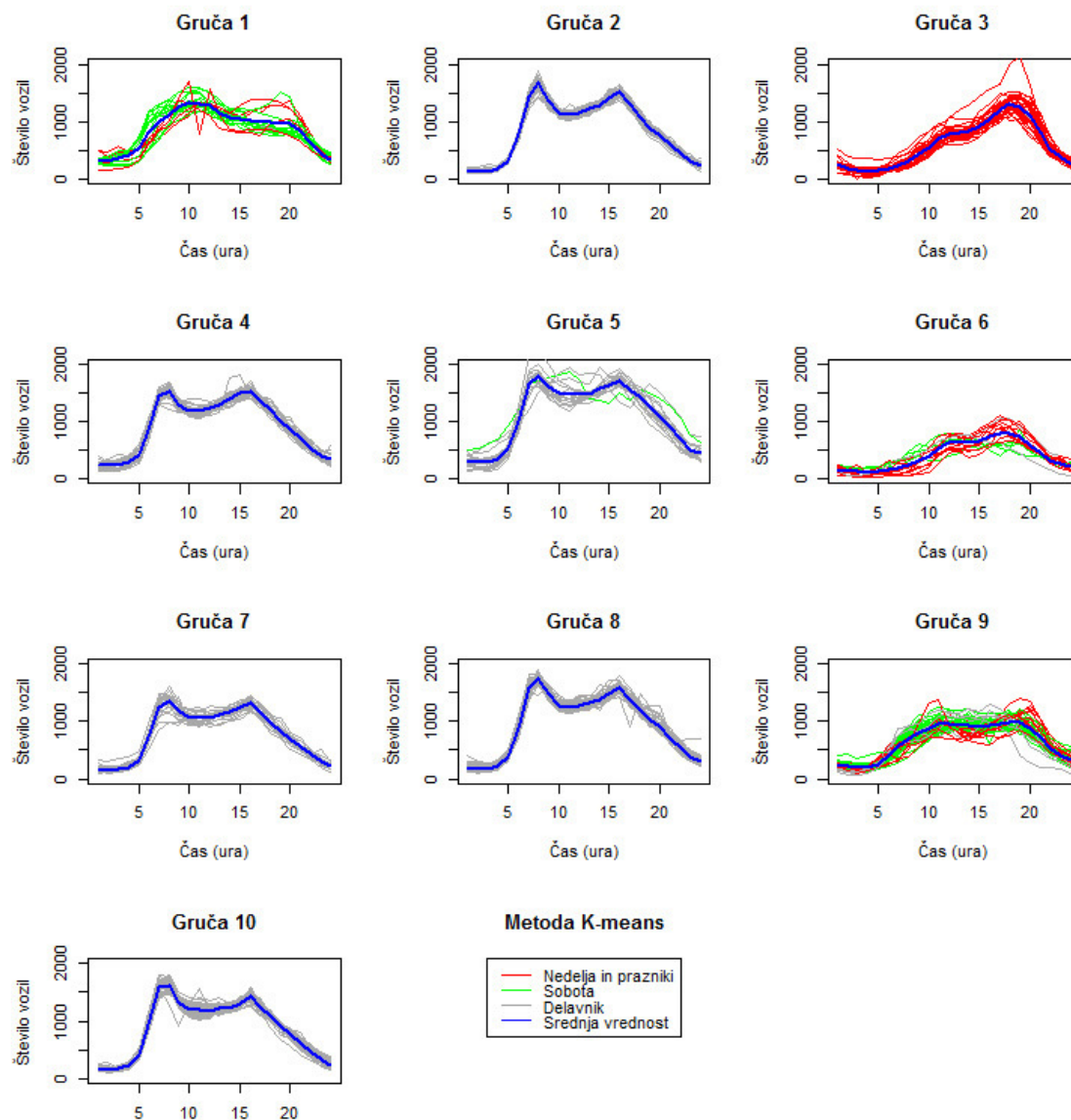




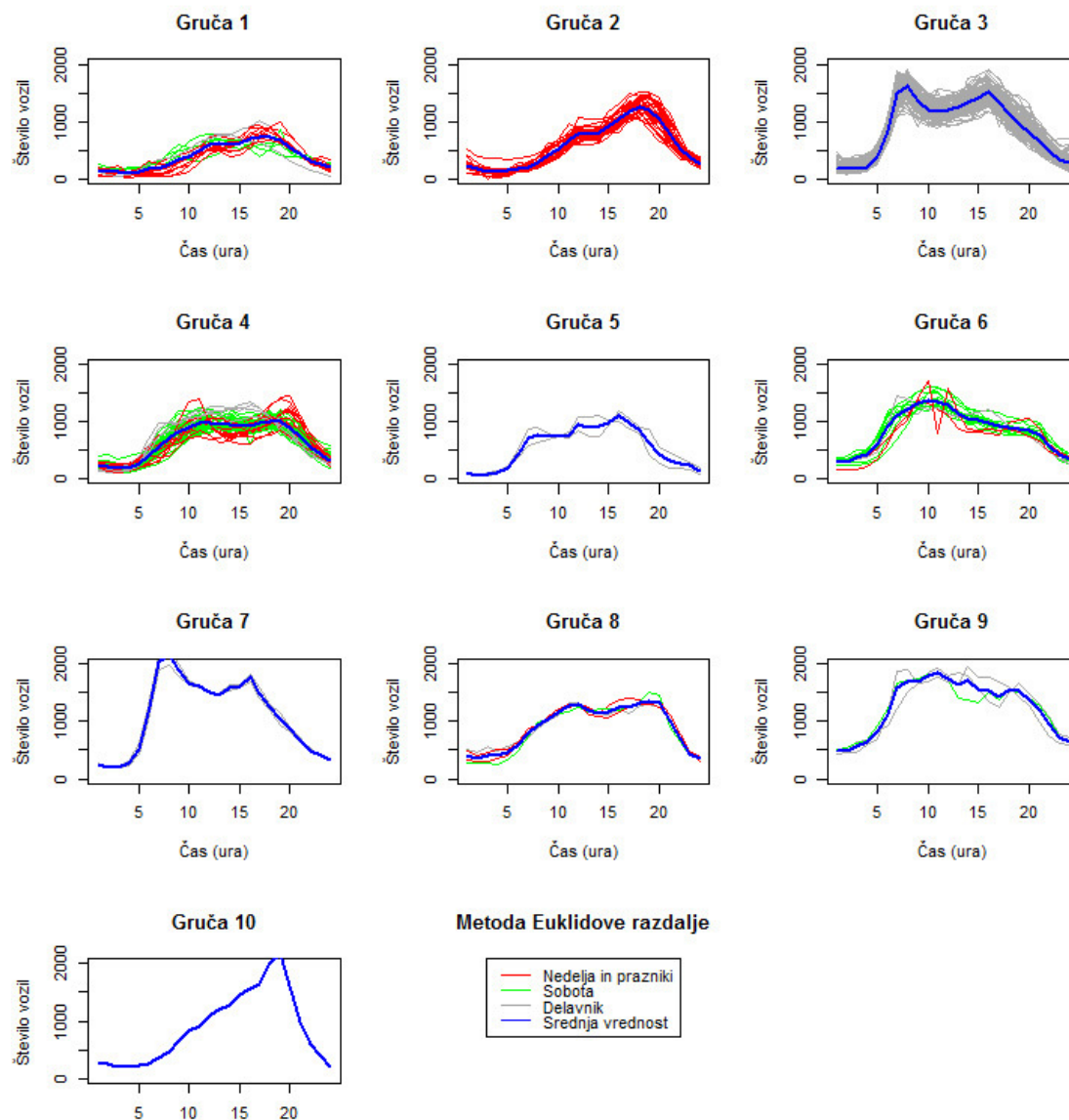
GRUČENJE PODATKOV

- Gručenje z uporabo metod podatkovnega rudarjenja
 - Metoda z delitvijo (k-Means)
 - Metoda Evklidove razdalje (Euclidian distance)
 - Metoda dinamičnega ukrivljanja časovne osi (Dynamic Time Warping)
- Porazdelitev časovnih vrst v
3,5,7,9,11,13,15,20,25,30 gruč

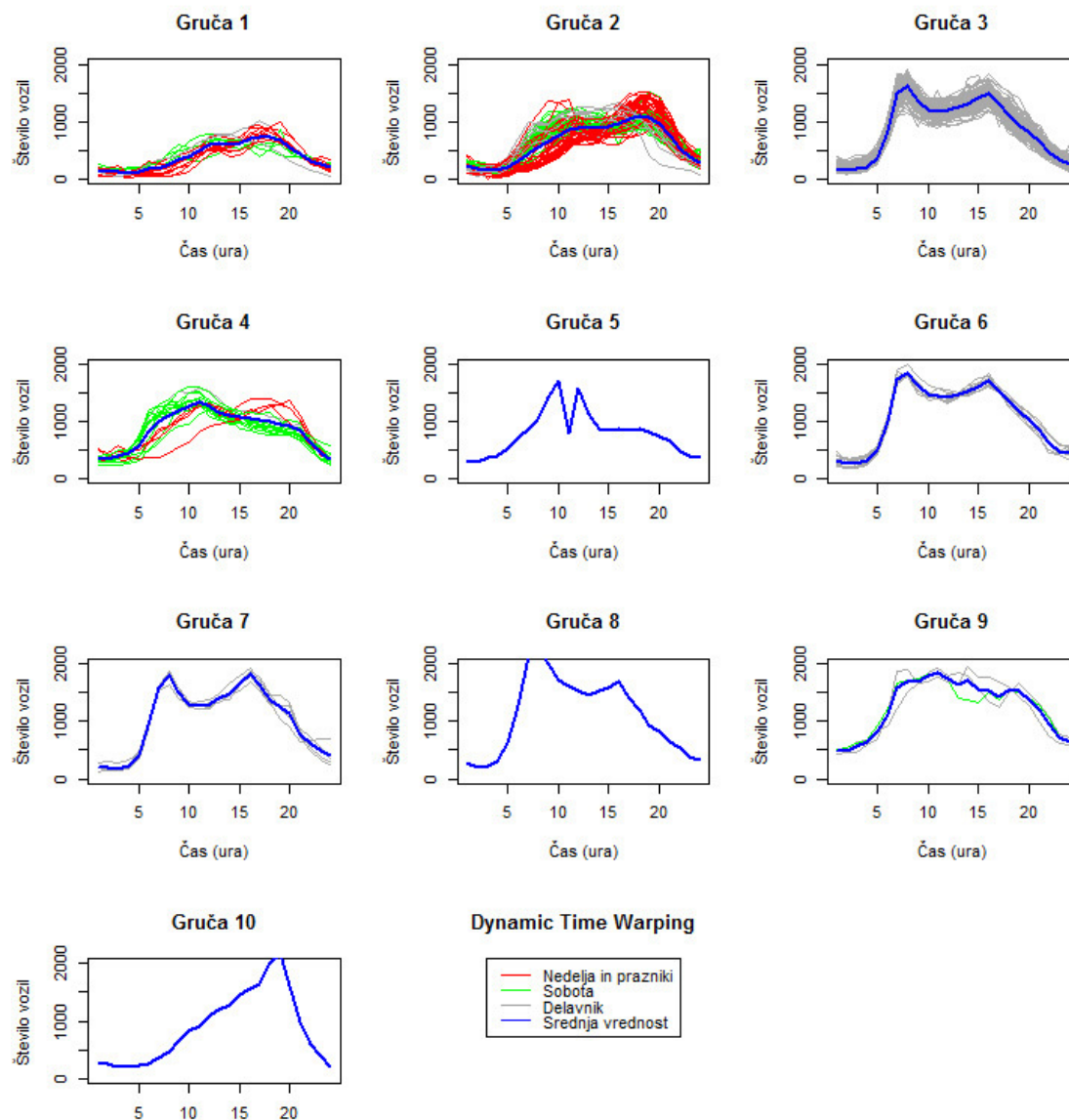
METODA Z DELITVIJO (K-MEANS)



METODA EUKLIDOVE RAZDALJE



METODA DINAMIČNEGA UKRIVLJANJA ČASOVNE OSI





RAZVOJ CENITVENE METODE

$$O[g] = \frac{\sum_{n=1}^k \left(\left| \frac{(T_g[n] - \text{avg}(T_g[n]))}{\text{avg}(T_g[n])} \right| \right)}{k}$$

$$\text{utež}[g] = \frac{\text{count}(T_g)}{\text{count}(T)}$$

$$O = \sum_{n=1}^m (O[g]_n * \text{utež}[g]_n)$$

Metoda	Število gruč	Odstopanje
k-Means	5	12,95 %
	10	10,70 %
Evklidova razdalja	5	15,35 %
	10	12,56 %
DTW	5	16,51 %
	10	13,71 %



VSEBINA PREDSTAVITVE

- MOTIVACIJA, DOSEDANJE REŠITVE, TEZA
- PRIPRAVA PODATKOV
- DOLOČITEV ČASOVNIH OKVIRJEV
- **NAPOVEDOVANJE PODATKOV in REZULTATI**
- ZAKLJUČEK



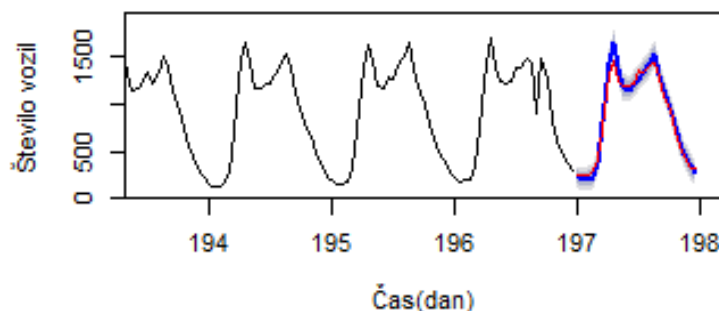
METODA ARIMA(p, d, q)

- AR model: samodejno-regresijski model
- MA model: model drsečih sredin
- Določitev parametrov p, d, q
 - Avtokorelacijska funkcija
 - Parcialna avtokorelacijska funkcija
 - Akaikejev informacijski kriterij
- Prisotnost ciklov
 - ARIMA (p,d,q) (P, D, Q)₂₄



NAPOVEDNA NAPAKA

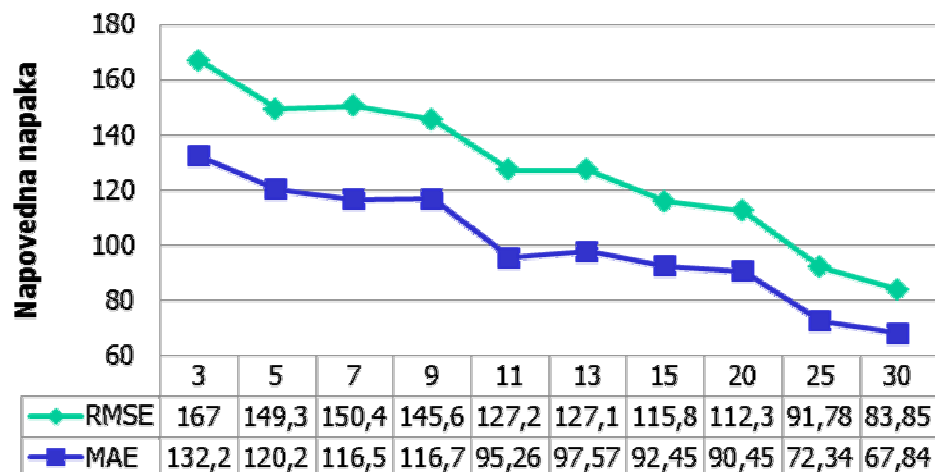
- Razlika med testno množico in napovednimi vrednostmi – napovedna napaka
 - 67% podatkov – učna množica
 - 33% podatkov – testna množica
- Napovedna napaka (RMSE, MAE, MAPE)



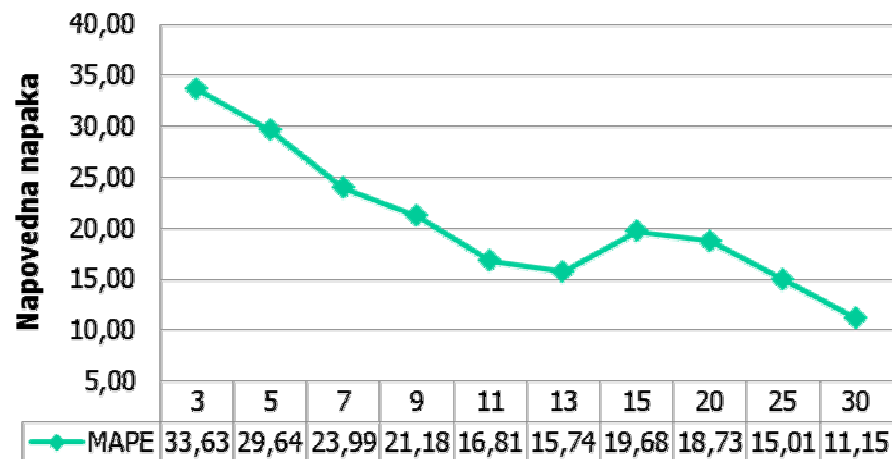


REZULTATI NAPOVEDNE NAPAKE

RMSE in MAE



MAPE





NAPOVEDNA NAPAKA

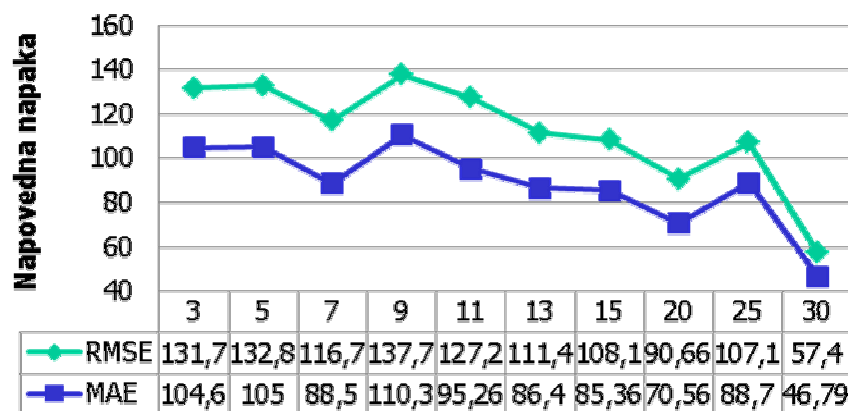
- Problem anomalij v časovnih vrstah zaradi nepredvidljivih dogodkov
- Razvoj izločitvene metode
 - Prepoznava vzorca v posamezni gruči
 - Določitev meje zaupanja
 - Izločili smo 10% časovnih vrst
 - Izločitev anomalij



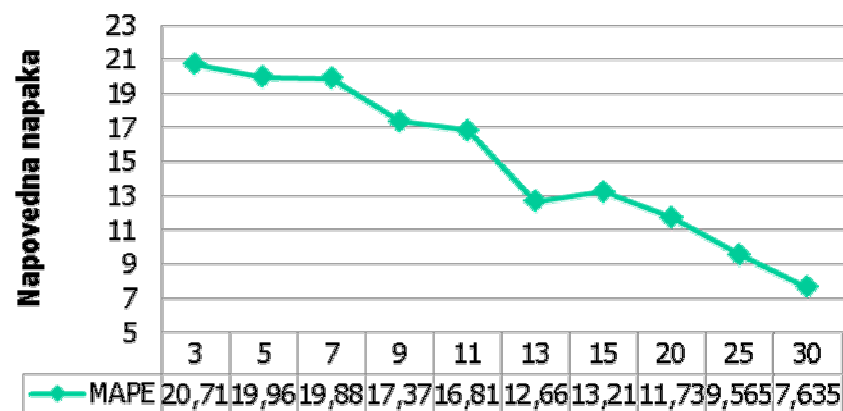
REZULTATI NAPOVEDNE NAPAKE

■ Gruče brez anomalij

RMSE in MAE



MAPE





VPRAŠANJA