

ALGORITMI ANALIZE MASIVNIH PODATKOV

DOMEN MONGUS

P01 - Uvod

Vsebina

- definicija masivnih podatkov,
- ključni podatkovni viri,
- analiza trendov razvoja in
- aktualni izzivi na področju

SAY BIG DATA



ONE MORE TIME

Definicija

- Tradicionalno 3xV
 - ▣ **Volume** (Velikost)
 - ▣ **Velocity** (Hitrost)
 - ▣ **Variety** (Raznorodnost)



- **Vedno premikajoča tarča!**
 - ▣ Masivnih podatkov ni mogoče učinkovito obdelati s tradicionalnimi metodami obdelave podatkov

Definicija

- Definicija s 5xV
 - ▣ Volume (Velikost)
 - ▣ Velocity (Hitrost)
 - ▣ Variety (Raznorodnost)
 - ▣ **Veracity** (Verodostojnost)
 - ▣ **Value** (Vrednost)
- **Podpora pri odločanju!**
 - ▣ Analize in napovedi so fokus obdelave podatkov!



Aplikacije



Tudi mi smo senzorji ...

- Kje se srečamo s tehnologijami masivnih podatkov?
 - ▣ Ste kdaj prejeli ciljno reklamno sporočilo?
 - Analize obnašanja na spletu
 - ▣ Uporabljate kartice popusta?
 - Nakupovalne navade
 - ▣ Ste kdaj plačali s kreditno kartico?
 - Zaznava goljufivih bančnih transakcij
 - ▣ Ste prejeli popust pri zavarovali avto?
 - Bonusni programi zavarovalnic (primer sledenja)
 - ▣ Uporabljate elektriko?
 - Napovedovanje porabe električne energije

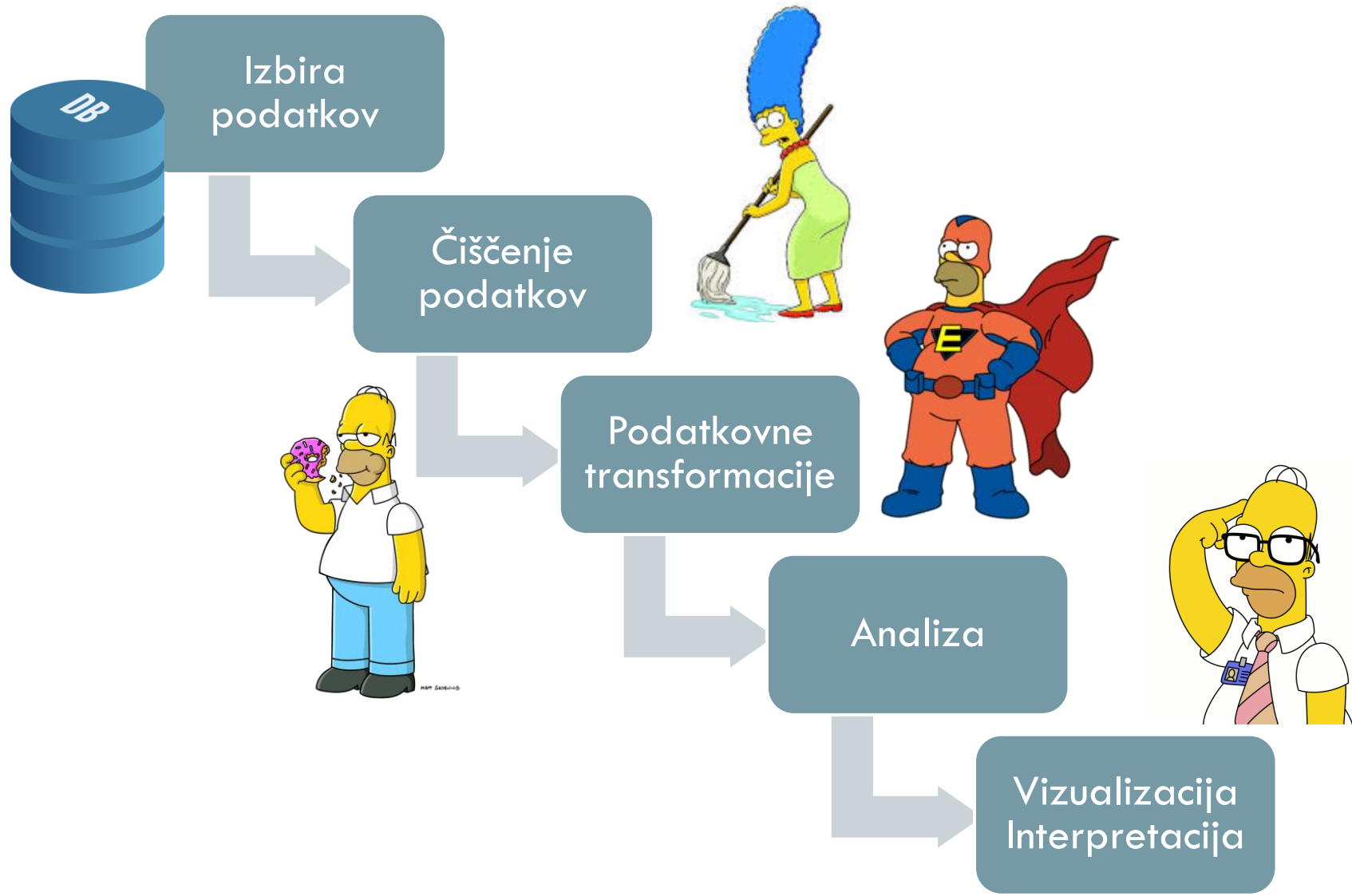
Osnovni izrazoslovje

- **Ciljna spremenljivka (ang. target variable)** je spremenljivka, ki jo napovedujemo
- **Razlagalna spremenljivka (ang. explanatory variable)** je vrednost, ki jo uporabljamo za napovedovanje
- **Značilnica (ang. feature)**, je značilna lastnost ali značilen vidik nečesa
- **Izdvajanje značilnic (ang. feature extraction)** je postopek strukturiranja značilnic za nadaljno rabo
- **Podatkovne transformacije (ang. data transformation)** je preslikava podatkov iz ene oblike/strukture v drugo

Osnovni izrazoslovje

- **Čiščenje podatkov (ang. data cleaning)** je proces zaznave in odstranjevanja manjkajočih ali pokvarjenih podatkov
- **Izbira podatkov (ang. data selection)** je postopek izbiranja ustreznih podatkov za naše analize
- **Agregacija podatkov (ang. data aggregation)**, je združevanje podatkov v skupne strukture
- **Podatkovno zlivanje (ang. data fusion)** je postopek združevanja komplementarnih podatkov
- **Podatkovno bogatenje (ang. data enrichment)** pomeni izboljševanje informacijske vrednosti podatkov skozi njihovo združevanje

Tradicionalen model procesa podatkovne analitike



Masivni podatki spreminjajo svet!

Nekoč

- ☐ Zastavi vprašanje
- ☐ Izvedi eksperiment
- ☐ Zberi podatke
- ☐ Analiziraj podatke
- ☐ Odgovori na vprašanje

Danes

- ☐ Zbiraj podatke
- ☐ Zastavi vprašanje
- ☐ Strukturiraj podatke
- ☐ Analiziraj podatke
- ☐ Odgovori na vprašanje

Podatkovni viri

Masivni podatki so nestrukturirani ali delno strukturirani

Masivni podatki vsebujejo vrednost

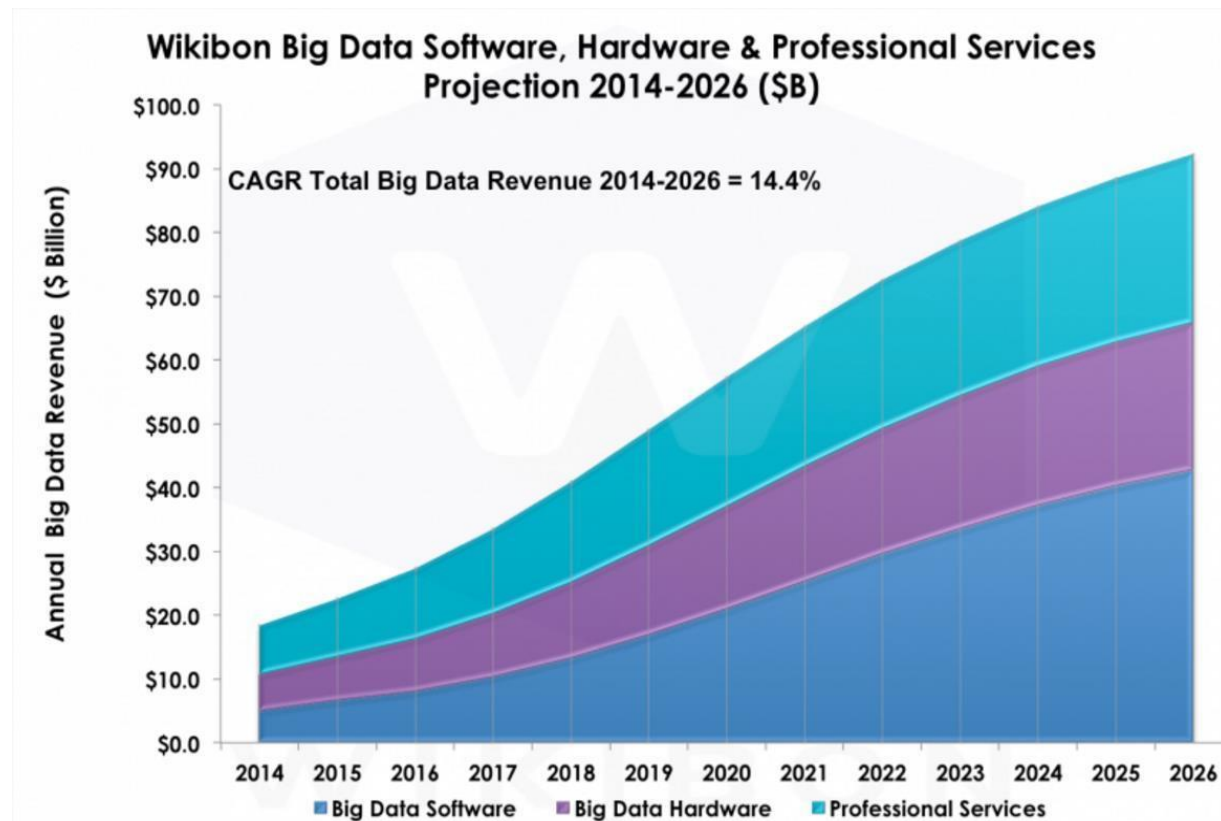
Masivni podatki niso enostavno razumljivi

- ☐ Senzorski podatki
- ☐ Logi naprav
- ☐ Spletne strani
- ☐ EMaili
- ☐ Socialna omrežja
- ☐ Mediji
- ☐ Tradicionalni dokumenti
- ☐ ...

Trenutni trendi

□ Ključni uporabniki

1. Finančni sektor
2. Proizvodnja
3. Prodaja
4. Mediji/zabava
5. Igralništvo
6. Zdravstvo
7. Telekomunikacije
8. Vladne organizacije



Source: © Wikibon Big Data Project, 2016