

M U N I
S C I

Bi8700 Vybrané kapitoly ze zpracování, analýzy a vizualizace dat

Úvod



Základní informace o předmětu

– <https://is.muni.cz/auth/predmet/sci/jaro2019/BI8700>

is > Katalog předmětů > PFF:BI8700 Vybr kap zprac, anal a viz dat - Informace o předmětu

Bi8700 Vybrané kapitoly ze zpracování, analýzy a vizualizace dat

Přírodovědecká fakulta

jaro 2019

Rozsah

0/1/0. 2 kr. Ukončení: z.

Vyučující

RNDr. Martin Komenda, Ph.D. (přednášející)

Mgr. Matěj Karolý (přednášející)

Mgr. Martin Víta (přednášející)

Mgr. Jakub Ščavnický (cvičící)

Garance

RNDr. Martin Komenda, Ph.D.

Centrum pro výzkum toxických látek v prostředí - Chemická sekce - Přírodovědecká fakulta

Kontaktní osoba: RNDr. Martin Komenda, Ph.D.

Dodavatelské pracoviště: Centrum pro výzkum toxických látek v prostředí - Chemická sekce - Přírodovědecká fakulta

Předpoklady

Předpokladem je zájem o oblast zpracování, analýzy a vizualizace dat.

Omezení zápisu do předmětu

Předmět je nabízen i studentům mimo mateřské obory.

Předmět si smí zapsat nejvýše 10 stud.

Momentální stav registrace a zápisu: zapsáno: 0/10, pouze zareg.: 0/10, pouze zareg. s předností (mateřské obory): 0/10

Mateřské obory

• Matematická biologie (program PFF, N-EXB)

Cíle předmětu

Předmět detailně představí vybraná témata z oblasti zpracování, analýzy a vizualizace dat. Vždy budou vybrány 4 aktuální projekty, která budou ve formě interaktivních workshopů pod vedením zkušených mentorů prezentovat použití ověřených metodik a metod dolování dat, analytických postupů a technik v praxi. Každá část bude vždy rozdělena na blok představující nezbytné teoretické pozadí a blok, který se bude věnovat praktickým výstupům a řešení výzkumných otázek ve spolupráci se studenty.

Výstupy z učení

Student porozumí potřebnosti systematickému používání metodik a metod pro dolování dat.

Student se seznámí s aktuálními trendy ve zpracování, analýze a vizualizaci dat.

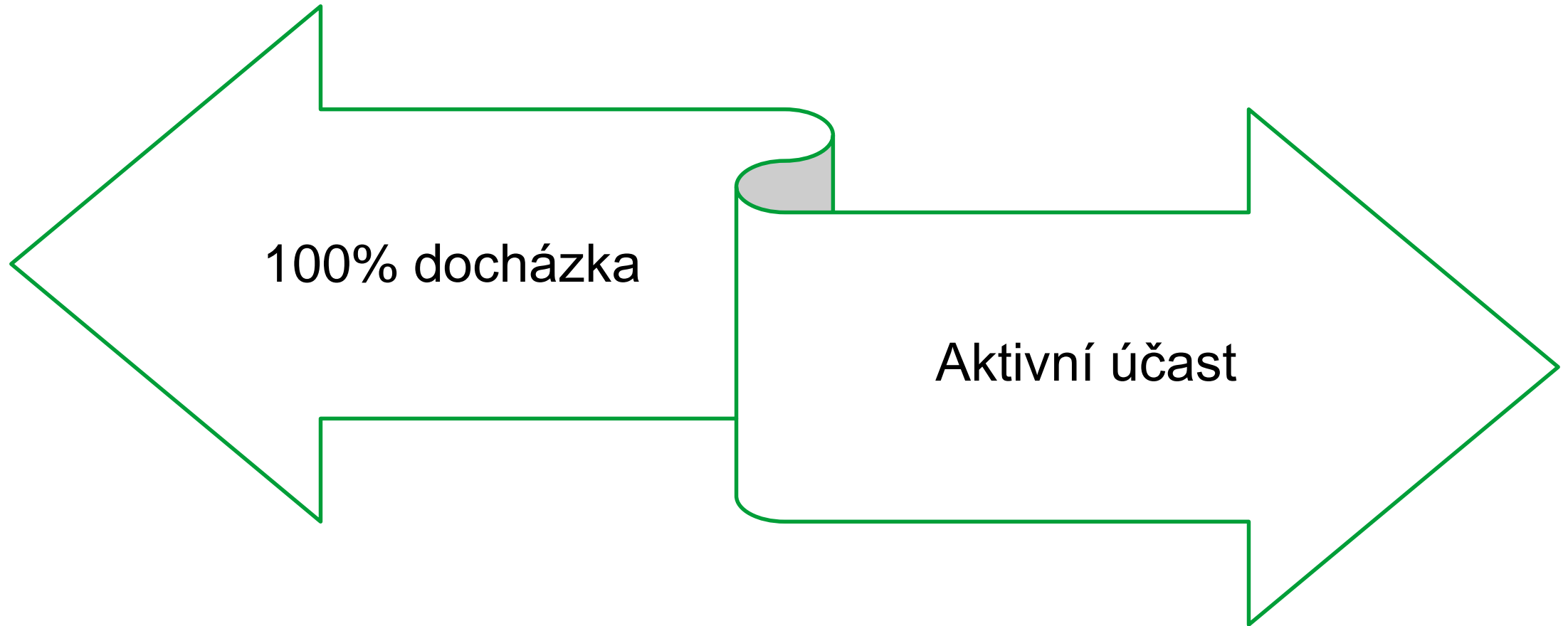
Student si osvojí nové postupy při řešení zadaných výzkumných problémů.

Bi8700 Vybrané kapitoly ze zpracování, analýzy a vizualizace dat (jaro 2019)

Výstupy z učení

- Student porozumí potřebnosti systematickému používání metodik a metod pro dolování dat.
- Student se seznámí s aktuálními trendy při zpracování, analýze a vizualizaci dat.
- Student si osvojí nové postupy při řešení zadaných výzkumných problémů.

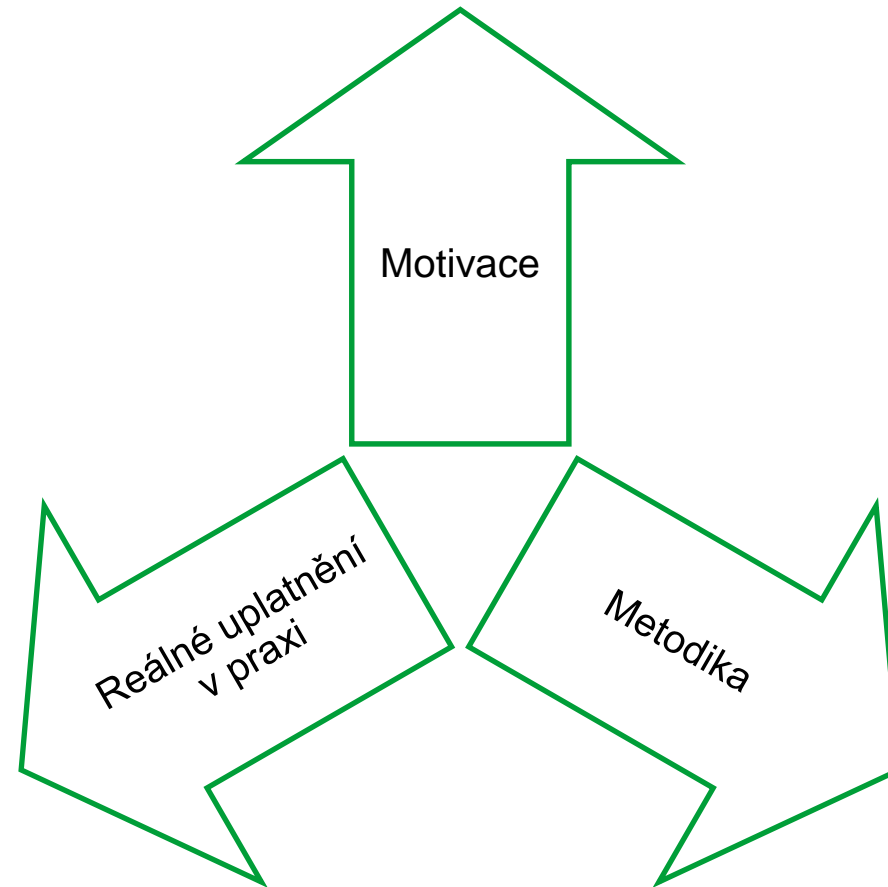
Podmínky pro udělení zápočtu



Tematické bloky

1. Mapování nejen medicínského kurikula
2. Strojové zpracování a analýza dat
3. Výběr správné vizualizace a „data storytelling“
4. Deep learning I.
5. Deep learning II.

Úvod do světa zpracování, analýzy a vizualizace dat



Motivace

- Získat užitečné informace z dostupných dat
- Zautomatizovat proces vytěžování, zpracování a vizualizace dat
- Interpretovat data správným způsobem
- Poskytnout interaktivní pohled v souladu s požadavky „konzumenta“

Co vám řešení projektu přinese?

– Multidisciplinární přístup

- Orientace ve specifické doméně (medicína, výuka, ekonomie, ...)
- Zpracování přirozeného jazyka, deskriptivní statistika, analýza dat, ...
- Vizualizace dat
- Testování a evaluace
- Nasazení v praxi

– Týmová spolupráce



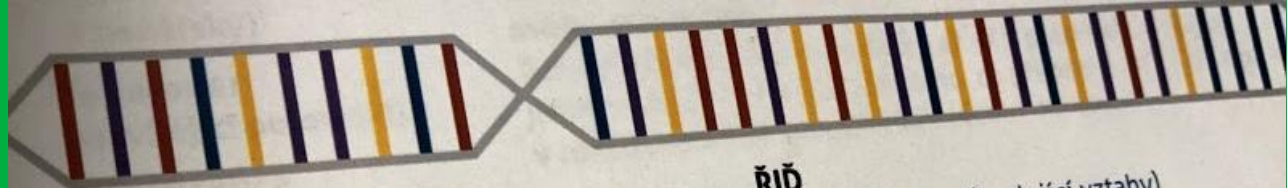


Jaké vlastnosti by měl mít člen týmu, který se věnuje zpracování a analýze dat?

CliftonStrengths®



*Weakness fixing
prevents failure,
but strengths building
leads to success.*



POSILUJ

1. **Analytical (analytický)**
2. **Focus (soustředěný/zaměřený)**
3. **Futuristic (vizionářský)**
4. **Developer (rozdávající potenciál)**
5. **Significance (uznávaný/respektovaný)**
6. Restorative (napravující)
7. Achiever (produktivní)
8. Communication (komunikativní)
9. Harmony (vytvářející harmonii)
10. Strategic (strategický)

ŘÍD

11. Relator (budující vztahy)
12. Discipline (disciplinovaný)
13. Competition (soutěživý)
14. Ideation (vymýšlející)
15. Input (zvědavý)
16. Responsibility (zodpovědný)
17. Activator (akční)
18. Intellection (intelektuální)
19. Self-Assurance (sebejistý)
20. Arranger (zařizující/organizující)
21. Positivity (pozitivní)
22. Maximizer (maximalizující)
23. Context (chápující souvislosti)
24. Deliberative (uvážlivý)
25. Individualization (mající o
26. Woo (společenský)
27. Consistency (konzistentní)
28. Learner (neustále se učící)
29. Belief (mající přesvědčení)
30. Command (velící)
31. Empathy (empatický)
32. Connectedness (propojený)
33. Includer (tmelící)
34. Adaptability (přizpůsobivý)

Jaké vlastnosti/silné stránky a proč by měl mít člověk, který je součástí analytického týmu?

M U N I

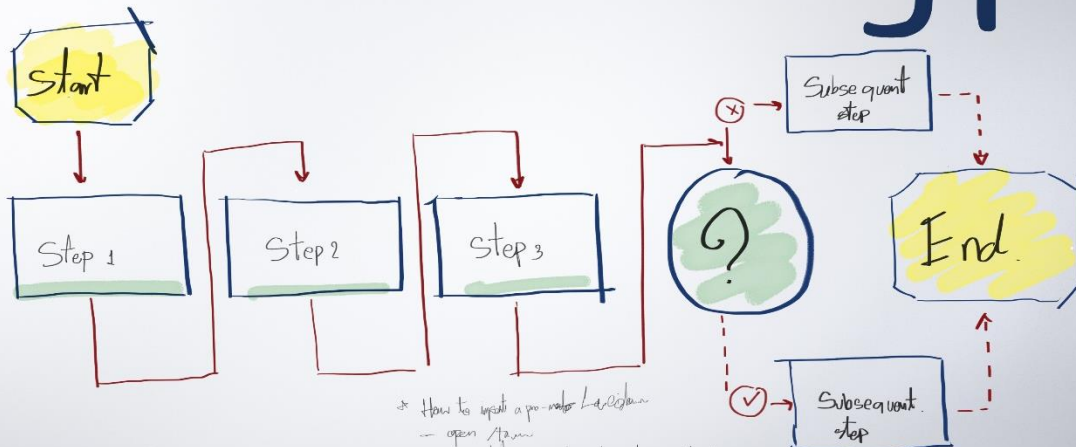
S C I

3

2

1

*Methodology



* How to insert a pre-numbered list item
- open Home
- font plate -> option 1 word -> list item

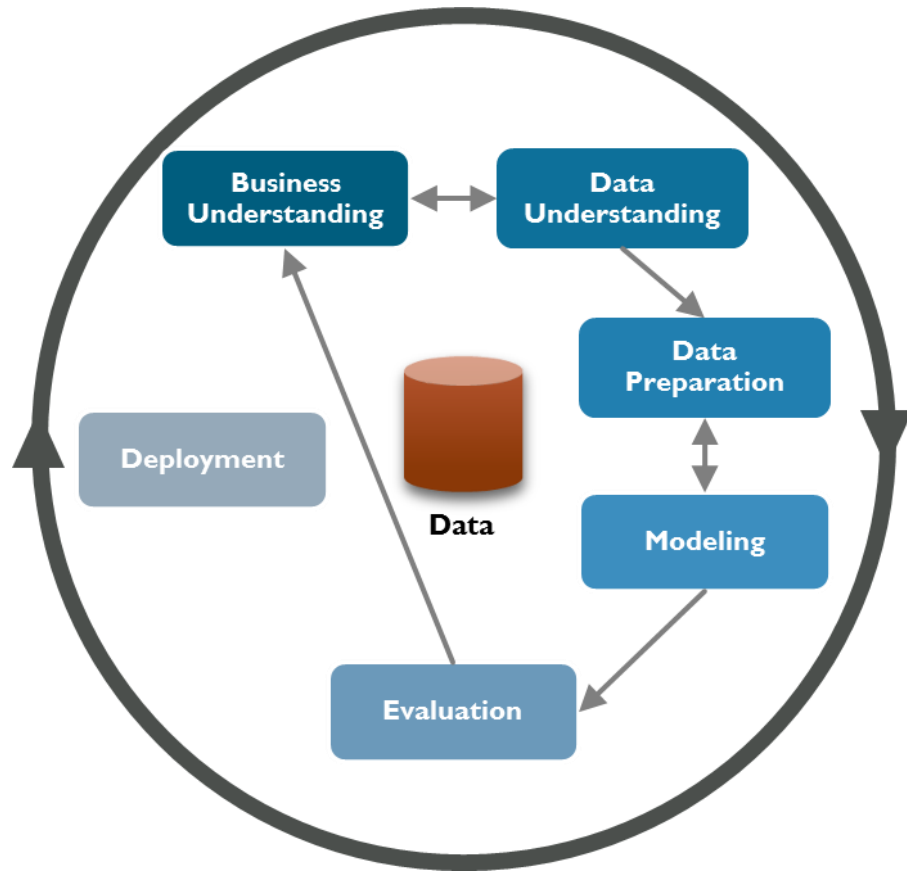




Metodiky pro vytěžování informací z dat

KDD	SEMMA	CRISP-DM
Pre KDD	-----	Business understanding
Selection	Sample	Data Understanding
Pre processing	Explore	
Transformation	Modify	Data preparation
Data mining	Model	Modeling
Interpretation/Evaluation	Assessment	Evaluation
Post KDD	-----	Deployment

Cross-Industry Standard Process for Data Mining



Cross-Industry Standard Process for Data Mining

- Porozumění problematice (Business Understanding)
 - Zadání úlohy a cíle data miningu, vymezení toho, co chce uživatel znát.

- Porozumění datům (Data Understanding)
 - Posouzení smyslu, kvality a významnosti dat.

- Příprava dat (Data Preparation)
 - Příprava dat zahrnující čištění dat, transformaci dat, práci s chybějícími hodnotami apod.

Cross-Industry Standard Process for Data Mining

- Modelování (Modeling)

 - Analytické zpracování, tvorba modelu, vizualizace dat.

- Vyhodnocení výsledků (Evaluation)

 - Ověření správnosti modelů.

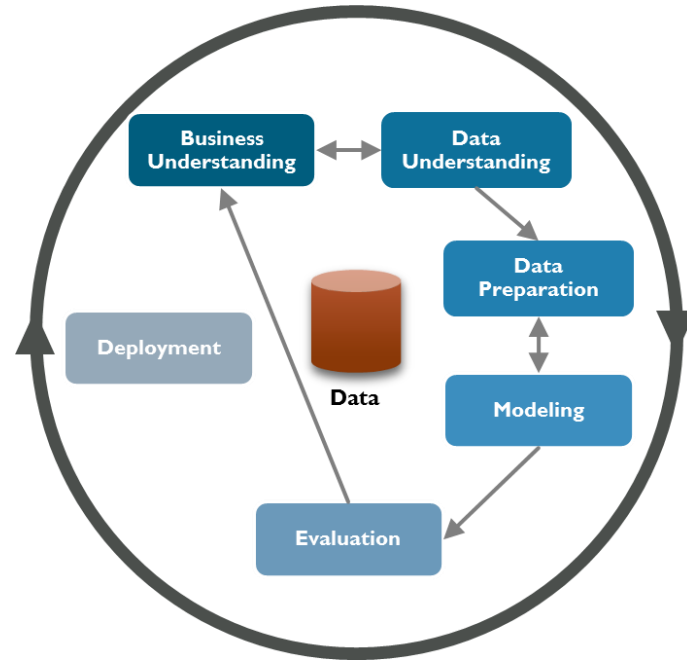
- Využití výsledků (Deployment)

 - Použití výsledků v praxi, prezentace výstupů.



Která fáze modelu CRISP-DM je nejdůležitější, časově nejnáročnější a proč?

Která fáze modelu CRISP-DM je
nejdůležitější, časově nejnáročnější a proč?



3

2

1



Jaké otázky bychom si měli položit před zahájením řešení projektu?



Bi8700 Vybrané kapitoly ze zpracování, analýzy a vizualizace dat (jaro 2019)

Jaké otázky bychom si měli položit před
zahájením řešení projektu?

M U N I

S C I

5

4

3

2

1

5 questions to ask before launching a project

Are we developing an analytics solution for a "one-shot deal", i.e., simply answering a specific business question, or for an ongoing process?

Where is the data?

What kind of preparation is needed for the data?

Do we need to reduce the data dimension?

Which technique to use?



5

4

3

2

1