

Nitrianske kompetenčné centrum kybernetickej bezpečnosti

Fakulta prírodných vied a informatiky · Univerzita Konštantína Filozofa v Nitre

Tr. A. Hlinku 1, 949 01 Nitra

Kritické myslenie pri práci s AI nástrojmi v dátovej analýze

Rozšírenie obsahu predmetu Počítačová analýza dát

Učebný text pre študentov: práca s generatívnou AI, best practices, identifikácia chýb AI a úlohy na precvičenie.

Mgr. Jaroslav Reichel, PhD.

jreichel@ukf.sk

Obsah

O tomto učebnom texte	3
Zaradenie do predmetu	3
1 Čo je generatívna AI a ako funguje	3
1.1 Dva spôsoby, ako AI v analýze používame	4
2 Best practices pri analýze dát s AI	4
2.1 Najprv otázka, potom AI	4
2.2 Od promptu ku kontextu: ako zadať analýzu	4
2.3 Validácia výstupov	8
2.4 Reprodukovateľnosť a transparentnosť	9
2.5 Nástroj nie je cieľ	9
3 Identifikácia chýb AI	10
3.1 Typológia chýb AI podľa oblasti	10
4 Rámec 4D - ako pristupovať ku každej úlohe	11
5 Úlohy pre študentov	12
6 Obhajoba projektu - rubrika k prezentácii	14
7 Na záver	15
8 Odporúčané zdroje a nástroje	15

O tomto učebnom texte

Tento text rozširuje obsah predmetu Počítačová analýza dát o tému, ktorá sa za posledné roky stala neoddeliteľnou súčasťou práce dátového analytika - používanie generatívnej AI (GenAI). Cieľom nie je naučiť „ovládať AI“, ale pracovať s ňou kriticky: rozpoznať, kedy pomáha, kedy zavádza, a ako jej výstupy overiť.

Zaradenie do predmetu

Modul je zaradený na **záver predmetu** Počítačová analýza dát, nadväzujúc na blok exploračnej (EDA) aj inferenčnej analýzy (ISA) a na semestrálny projekt. Stavia na zručnostiach osvojených pri práci v jazyku Python (Jupyter Notebook: pandas, numpy, matplotlib) a rozširuje ich o prierezovú kompetenciu z profilu absolventa - **kriticky riešiť problémy**. Namiesto ďalšej novej metódy sa pri dovtedy osvojených postupoch (od početností a popisných charakteristík cez vizualizáciu až po testovanie hypotéz a vzťahy medzi premennými) kladie otázka, ako rozpoznať, kedy AI pomáha a kedy zavádza.

VZDELÁVACIE CIELE MODULU

- Porozumenie tomu, ako GenAI funguje a prečo pri analýze dát robí určité typy chýb (nevhodný test, chybný výpočet, skreslený graf, halucinovaný zdroj).
- Osvojenie best practices pri promptovaní a validácii výsledkov.
- Schopnosť rozpoznať typické chyby AI v kóde, štatistike, vizualizácii aj interpretácii.
- Precvičenie kritického myslenia na úlohách s modelovými AI výstupmi.

1 Čo je generatívna AI a ako funguje

Generatívna AI (napríklad ChatGPT, Claude alebo Gemini) je systém postavený na veľkom jazykovom modeli (LLM - Large Language Model). Model bol natrénovaný na obrovskom množstve textu a jeho úlohou je predpovedať, aké slovo (presnejšie token) nasleduje. Z tejto jednoduchej úlohy vzniká schopnosť písať kód, vysvetľovať pojmy aj formulovať súvislé odpovede.

Kľúčové pochopenie pre kritické myslenie: model negeneruje pravdu, generuje pravdepodobný text. Odpoveď, ktorá znie najprirodzenejšie, nemusí byť správna. Práve preto vznikajú **halucinácie** - model seabavedomo uvedie neexistujúcu funkciu, vymyslenú štúdiu alebo nesprávny výpočet, pretože taký text je „pravdepodobný“, hoci nie pravdivý.

1.1 Dva spôsoby, ako AI v analýze používame

V praxi rozlišujeme dva odlišné prístupy. Pri každej úlohe si treba vybrať vedome, pretože sa líšia mierou kontroly aj rizika.

Kritérium	AI generuje kód	AI priamo spracuje dáta
<i>Ako to funguje</i>	AI napíše skript, ktorý sa spustí lokálne	Dáta sa nahrajú do AI, tá vráti výsledok
<i>Kontrola</i>	Plná - viditeľný každý krok	Black-box - nie je vidno, čo sa stalo
<i>Reprodukovateľnosť</i>	Vysoká - kód možno zopakovať	Nízka
<i>Súkromie</i>	Bezpečné (dáta zostávajú lokálne)	Rizikové (dáta opúšťajú lokálne prostredie)
<i>Kedy použiť</i>	Citlivé dáta, opakovaná či produkčná analýza	Rýchly prieskum necitlivej vzorky

POZOR NA SÚKROMIE

Pri práci s osobnými údajmi, mzdami, zdravotnými alebo inak citlivými dátami sa tieto **nikdy nenahrávajú priamo do verejných AI nástrojov**. Namiesto toho AI napíše kód, ktorý sa spustí v lokálnom prostredí, kde dáta zostávajú.

2 Best practices pri analýze dát s AI

2.1 Najprv otázka, potom AI

Najdôležitejšie pravidlo celého textu: väčšina práce dátového analytika nie je tvorba grafov ani kódu - je to presné definovanie toho, čo vlastne chceme zistiť. AI je výborná v mechanickej časti (napísať kód, vykresliť graf), no definíciu problému jej nemožno zveriť. Ak sa jej zverí aj otázka, výsledkom je presná odpoveď na nesprávne položený problém.

2.2 Od promptu ku kontextu: ako zadať analýzu

Využívanie moderných AI nástrojov pri analýze dát nie je o jednom dokonalom prompte, ale o kontexte. Tento posun sa označuje ako context engineering: kým prompting rieši, ako otázku položiť, context engineering rieši, aké informácie má model k dispozícii. Namiesto jednej dlhej požiadavky je preto často lepšie začať krátkym rozhovorom, v ktorom AI doplní kontext, položí spresňujúce otázky (je vhodné AI nástroj vyzvať na polozenie takýchto otázok) a spoločne sa ujasní smer analýzy.

Nech už ide o jednu požiadavku alebo rozhovor, dobré zadanie nesie štyri zložky kontextu:

- aké dáta sú k dispozícii (počet riadkov, názvy a typy premenných),
- čo presne sa požaduje (nie „urob analýzu“),
- v akom nástroji či jazyku (napr. Python a pandas),
- v akom rozsahu/formáte má byť výstup.

Aj pri jednorazovom (one-shot) zadaní sa oplatí rozpísať postup po krokoch a vymenovať možnosti, ktoré má model zvážiť - čo spraviť najprv, ako a od akých chýb očistiť dáta, ktoré metriky spočítať, aké grafy vykresliť, či hľadať korelácie, anomálie, trendy a špičky. Platí jednoduché pravidlo: čo nespomeniete, to model spravidla nevyužije. Pokojne preto píšete dlhšie a konkrétnejšie; čitateľ si z ponuky vyberie, čo potrebuje. Do one-shot promptu sa navyše zmestí aj stručný kontext a veta, že AI sa môže najprv opýtať.

ZAČNITE ROZHOVOROM, NIE PRÍKAZOM

Užitočný úvodný krok: „Skôr než začneš, polož mi 3 - 5 otázok, ktoré potrebuješ na to, aby si správne pochopil cieľ a dáta tejto analýzy.“ Model tak vyplní medzery, ktoré by inak doplnil dohadom. Odpovede následne tvoria kontext pre samotnú analýzu.

Situácia EDA (exploračná analýza)

Cieľom EDA je objavovanie, preto výstup nešpecifikuj rigidne (žiadna pevná schéma). Ujasni skôr otázku alebo rozhodnutie, ktoré analýza podporí, premenné a ich typy a to, ako naložiť s chýbajúcimi a odľahlými hodnotami. Postup však pokojne rozpiš po krokoch a vymenuj možnosti. Nasledujúci príklad je zámerne dlhý a podrobný, aby ukázal, čo všetko sa doň dá vložiť. V praxi využijete len časť.

PRÍKLAD - ROZŠÍRENÝ ONE-SHOT PROMPT PRE EDA

„Si dátový analytik. Pomôž mi s exploračnou analýzou. Ak ti v zadaní niečo chýba alebo je nejednoznačné, najprv sa opýtaj.“

Kontext a cieľ: mám CSV s 500 zamestnancami - premenné vek (číselná), plat (číselná, EUR), pohlavie (muž/žena), oddelenie (6 kategórií), odpracované roky (číselná). Cieľ: pochopiť, ako sa líšia platy podľa pohlavia a oddelení - ako podklad pred testovaním hypotéz. Pracuj v Pythone (pandas, numpy, matplotlib/seaborn), kód komentuj po slovensky.

Postupuj takto:

- 1. Prehľad dát:** rozmery, typy stĺpcov, ukážka riadkov, počet unikátnych hodnôt kategórií.
- 2. Kvalita dát (každú zmenu vypíš):** duplicity; chýbajúce hodnoty (počet aj podiel, navrhní a zdôvodni doplnenie alebo vynechanie); nelogické hodnoty (záporný plat, vek mimo 15 - 80, plat 0); nekonzistentné kategórie (rôzny zápis „muž“/„M“); odľahlé hodnoty (IQR alebo z-skóre - len ich označ, neodstraňuj automaticky).
- 3. Popisné štatistiky:** priemer, medián, modus, smerodajná odchýlka, min/max a kvartily; pre kategórie početnosti a podiely; plat rozdeľ podľa pohlavia aj oddelenia (vrátane mediánu, nielen priemeru).
- 4. Rozdelenia a tvar:** histogramy a boxploty platu a veku; šikmosť a špicatosť; posúď, či sa plat blíži normálnemu rozdeleniu (dôležité pre neskoršie testy).
- 5. Vzťahy a vzory:** korelačná matica (Pearson aj Spearman); boxploty platu podľa oddelenia a pohlavia; rozptylový graf plat verzus odpracované roky; hľadaj anomálie, neočakávané zhľady, extrémny a špičky (napr. oddelenie s neúmerne vysokým platom).

6. Segmentácia: porovnaj kombinácie (plat podľa pohlavia krát oddelenie) a označ najväčšie rozdiely.

7. Zhrnutie: 5 - 7 kľúčových zistení; ku každému uveď, na ktorom čísle alebo grafe stojí; vyznač, čo je len korelácia (nie príčina) a čo treba ešte overiť testom.

Pravidlá: ku každému kroku pripoj krátky komentár; uveď prijaté predpoklady; nepredkladaj koreláciu ako kauzalitu.“

Situácia ISA (inferenčná analýza)

Pri ISA musí byť jasná konkrétna hypotéza. Pred výpočtom ujasni: nulovú a alternatívnu hypotézu (H_0/H_1), premenné a ich typy, či sú vzorky závislé alebo nezávislé, smer testu (jedno- alebo obojstranný), hladinu významnosti a predpoklady testu. Vhodný postup: nechaj AI najprv navrhnúť vhodný test a overiť jeho predpoklady, až potom počítať a interpretovať.

PRÍKLAD - ROZŠÍRENÝ ONE-SHOT PROMPT PRE ISA

„Si dátový analytik. Pomôž mi otestovať hypotézu. Ak ti niečo chýba alebo je nejednoznačné, najprv sa opýtaj.

Kontext a hypotéza: spokojnosť zákazníkov (škála 0 - 10) v dvoch nezávislých skupinách, A ($n = 50$) a B ($n = 45$). H_0 : stredné hodnoty sa nelíšia; H_1 : líšia sa. Obojstranný test, $\alpha = 0,05$. Pracuj v Pythone (pandas, scipy), komentuj po slovensky.

Postupuj takto:

1. Popis dát: veľkosti skupín, popisné štatistiky a boxploty oboch skupín; vizuálne porovnaj rozdelenia.

2. Overenie predpokladov: normalita v oboch skupinách (Shapiro - Wilk a Q - Q graf), homogenita rozptylov (Levene).

3. Výber testu: na základe predpokladov navrhni a zdôvodni test - Studentov t-test (rovnaké rozptyly), Welchov t-test (nerovnaké) alebo neparametrický Mann - Whitney (porušená normalita). Najprv vysvetli voľbu, až potom počítaj.

4. Výpočet a interpretácia: testová štatistika, stupne voľnosti, p-hodnota a rozhodnutie voči α ; uveď interval spoľahlivosti rozdielu.

5. Praktická významnosť: veľkosť účinku (Cohenovo d alebo rank-biserial) a slovná interpretácia - nielen či je výsledok „štatisticky významný“.

6. Záver a limity: čo z výsledku vyplýva pre pôvodnú otázku a aké sú obmedzenia (veľkosť vzorky, predpoklady, riziko viacnásobného testovania).

Pravidlá: neprezentuj koreláciu ako kauzalitu; nehľadaj test, ktorý „vyjde“ (p-hacking); ak meníš postup, povedz prečo.“

Ďalšie fázy: čistenie, vizualizácia, interpretácia

Rovnaký princíp - rozpísať kroky a vymenovať možnosti - platí pre každú fázu. Nasledujúce ukážky sú kratšie a obsahujú zástupné polia [doplň: ...], ktoré sa nahradia konkrétnym zadáním.

PROMPT - ČISTENIE DÁT

„Si dátový analytik. Vyčisti dataset [doplň: zdroj/CSV] v Pythone (pandas); každú zmenu vypíš a zdôvodni. Postupuj:

1. **Duplicity:** nájdi a odstráň úplné aj čiastočné duplicity (podľa [doplň: kľúčové stĺpce]).
2. **Chýbajúce hodnoty:** vypíš počet a podiel po stĺpcoch; navrhni a zdôvodni stratégiu (vynechať / medián / modus / interpolácia) zvlášť pre každý stĺpec.
3. **Typy a formáty:** zjednoť dátové typy, dátumy, desatinné oddeľovače a jednotky; orež medzery v texte.
4. **Kategórie:** zjednoť nekonzistentné zápisy (napr. „muž“/„M“/„m“) a vypíš použité mapovanie.
5. **Nelogické hodnoty:** označ a ošetri hodnoty mimo rozsahu [doplň: pravidlá, napr. vek 15 - 80, nezáporný plat].
6. **Odľahlé hodnoty:** identifikuj (IQR alebo z-skóre) a len ich označ - neodstraňuj bez môjho súhlasu.
7. **Výstup:** čistý dataset + tabuľka „pred/po“ (počet riadkov, chýbajúce, duplicity) a zoznam vykonaných úprav.

Pravidlá: nič nemeň potichu; ak si nie si istý stratégiou, opýtaj sa.“

PROMPT - VÝBER A TVORBA VIZUALIZÁCIE

„Pomôž mi zvoliť a vytvoriť vhodné grafy pre [doplň: premenné/otázku] v [doplň: matplotlib/seaborn]. Postupuj:

1. Pre každú otázku navrhni vhodný typ grafu podľa typu premenných (kategória vs. číselná, jedna vs. dve, čas) a krátko zdôvodni voľbu.
2. Dodrž zásady poctivého grafu: os Y od nuly (ak nie, zdôvodni), rovnomerné škály, vhodný typ, čitateľné popisy, farby prístupné aj pri farbosleposti.
3. Vyhni sa skresleniam: 3D efekty, koláč pre vývoj v čase, orezaný rozsah, ikony namiesto čísel.
4. Ku každému grafu uveď, čo má ukázať a čo NEukazuje.
5. Výstup: kód + jedna veta interpretácie ku každému grafu.

Pravidlá: žiadne ozdoby na úkor presnosti; ak dáta nie sú vhodné na požadovaný typ grafu, povedz to.“

PROMPT - INTERPRETÁCIA A REPORT

„Napiš interpretáciu výsledkov [doplň: vlož výstupy/tabuľky] pre [doplň: cieľové publikum].
Postupuj:

1. Zhrň hlavné zistenia v 5 - 7 bodoch; ku každému uveď konkrétne číslo alebo graf, na ktorom stojí.
2. Pri každom vzťahu rozlíš koreláciu od kauzality; ak ide len o súvislosť, napíš to.
3. Doplň praktickú významnosť (veľkosť účinku, kontext odboru), nielen p-hodnotu.
4. Uveď neistoty a limity (veľkosť vzorky, predpoklady, chýbajúce dáta).
5. Ak cituješ zdroje, uveď len overiteľné - nič si nevymýšľaj.
6. Výstup vo formáte [doplň: odsek / odrážky / manažérske zhrnutie].

Pravidlá: žiadne nepodložené tvrdenia ani vymyslené citácie; oddel', čo dáta ukazujú, od toho, čo je interpretácia.“

Šablóna promptu na vyplnenie

Kostru, ktorú si stačí skopírovať a doplniť hranaté polia. Nie každý riadok je vždy potrebný - vynechajte, čo sa na danú úlohu nehodí.

ŠABLÓNA - SKOPÍRUJ A DOPLŇ

Rola: Si dátový analytik.

Kontext a cieľ: [doplň: aké dáta mám - stĺpce a ich typy; čo chcem zistiť a prečo].

Nástroj: [doplň: napr. Python + pandas; jazyk komentárov].

Postup (kroky): [doplň: 1... 2... 3... - čo spraviť a v akom poradí].

Čo zvážiť: [doplň: kvalita dát, metriky, rozdelenia, grafy, korelácie, anomálie, segmentácia].

Výstup: [doplň: formát a rozsah].

Pravidlá: ak ti niečo chýba, najprv sa opýtaj; nepredkladaj koreláciu ako kauzalitu; neuvádzaj vymyslené zdroje.

2.3 Validácia výstupov

Žiadny AI výstup sa nepreberá bez kontroly. Podľa typu výstupu:

Kód: spustiť na vlastných dátach aj na jednoduchom príklade so známym správnym výsledkom; otestovať hraničné prípady (prázdne dáta, chýbajúce hodnoty).

Analýza: porovnať aspoň podvzorku s ručným výpočtom; overiť, či výsledok dáva zmysel v kontexte.

Interpretácia: overiť existenciu každého citovaného zdroja; doplniť praktickú významnosť, nielen p-hodnotu.

2.4 Reprodukovateľnosť a transparentnosť

- Prompty aj vygenerovaný kód je potrebné uchovať - analýza musí byť zopakovateľná.
- Transparentne sa uvádza, čo bolo vygenerované pomocou AI a ako to bolo overené.
- Nepredstiera sa, že hypotéza existovala vopred, ak vznikla až po pohľade na výsledky.

2.5 Nástroj nie je cieľ

Dnes je ťažisko dátovej analýzy primárne na jazyku Python a generatívnej AI. Praktické dôvody sú, že Python aj AI sú dostupné zadarmo, aktívne sa vyvíjajú a reálne sa používajú v praxi, kým licencované štatistické softvéry ako Statistica alebo SPSS zostávajú skôr pre špecifické úlohy.

Zároveň však platí: **voľba nástroja sama osebe ešte nezaručuje lepšie pochopenie**. Klikanie cez menu v štatistickom softvéri, kde sa každý krok vyberá ručne, niekedy núti premýšľať nad tým, čo sa počíta, viac než hotový kód od AI, ktorý stačí spustiť. Rozhodujúce preto nie je, akým nástrojom sa analýza vykoná, ale či autor rozumie tomu, čo sa deje pod kapotou.

AI dokáže vyrobiť hladký výsledok aj tam, kde pochopenie chýba. Preto sa v tomto module predmetu nehodnotí, či bola použitá AI, ale či autor vie vysvetliť a obhájiť každý krok svojej analýzy.

3 Identifikácia chýb AI

Najnebezpečnejšie nie sú zjavné chyby - tie sa zvyčajne zachytia. Nebezpečná je **falošná istota**: výstup, ktorý vyzerá presne a presvedčivo, ale stojí na chybnom základe. Rozlišujú sa tri pasce.

3.1 Typológia chýb AI podľa oblasti

Pri každej oblasti analýzy vznikajú iné typické chyby:

Chyby v kóde

- Syntakticky správny, ale logicky zlý kód (počíta nesprávnu vec).
- Zastaralé alebo neexistujúce funkcie (halucinácia knižnice).
- Neefektívny kód, ktorý na väčších dátach zlyhá na výkone.

Chyby v štatistike a metóde

- Nevhodný test pre daný typ dát alebo porušené predpoklady testu.
- Nevhodná metóda imputácie chýbajúcich hodnôt bez znalosti kontextu.
- **P-hacking** - skúšanie ďalších a ďalších testov, kým nevyjde $p < 0,05$. AI ich ponúkne nekonečne veľa bez etických brzd.

Chyby vo vizualizácii

Tu sa uplatní **kontrolný checklist na grafy**. Najčastejšie chyby AI grafov:

- Os Y nezačína nulou - malé rozdiely vyzerajú dramaticky.
- Zlý typ grafu (koláč pre vývoj v čase), 3D efekty, ikony namiesto čísel.
- Cherry-picking - chýbajúce roky alebo kategórie.
- Neprístupné farby (nerozlíši ich človek s farbosleposťou).

Chyby v interpretácii

- „Štatisticky správne, obsahovo prázdne“ - chýba praktická významnosť.
- Zámena korelácie za kauzalitu.
- Prehnaná istota pri neistých výsledkoch.
- Vymyslené štúdie a citácie (halucinácie).

4 Rámec 4D - ako pristupovať ku každej úlohe

Pre orientáciu v rozhodovaní slúži rámec **4D** z kurzu **Anthropic AI Fluency**. Pri každej úlohe sa kladú štyri otázky:

Zložka (4D)	Otázka a príklad v dátovej analýze
Delegation (Delegovanie)	Čo možno zveriť AI a čo nie? Kód na histogram áno; rozhodnutie, ktorá otázka je dôležitá, nie.
Description (Popis / prompt)	Ako sformulovať požiadavku? Uviesť dáta, cieľ, nástroj a formát výstupu.
Discernment (Rozlišovanie)	Ako rozpoznať chyby? Skontrolovať kód, predpoklady testu, graf cez checklist.
Diligence (Validácia)	Ako overiť správnosť? Spustiť kód, porovnať s ručným výpočtom, overiť zdroje.

SENIOR VS. JUNIOR - PREČO TO SPOMÍNAME

AI sľubuje, že veľkú časť času preskočí. Lenže skúsený analytik **skraca** cestu, ktorú pozná - vie, ktoré kroky smie vynechať. Začiatočník **niekedy nevie, že nejaká cesta vôbec existuje**, a preskočí časti, ktoré ešte nepochopil. Keď AI preskočí pochopenie problému za analytika, vznikajú falošné istoty. Preto je vhodné najprv pochopiť postup ručne, až potom ho zrýchľovať pomocou AI.

5 Úlohy pre študentov

Nasledujúce úlohy precvičujú kritické hodnotenie AI výstupov. Pri každej je uvedený cieľ, postup a kritériá hodnotenia. Úlohy možno riešiť samostatne alebo vo dvojiciach.

Úlohy zámerne celiť na časté zlyhania v oblasti analýzy dát: nevhodný výber testu, falošné asociácie z chybného spárovania premenných a vymyslené zdroje.

Úloha 1 - Audit AI kódu

Cieľ: Overiť, či vygenerovaný kód robí naozaj to, čo zadanie pýta.

Postup:

1. Zadať AI prompt na napísanie skriptu (napr. odstránenie outlierov metódou IQR).
2. Prečítať kód riadok po riadku a opísať vlastnými slovami, čo robí.
3. Spustiť ho na malom datasete so známym správnym výsledkom.
4. Nájsť aspoň jednu chybu alebo riziko (zastaraná funkcia, edge case, neefektívnosť).

KRITÉRIÁ HODNOTENIA

- Správnosť identifikácie toho, čo kód reálne robí.
- Otestovanie na kontrolnom príklade.
- Nájsť a zdôvodnená aspoň jedna chyba alebo riziko.

Úloha 2 - Detektív manipulatívneho grafu

Cieľ: Aplikovať kontrolný checklist na AI graf.

Postup:

1. Nechať AI vytvoriť graf z poskytnutých dát.
2. Prejsť body checklistu (os Y, škály, typ grafu, čo graf neukazuje) a pri každom zaznačiť zistenie.
3. Identifikovať aspoň dve manipulačné techniky (alebo riziká) v grafe.
4. Navrhnuť opravený, poctivý graf tých istých dát.

KRITÉRIÁ HODNOTENIA

- Použitie celého kontrolného checklistu.
- Správne pomenované manipulačné techniky.
- Návrh poctivejšej alternatívy s odôvodnením.

Úloha 3 - Korelácia verzus kauzalita

Cieľ: Rozpoznať, kedy AI zamení súbežnosť za príčinu.

Postup:

1. Predložiť AI dvojicu korelovaných premenných a požiadať o interpretáciu.
2. Posúdiť, či AI tvrdí príčinnú súvislosť tam, kde ide len o koreláciu.
3. Navrhnuť možný tretí (mäťúci) faktor.
4. Sformulovať korektnú, opatrnú interpretáciu.

KRITÉRIÁ HODNOTENIA

- Správne odhalenie zámeny korelácie za kauzalitu.
- Zmysluplný návrh tretieho faktora.
- Korektná a opatrná preformulácia záveru.

Úloha 4 - Overenie zdrojov a integrita

Cieľ: Skontrolovať, či AI necituje neexistujúce zdroje, a zamyslieť sa nad akademickou integritou.

Postup:

1. Požiadať AI o interpretáciu výsledkov s odkazmi na odbornú literatúru.
2. Overiť, či každý uvedený zdroj reálne existuje a tvrdí to, čo AI uvádza.
3. Označiť halucinované citácie.
4. Napísať krátku reflexiu: kde je hranica medzi legitímnou pomocou AI a podvodom?

KRITÉRIÁ HODNOTENIA

- Overenie všetkých citácií.
- Odhalené halucinované zdroje.
- Premyslená reflexia o akademickej integrite.

6 Obhajoba projektu - rubrika k prezentácii

Semestrálny projekt dnes takmer vždy vzniká s pomocou AI - a to je v poriadku. **Cieľom prezentácie preto nie je dokázať, že analýzu napísal sám študent, ale že jej rozumie.** Pri obhajobe sa hodnotí, či študent vie vysvetliť a zdôvodniť, čo jeho analýza robí a prečo. Nasledujúca rubrika zhrňa, čo má silná obhajoba ukázať.

Kritérium	Čo má študent pri obhajobe ukázať
Pochopenie postupu	Vie vlastnými slovami vysvetliť každý krok analýzy, nielen prečítať kód. Na otázku „prečo tu robíš toto?“ má odpoveď.
Voľba metódy	Zdôvodní, prečo zvolil danú metódu či test a či sedia jeho predpoklady (nie „lebo to navrhla AI“).
Vizualizácia	Grafy prejdú kontrolným checklistom (os Y od nuly, poctivé škály, vhodný typ, čo graf neukazuje).
Rola AI	Transparentne povie, kde AI pomohla a kde musel zasiahnuť vlastným úsudkom; uvedie aspoň jedno miesto, kde AI navrhla niečo nesprávne alebo nevhodné.
Overenie a zdroje	Doloží validáciu (porovnanie, kontrolný príklad) a overené zdroje - žiadne halucinované citácie.
Praktická významnosť	Interpretuje výsledok v kontexte problému, nielen ako p-hodnotu alebo „štatisticky významné“.

PRE VYUČUJÚCEHO

Rubrika presúva ťažisko hodnotenia z otázky „kto to napísal“ na otázku „kto tomu rozumie“. Otázky pri obhajobe sa zameriavajú na zdôvodnenie rozhodnutí (prečo táto metóda, prečo táto kategória), nie na memorovanie syntaxe.

7 Na záver

Generatívna AI robí z analýzy dát zdanlivo jednoduchú úlohu - stačí dáta „hodiť“ do nástroja a o pár minút sú na svete pekné grafy aj plynulý záver. Lenže pekný graf z AI ešte nie je dátová analýza.

Rýchly výstup je lákavý a občas naozaj užitočný. Podmienkou však je vedieť, čo sa od analýzy očakáva, vedieť sa AI správne pýtať a overiť, či spravila to, čo sa žiadalo. AI totiž vracia mimoriadne sebavedomý výstup, ktorý môže byť úplne mimo: dokáže vynechať podmienky, ktoré sa zdajú samozrejmé - a niekedy ich obíde aj vtedy, keď sú zadane explicitne, lebo pri prvom probléme ich potichu zahodí. Výsledkom sú pekné grafy a logicky znejúci text, ktorý nestojí na pravde.

Dôsledok je dvojaký: v lepšom prípade hodiny strácané overovaním nezmyselných „insightov“, v horšom rozhodnutia postavené na halucinácii. Práve preto je tým, čo z človeka robí analytika, kritické myslenie, nie ovládanie nástroja.

NA ZAMYSLENIE

„Pekný graf z AI ešte nie je dátová analýza.“

Prístup „hod' to do AI“ je lákavý, no výstup bez pochopenia je len rýchlejšia cesta k omylu - o to nebezpečnejšia, že vyzerá presvedčivo.

8 Odporúčané zdroje a nástroje

AI nástroje: ChatGPT, Claude, Gemini - vhodné na generovanie kódu a prvý návrh analýzy.

Dôveryhodné zdroje dát: data.gov.sk, Eurostat, Our World in Data.

Odporúčaná literatúra k téme

AI v analýze dát:

- Evaluating the Accuracy and Explanatory Quality of Large Language Models (ChatGPT, Claude, DeepSeek, Gemini, Grok, Le Chat) in Statistical Test Selection for Hypothesis Testing Decisions (2025). Cureus.
- LLM-Based Data Science Agents: A Survey of Capabilities, Challenges, and Future Directions (2025). arXiv:2510.04023.
- Evaluating large language models for selection of statistical test for research: a pilot study (2024). PMC.
- Data Analysis in the Era of Generative AI (2024). arXiv:2409.18475.
- What Should Data Science Education Do with Large Language Models? (2024). Harvard Data Science Review, 6(1).
- Dakan, R. - Feller, J. (2025). AI Fluency: Framework & Foundations. Anthropic.
- A Survey of Context Engineering for Large Language Models (2025). arXiv:2507.13334.