

Evaluatie van Moderne AI-tools voor Data Scientists

Inleiding en Onderzoeksvraag

Context: Generatieve AI-tools worden steeds vaker gebruikt in data science. Toch ontbreekt systematisch onderzoek naar de nauwkeurigheid van recente en populaire modellen bij data cleaning en EDA.

Onderzoeksvraag: Kunnen AI-tools betrouwbare assistenten zijn voor data cleaning en EDA?

Methodologie

AI-tools: ChatGPT, Claude, DeepSeek, Grok.

Cases:

- Data cleaning (Titanic).
- Visuele & numerieke EDA (Penguins).

Evaluatie: nauwkeurigheidsscore (0-2) + kwalitatieve beoordeling.

Resultaten

Data Cleaning

Nauwkeurigheidspercentage

ChatGPT
& Claude

100%

ChatGPT correct maar beknopt;
Claude meer uitleg en controles.

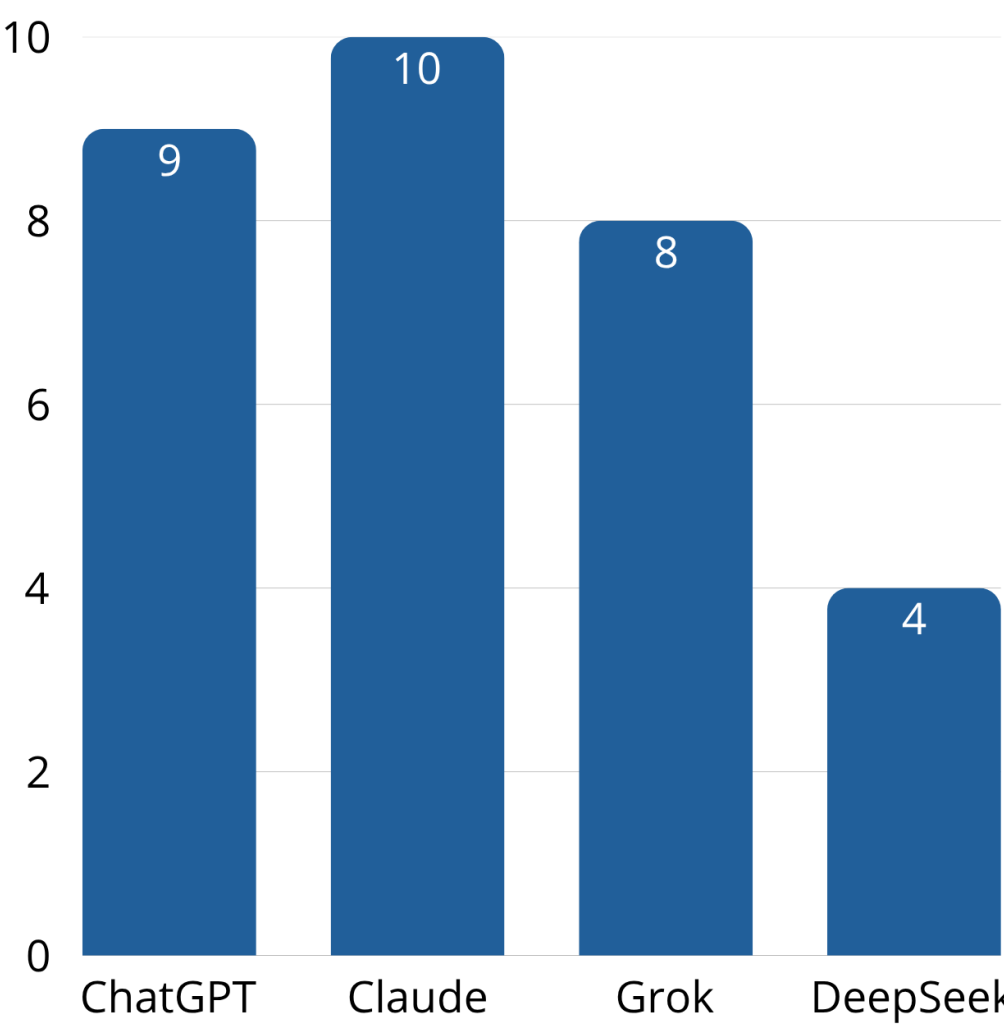
Grok
& DeepSeek

88%

Grok en DeepSeek faalden bij
complexere code.

Visuele EDA

Nauwkeurigheidsscore op 10



- Claude levert de meest complete en correcte interpretaties.
- DeepSeek toont structurele zwaktes bij visuele analyses.

Numerieke EDA



Alle cases
succesvol
afgerond

- ChatGPT: correct en bondig
- Claude: correct en bondig
- Grok: correct en uitgebreid
- DeepSeek: correct en uitgebreid

Conclusie

- Claude presteert het beste.
- AI-tools zijn inzetbaar als assistenten bij data cleaning en EDA.
- Kritische gebruikerscontrole blijft cruciaal.

Beperkingen & Aanbevelingen

- Beperkt tot 2 datasets, 1 programmeertaal (R) & 1 beoordelaar.
- Geen onderzoek naar gebruikerservaring of interactie.
- Toekomstig onderzoek: complexere data, menselijke vergelijking & interactie.



UHASSELT