

A Few Comments on AI & Data Science



AI and Data Science: The Data?

- Requires lots of (public) data

<https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt>

facebook  Twitter  LinkedIn  Rydo  Wetter  Regenradar  Digital Manual | ID.4 

Samsung workers made a major error by using ChatGPT

By Lewis Maddison published 6 days ago

Samsung meeting notes and new source code are now in the wild after being leaked in ChatGPT



- Is your data part of it and... should it?

Does a Language Model Recognize Data?



Based on the information provided, it appears that the data set consists of four numerical features (col1, col2, col3, col4) and a categorical class label. Without any additional context or information about the nature of the data, it is difficult to identify any specific patterns or trends in the data.

To answer your original question, yes, as a language model, I am capable of recognizing and finding patterns within data. However, since I am not specifically designed for data analysis, my abilities in this area may be limited compared to specialized tools or software designed for this purpose. Nonetheless, I can provide guidance on general approaches and techniques for pattern recognition and data analysis.



Can you also build a decision tree?



Yes, I can build a decision tree. A decision tree is a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is commonly used in machine learning and data mining to help solve classification and regression problems. Would you like me to build a decision tree for you? If so, please provide me with more information about the problem you want to solve.

I meant a decision tree for the data I just shared!



Sure! I can definitely help you build a decision tree for your data.

First, let's define the target variable for your decision tree. Based on your data, it seems like you want to predict whether a person will purchase a car or not. Therefore, the target variable for your decision tree will be the "Car Purchase" column.

AI for Data Science: Generalization vs. Hallucination

- Learns from existing data points

But can my end user also do this interactively during execution? I don't want to store my own auth within the node, I'd rather have it that the consumers provide their own

Yes, you can allow your end user to provide their own Google authentication interactively during the execution of the KNIME workflow. To do this, you can use the "Interactive Authentication" node instead of the "Google Authentication" node in your workflow. This node will prompt the user to enter their Google account credentials when the workflow is executed, and the entered credentials will be used to authenticate the API requests.

In this way, **...use the "Interactive Authentication" node...**
you don't have to store any authentication information in the node. This provides a higher level of security and flexibility to your KNIME workflow.

- Generalization assumes that reality generalizes well (does it?)
- Not to forget: Generalization always includes Discrimination (some of it is bad)

AI and Data Science: Sharing

AI doesn't give credit.

- Will people be willing share as before?
- Back to more locked & non public content
- “SEO is dead”?

AI for Data Science – Some Early Thoughts:

Likely useful

- For education purposes (“how do I do that?” – mostly...)
- For guidance (“why don’t you try this next”)
- For skipping the annoying 60% (“can you build something like this...”)
- For sanity checking processes (“do you find something wrong with this...”)

Likely less useful

- For novel insights into new data
- For highly sparse information / data
- For hypothesis generation (The “that’s strange...”-effect)
- For the analysis of proprietary, private, competitively relevant data

Forecasting the Future

*"In [...] three to eight years
we will have a machine
with the general intelligence
of an average human being."*

Marvin Minsky, 1970

Cool Stuff ;-)



would you recommend KNIME?



Yes, KNIME is a powerful and popular open-source data analytics platform that provides a wide range of functionalities for data preparation, data blending, modeling, and visualization. It also has a user-friendly graphical interface that allows users to create, execute, and manage data workflows easily. KNIME supports a large number of data formats, including CSV, Excel, XML, JSON, and databases, and it also integrates with several machine learning libraries such as TensorFlow and scikit-learn. Additionally, KNIME has a large community of users and developers, which means that you can find many resources and support if you encounter any issues or have questions.

