AI Society

Week 8: AI Student

[March 27, 2025]

Use case is king – not tech

Project themes

Chatbot for x

ML/DL model for doing y

Data ingestion pipeline for z

Project suggestions

Chatbot for helping homeless in Stockholm

Upload files to Google Drive, perform processing and store cleaned data in database

Regression using XGBoost

Fashion-MNIST classification

Chatbot for homeless in Stockholm

1. Write a simple system prompt telling the LLM to answer in Swedish and try it out with an example prompt.
2. Think about 10 questions a homeless person might have and adjust the ability of the system and user prompt template in such a way that the chatbot can answer these questions
3. Choose one of 3 options:

- Build a web page where you can chat to the chatbot
- Use the streaming endpoint to get continuous text output
- Add an llm call that looks at the question given to the chatbot and deem if the question is about programming. In that case, the chatbot should deterministically respond that it is a chatbot not made to help with programming. Example: user- “Write a python script for saying hello world” -assistant “Jag kan inte hjälpa dig med frågor relaterade till programmering”

Predicting Concrete Compressive Strength using XGBoost

1. Read up on concrete
2. Do exploratory data analysis to see if what you understand about concrete.
3. Split data into 80/20 train test split
4. Perform predictions
5. In which cases are our predictions bad? What do these cases have in common and how can we give the model more information?
6. Tune hyperparameters

Data:
https://drive.google.com/file/d/122c4O8vwdpNX_KxkWko54k1H5hGNhYUI/view?usp=sharing

Data engineering – ETL

1. Upload csv to Modal
2. Go through the data, figure out which company hires the most “principal” engineers (look for the keyword “principal” inside the job title column)
3. Schedule a job in Modal that fetches the data everyday, takes out all of those jobs, saves it as Parquet to a new location on Modal with the name <date>-principal.parquet

Data:
https://drive.google.com/drive/folders/1V5ZKMklj43cnD50qPKqhFaMQQFfa1kcm?usp=sharing

Deep learning – clothes classification

1. Get the data from https://github.com/zalandoresearch/fashion-mnist
2. Train a convolutional neural network to work on the dataset
3. In which cases are our predictions bad? What do these cases have in common and how can we give the model more information?
4. Tune hyperparameters
5. Optionally try out a different architecture (VGG, ResNet)

Email **timothylindblom@kthais.com**

Submission is preferably a github link otherwise a zip folder.

Mandatory:

README.md that explains what the project is about and how to run it.