Credit:Esoon Ko
IT Högskolan

Project ver:1.0

Group work: 2-3 individuals in group

Two possible tasks for you to choose from:

Group work: 2-3 individuals in group

Two possible tasks for you to choose from:

TASK 1 (Grade G)

Scenario:

You are a new Data Team at a small firm Something Weather AB (SWAB).
The firm has built an ML model as a POC with a dataset they purchased.
They want your team to build the company's data platform so that their ML model can use the data from the platform for future training. They also want your team to deploy the model so that their end user can use the model.
Your team is given a deadline of 9 october to both launch the products and hold a presentation on your products, workflow process and why you did things the way you did.

For grade G:

- Build a data platform that takes in data from the given public API and processes data for consumption by the ML model for future training or analysis. The platform must have schedules for the pipelines as well as possibility to monitor and logging.
- Deploy the given ML model to the cloud for use by end user.

Other Requirements:

- Use agile methods during the project. Document the proof of the methods you used.
- Containerize services through docker or other containerization technology such as Kubernetes.

Group work: 2-3 individuals in group

Two possible tasks for you to choose from:

TASK 1 (Grade VG)

Scenario:

You are a new Data Team at a small firm Something Weather AB (SWAB).

The firm has built an ML model as a POC with a dataset they purchased.

They want your team to build the company's data platform so that their ML model can use the data from the platform for future training. They also want your team to deploy the model so that their end user can use the model.

Your team is given a deadline of 9 october to both launch the products and hold a presentation on your products, workflow process and why you did things the way you did.

For grade VG:

- Build a data platform that takes in data from the given public API and processes data for consumption by the ML model for future training or analysis. The platform must have schedules for the pipelines as well as possibility to monitor and logging.
- Deploy the given ML model to the cloud for use by end user.
- Find an additional relevant data API and create an additional data pipeline. In other words you need to have 2 data source APIs. Use the data from this new data source in combination with the existing data in order to gather new insight through a dashboard. Optionally train a new ML model that uses both data sources. Examples include: Bike traffic data. The data DOES NOT HAVE TO BE RELATED, IE THE INSIGHT GAINED MAY BE "No relation found between the data."

Other Requirements:

- Use agile methods during the project. Document the proof of the methods you used.
- Containerize services through docker or other containerization technology such as Kubernetes.
- Apply CI/CD techniques and automated testing + linting into your project.

Group work: 2-3 individuals in group

Two possible tasks for you to choose from:

TASK 1

Materials:
https://drive.google.com/drive/folders/1ArHYeS7L3rRXM_Sgmst1sDkmSVDmDPXV?usp=sharing

Code:
Github Repo - https://github.com/esoonko/Data-engineering-AI23
lectures/code/project/ml-model

Group work: 2-3 individuals in group

Two possible tasks for you to choose from:

TASK 2 (Grade VG)

Scenario:

You are a new Data Team at a small firm {INSERT YOUR FICTIONAL COMPANY AB}.

The firm wants to build ML functionality as part of their business model.

They want your team to create a ML POC with any data API of your choice.

They want your team to also build the company's data platform so that their ML model can use the data from the platform for future training. They also want your team to deploy the model so that their end user can use the model.

Your team is given a deadline of 9 October to both launch the products and hold a presentation on your products, workflow process and why you did things the way you did.

For grade VG:

- Use any data API of your choice to gain insight and train a new ML model to solve a problem you identified.
- Build a data platform that takes in data from the given public API and processes data for consumption by the ML model for future training or analysis. Create analysis dashboard from this data. The platform must have schedules for the pipelines as well as possibility to monitor and logging.
- Deploy the given ML model to the cloud for use by end user.

Other Requirements:

- Use agile methods during the project. Document the proof of the methods you used.
- Containerize services through docker or other containerization technology such as Kubernetes.
- Apply CI/CD techniques and automated testing + linting into your project.

Group work: 2-3 individuals in group

Additional Requirements/Information

- You may use Cloud Run with docker as container or use Cloud Run Function as it itself is a form of containerization.
- You are allowed to create databases manually but the database schema should be saved as evidence. If you used a script to create the tables the script itself suffice.
- Logging and monitoring is required but usually comes pre-packaged in airflow or GCP. Show a screenshot of logging and monitoring environment.
- For VG you must have at least one test on one data pipeline - example: Ingestion pipeline should have a test that runs automatically via GitHub actions or Cloud Build.

Group work: 2-3 individuals in group

Hand in

- All code should be downloaded and zipped. Include the .git folder. Everything should be in the same zip.
- The presentation slides.
- Screen shot your Jira/Kanban/Agile framework tool in order to document project work.
- 1 A4 document with description that describes what the project does and why certain methods and tools are chosen. This is to demonstrate your reasoning and motivation of work method.
- Screen shot your environment, for example airflow or Google Cloud Platform components

Group work: 2-3 individuals in group

Presentation - Aim for 10 min

Have the following information:
- What task you aimed to solve and how you decided to solve the problem.
- Display the tools and methods you used as well as architecture for your solution.
- Show how the project was worked on by the group. Show your Jira/Kanban/agile framework tool to help you with this.
- Show a demo on your solution

7th and 9th October. You may choose which date to present.

Group work: 2-3 individuals in group

Pair up with people that have:

- Similar grade ambition
- Location