
**SINGAPORE POLYTECHNIC
SCHOOL OF COMPUTING**

DIPLOMA IN APPLIED AI AND ANALYTICS (DAAA)

Module Code	:	ST1516
Module Name	:	DEVOPS AND AI AUTOMATION
Semester	:	2023/2024 Semester 2
Assignment Name	:	CA2 MLOPS FOR DL WEB APPLICATION
Assignment Type	:	Individual
Weighting	:	40%
Deadline	:	Part A - 2 Feb 2024 (Friday), 09:00 hrs Part B - 13 Feb 2024 (Tuesday), 09:00 hrs
Mode of Submission	:	Online Submission via Brightspace. Please follow the instructions in Brightspace on how to submit.

LATE SUBMISSION

50% of the marks will be deducted for assignments that are received within ONE (1) calendar day after the submission deadline. No marks will be given thereafter.

Exceptions to this policy will be given to students with valid LOA on medical or compassionate grounds. Students in such cases will need to inform the lecturer as soon as reasonably possible. Students are not to assume on their own that their deadline has been extended.

PLAGIARISM

No marks will be awarded, if the work is copied or you have allowed others to copy your work. SP's academic policy on Copyright and Plagiarism applies. Please refer to this link: <https://www.sp.edu.sg/sp/student-services/osc-overview/student-handbook/intellectual-property-copyright-and-plagiarism>

MLOPS FOR DEEP LEARNING WEB APPLICATION

BACKGROUND

MLOps stands for Machine Learning Operations. It's a primary function of Machine Learning engineering, focusing on streamlining the process of taking machine learning models to production, and then maintaining and monitoring them. MLOps is a collaborative function, often comprising data scientists, devops engineers and IT.

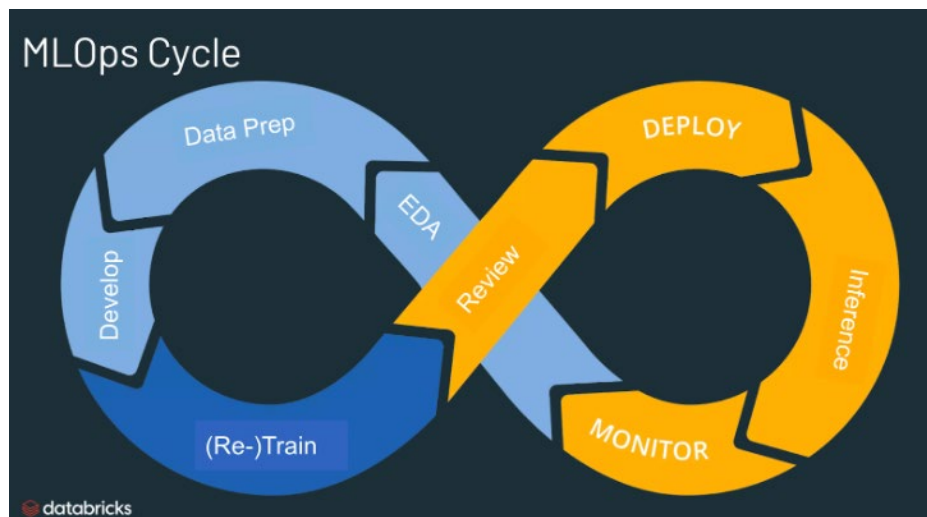


Image Credit: Databricks (<https://www.databricks.com/glossary/mlops>)

Data scientists and ML engineers can collaborate and increase the pace of model development and production, by implementing continuous integration and deployment (CI/CD) practices.

OBJECTIVE

To develop a web application for Deep Learning using Model Serving, GitLab, Docker and Flask Framework, incorporating MLOps best practices (see above). The web application will be used to demonstrate the particular Deep Learning concept based on the application domain specified in the next section.

APPLICATION DOMAIN

You are to reuse the dataset and algorithms that you have chosen for your CA1 assignment in ST1504 DELE (Part A - CNN) as the application domain. Note: there is no requirements to further improve the models' accuracy that you have already done as part of the CA1 assignment submitted, if your models' accuracy are at least 70%.

ASSIGNMENT REQUIREMENT

Part 1 (GitLab and SCM)

For this part, you are required to do the following:

1. Set up a GitLab project using your GMail account. The naming convention for your project should start with: CA2-[class]-[AdminID]-[Name]. The class naming convention is: DAAA2B0X.
Note: Failure to adopt the naming convention will result in 2 marks deduction
2. Set up a SCRUM board for your development process.
3. Setup at least 3 branches in the GIT repository. The Main branch must be merged after each branch is completed.

Part 2 (Model & Web Application Development using Docker)

For this part, you are required to do the following:

1. Develop a Deep Learning (DL) model serving for inference using Docker, based on the application domain. You can consider other serving platforms or libraries beside Tensorflow Serving.
2. Deploy 2 models used for inference such that you can switch between the models in your web application.
3. Design an input function for your web application such that users can interact with your web application, make use of the model and see the outputs generated by the model.
4. Design and create wireframe(s) for your web application
5. Make use of Flask Framework to develop your web application as container using Docker
6. Store the prediction history with search and filter functions
7. Present the prediction history with a good level of information in a table form
8. Create a simple login credential to secure your web application

Part 3 (Automatic Testing)

For this part, you are required to do the following:

1. Setup Unit testing and it should cover at least 3 of the following:
 - a. Validity testing
 - b. Range testing
 - c. Consistency testing
 - d. Unexpected failure testing
 - e. Expected failure testing
2. Create REST APIs for crucial operations and prediction functionalities
3. Testing on all the REST APIs created

Part 4 (Continuous Integration and Deployment)

You are required to perform the following for this section:

1. Deploy the web application and model serving to internet using one of the available free-tier platforms:
 - Render
 - Other platform
2. Demonstrate how do you perform continuous integration and deployment (CI/CD)

Part 5 (Robotic Process Automation – Optional)

This part is optional and served as bonus. You are to research and find out how to leverage on Robotic Process Automation (RPA) tools like UiPath or Selenium to perform Testing Automation.

For instance, you can consider using UiPath to perform automated GUI testing as part of your test suite. For more information, please refer to this link

<https://www.uipath.com/blog/rpa/gui-automation-key-to-automating-interfaces>

Part 6 (Presentation)

You are required to perform the following for the presentation:

1. Create a Powerpoint presentation (maximum slide: 25).
2. Do a 10-minute presentation using PowerPoint slides and record it as video in MP4 format.
3. Demonstrate the key features of your web application

Note:

- You must present your work and demo to ascertain the authenticity of your work is in order. Zero marks will be awarded for all other components if the authenticity of your work cannot be ascertained.
- You may be asked to do an interview with your tutor if your tutor has some questions from your presentation in the video recording.

SUBMISSION PART A (Model Development)

You are required to submit the following requirement for Part A submission:

- The Deep Learning (DL) models used for inference, containerized using Docker and deployed to the internet.
- The deployment of 2 models used for inference such that you can switch between the models in web application.
- Document the deployment process in the Jupyter Notebook.

Procedure:

1. Submit a ZIP file containing the following deliverables with the naming convention: CA2-[class]-[AdminID]-[Name].zip - the class naming convention is: DAAA2B0X. Failure to adopt the naming convention will result in 2 marks deduction.
 - Jupyter Notebook and its HTML exported version
2. Submit your assignment using the CA2 Part A assignment link under the Assignment folder in Brightspace. Please remember to include your student name and admission number on the top of your Jupyter Notebook.

Note:

- Late submission policy applies and will be accounted towards the model development component of the evaluation criteria.

SUBMISSION PART B

You are required to submit the remaining requirement stated in this assignment:

1. Submit a ZIP file containing the following deliverables with the naming convention: CA2-[class]-[AdminID]-[Name].zip - the class naming convention is: DAAA2B0X. Failure to adopt the naming convention will result in 2 marks deduction.
 - Powerpoint slide in both PPTX and PDF format
 - Code repository (excluding the env folder)
 - Video recording in MP4 format (maximum duration: 10 minutes)
2. Submit your assignment using the CA2 Part B assignment link under the Assignment folder in Brightspace. Please remember to include your student name and admission number on the top of your Jupyter Notebook and in the title slide of your Powerpoint.

EVALUATION CRITERIA

You are encouraged to use your creativity and apply the concepts in this assignment. If you simply clone or adapt the web application from the module practicals, low mark will be awarded.

No	Category	Description	Marks
1	DevOps Process	<ul style="list-style-type: none"> • Gitlab and branches setup • Setup of Scrum Board 	10
2	Model Development	<ul style="list-style-type: none"> • Models Deployment • Setup of container using Docker 	15
3	Web Application Development	<ul style="list-style-type: none"> • Web Application Development • Prediction history setup • Login credential • Setup of container using Docker 	35
4	Automatic Testing	<ul style="list-style-type: none"> • Setup of relevant test cases • Web API setup for testing 	10
5	MLOps	<ul style="list-style-type: none"> • Demonstration of CI/CD 	15
6	Presentation	<ul style="list-style-type: none"> • Powerpoint slide submission • Presentation & demonstration 	15
7	Robotic Process Automation (Bonus)	<ul style="list-style-type: none"> • Demonstrate use of RPA <p>Note: total mark capped at 100</p>	10
		TOTAL	100