International Programs Department

Data Science in Production

Description

This course is about defining, scoping, and implementing an AI project and its industrialization. It will include the theoretical principles useful for project leaders as well as future project team members. It will include how to implement a methodology in compliance with the agreed scope, schedule, budget, and quality requirements, to meet customer expectations. The course addresses mostly the technical aspects of projects and how to industrialize the production of its applications.

Learning Objectives and Outcomes

• Know how to deliver AI models and expose them to other system.

• Setup continuous validation of model performance against new data.

• Apply MLOps principles.

• Know CI/CD and TDD can be applied in a AI oriented project.

Course Schedule and Contents

• Session 0 : Lead an AI project

- identifying the target

- deal with complexity, ethics and explain ability

- identify the major issues regarding an AI project (availability of data, feasibility, time to design model, infrastructure costs) and put them in balance with the benefits.

• Session 1 : Observing real world problems regarding AI models lifecycle

- Reliability issues (data structures are changing with time in an unexpected way)

- Performance issues (the model is not as good as it was on new data)

- Identify the key steps, and design a pipeline to automate retraining

- Setup automated non-regression test to have early feedback on breaking changes

• Session 2 : Optimize your time during development phases

- try to find the time-consuming operations during design phases and avoid to repeat them

- keep track of different executions, by versioning the code and keeping a board with the results (to check improvement and regressions)

• Session 3 : Conventional tools for software production & general purpose automation tools

- CI/CD tools (Jenkins)
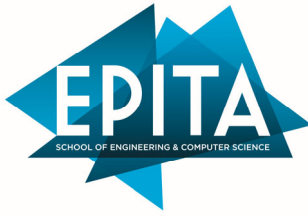
- batch tools (Luigi)

• Session 4 : Practice on session 2 using an AI lifecycle management tool

- local design tracking example

- expose model using an API

• Session 5 : Practice on session 1 & 3 Continuous retraining of ML models example.

- Setup the infrastructure

- Configure the steps and build the project

Assignment will be the continuation of the practical work engaged in session 5 & 6

Grading

Quiz : 20 %

Practical Work : 30%

Assignment : 50%

Policies

- I expect you to turn-in your reports on time to receive proper credit/grade.
- Any work submitted must be your own.
- I expect everyone to contribute equally to group assignments
- Attendance in every class is expected and class participation and discussion is strongly encouraged.
- Late work will not be accepted unless prior arrangements have been made directly with me.
- Cases will be decided on an individual basis.

Good Luck!