

## Deployment (anything REST + batch pipeline)

- Task description:
  1. Prepare model for deployment in two modes: online (as REST API) and batch (as scheduled pipeline)
  2. For REST use any tool you feel comfortable with (flask, TFServing, cloud helpers, MLFlow Models - anything that's REST)
  3. For batch use workflow manager and simple scheduler (cron)
  4. Batch and online prediction should use the same artifacts (model, transformers, etc.).
- Criteria:
  1. Model deployment is straightforward
  2. REST endpoint is stable and response time is manageable ( $\ll 1$  sec)
  3. Batch predictions: works.
  4. All code is packed as a python package
  5. Perform at least one test (for example, check input/output data or your code)
  6. \*last task of batch pipeline producing a ready to use Docker image for online prediction.
- Materials:
  1. RestAPI:
    - a. <https://flask.palletsprojects.com/en/2.0.x/>
    - b. <https://blog.keras.io/building-a-simple-keras-deep-learning-rest-api.html>
    - c. <https://www.mlflow.org/docs/latest/models.html#deploy-mlflow-models>
  2. Python packaging tutorials:
    - a. <https://packaging.python.org/tutorials/packaging-projects/>
    - b. <https://python-packaging.readthedocs.io/en/latest/minimal.html>
    - c. [https://python-packaging-tutorial.readthedocs.io/en/latest/setup\\_py.html](https://python-packaging-tutorial.readthedocs.io/en/latest/setup_py.html)
    - d. <https://blog.ionelmc.ro/2014/05/25/python-packaging/>
  3. Testing:
    - a. <https://github.com/ericmj1/data-testing-tutorial>
    - b. <https://docs.pytest.org/en/6.2.x/>
    - c. <https://docs.python.org/3/library/unittest.html>