

Lab: Capstone

Capstone Project

Now apply everything we learned so far yourself.

That will include:

- * Thinking of a problem that's interesting for you and finding a dataset for that
- * Describing this problem and explaining how a model could be used
- * Preparing the data and doing EDA, analyzing important features
- * Training multiple models, tuning their performance and selecting the best model
- * Exporting the notebook into a script
- * Putting your model into a web service and deploying it locally with docker
- * Bonus points for deploying the service to the cloud

Datasets

- * <https://www.kaggle.com/datasets> and <https://www.kaggle.com/competitions>
- * <https://archive.ics.uci.edu/ml/index.php>
- * <https://data.europa.eu/en>
- * <https://www.openml.org/search?type=data>

Deliverables

For this project, your repository/folder should contain the following:

- * `README.md` with
 - * Description of the problem
 - * Instructions on how to run the project
- * Data
 - * You should either commit the dataset you used or have clear instructions how to download the dataset
- * Notebook (suggested name - `notebook.ipynb`) with

- * Data preparation and data cleaning
- * EDA, feature importance analysis
- * Model selection process and parameter tuning
- * Script `train.py` (suggested name)
 - * Training the final model
 - * Saving it to a file (e.g. pickle)
- * Script `predict.py` (suggested name)
 - * Loading the model
 - * Serving it via a web service (e.g. with Flask)
- * `Pipenv` and `Pipenv.lock`
 - * or equivalents: conda environment file, requirements.txt or pyproject.toml
- * `Dockerfile` for running the service
- * Deployment
 - * URL to the service you deployed or
 - * Video or image of how you interact with the deployed service

Function for computing the hash of your email:

```
```python
from hashlib import sha1

def compute_hash(email):
 return sha1(email.lower().encode('utf-8')).hexdigest()
```
```

Tip: you can use <https://nbviewer.org/> to render notebooks if GitHub doesn't work