Application template Small-Scale Initiatives in Machine Learning

(OpenSSI 2021)

Training and consultancy

Title: *Short title for this proposal*

Lead applicant (LA): *The name of the lead applicant.*

Position and Affiliation LA: *Used to determine the eligibility of the LA*

Email address LA: *This address will be used for all official communication.*

Postal address LA: *For official correspondence with the LA*

Start date: 1 May 2021

End date: *not later than 31 December 2021*

Requested RSE-PYRs: 0.25 FTE *(default and maximum)*

Discipline area: *select the one that fits the best*

- ❑ *Physical Sciences and Engineering* (mathematics, physics, astrophysics, chemistry, engineering, computer science, earth sciences, climatology, environmental sciences, agricultural sciences)
- ❑ *Life Sciences* (biology, medical and health sciences incl. neuroscience)
- ❑ *Social Sciences and Humanities* (economics, political sciences, law, sociology, anthropology, education sciences, communication studies, psychology, linguistics, philosophy, literary studies, archaeology, history, human geography)

Research team

	Name	Email	Position	Affiliation	Expertise	Availability for this project [hours/week]	Will attend the workshop 17-21 May 20201
Lead Applicant							YES
Research Team member 1							YES
Research Team member 2							YES
Research Team member 3							YES
Research Team member 4							YES

Research Question (max. 250 words)

Describe the research question you are planning to work on. Provide some context for the importance and/or impact of a successful outcome of this proposal.

Current state of the art (max. 250 words)

Describe the current state of the art in your domain wrt. this research topic: is machine learning being used? or other types of statistics? Is this problem typically ignored, or are there simplifying approximations?

Description of available data

For each dataset that you are considering using, please provide the following:

Dataset name	
Short description	*max. 50 words*
Type of data	*text, images, audio, video, numerical, categorical, mixed…*
Format of the data	*html, txt, csv, pdf, HDF, NetCDF, …*
Size of the dataset	*approximate number of records: 1000s, millions, ..?*
Size of the dataset	*approximate size in bytes / MB / GB*
More information	*website, if available*

And please confirm that:

- ❑ The datasets will be available at the start of the project.
- ❑ The datasets can be used for this research (AVG/GDPR, copyrights, …).

Programming experience (max 100 words)

Briefly sketch the programming experience within your team.

Description of envisioned machine learning solution and deliverables (max 250 words)

Designing the machine learning solution for your research question is part of the consultancy project, and you are not expected to provide a detailed plan here. However, if you already have an idea of the best approach, network design, or required expertise, you can describe it here.

Additionally, briefly describe the deliverables (for instance title and topic):

- *1 paper and*
- *1 blogpost and*
- *code release.*