

Capstone Project Proposal Template

Notes:

- This should take no more than one hour to complete – the clearer you are about the business problem you're working to solve with your ML-driven solution, the easier your proposal will be to complete
- This will be uploaded to your repo, which will be a part of your final submission
- Due date for submission is end-of-day 3/13 for Cohort 3b

Instructions:

1. Download this document as a Word Doc
2. Answer each question using a few sentences, at most
3. Save your completed proposal as a PDF
4. [Create a project GitHub repo](#) (if you have yet to do so)
5. [Add your instructor as a collaborator](#) (username `jvntra`) to your project repo
6. Add your mentor as a collaborator
7. Push your proposal PDF (created in Step 3) up to your repo
8. Copy the URL corresponding to the location of the PDF in your repo
9. Submit the copied URL using [this link](#) for Cohort 3b

[Deep learning for particle classification in High energy Physics]

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
The main goal of the project is to provide an automated solution implementing deep learning to classify elementary particles in particle collisions in the LHC.
- What industry/realm/domain does this apply to?
It will be of use mainly for experimental physics laboratories.
- What is the motivation behind your project?
I have just finished a bachelor's degree in physics, and I would like to dedicate the future of my career in experimental particle physics. This project is intended to be one of my first's contributions to the field, putting in practice what I have learn so far in the AI Academy and in college.

Data Understanding

- What data will you collect?
Particle collision images that have already been classified with statistical tools and other sorts of methods and are available to everyone in the CERN's database.

- Is there a plan for how to get the data (API request, direct download, etc.)?
It is possible to download the files directly from CERN's Open Data web page or from Kaggle.
- What are the features you'll be using in your model?
Physical properties measured by different particle detectors such as momentum, energy, mass, etc.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
I would have to explore the data to decide if the number of features will be a problem in terms of processing, and if they are, it will be necessary to run a PCA to reduce dimensionality. Also, data will need to be standardized. Also, identify linear dependencies between variables to select a linearly independent set of features.
- What are some of the cleaning/pre-processing challenges for this data?
To this moment it is not clear to me what transformations will the data require, but the main problem will be understanding what the data means and reducing the scope of the project to make it feasible.

Modeling

- What modeling techniques are most appropriate for your problem?
I will address the problem with a neural network. However, the architecture is not clear yet.
- What is your target variable?
The type of elemental particle (this is what we want to identify)
- Is this a regression or classification problem?
It is a classification problem, but it may need some regression analysis to identify linear dependencies.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
It will depend on the selected method, but mainly RMSE, and Loss functions.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
The intention is to use deep neural networks, but decision trees and other unsupervised learning methods (such as K-means) will be tested to determine which will work better.