

Problem Definition

- Regression/Classification?
- Supervised/Unsupervised/semi-supervised
- Performance Measure? MSE, RMSE, other measures defined by problem-related standard
- Check the literature for what they have done, and best practices on (Model selection & data preprocessing)



Decide on System Arch:

- PC? Cloud? Fog? Edge? Mix?



Getting Data

- Search for publicly available datasets on Kaggle, UCI library, and github repo's,
- Reliability of the data
- Reproducibility of the results
- reflecting on out choice of MIMIC



Data Preprocessing

- Data Cleaning/Cleansing, replace/get rid of missing values
- depending on the problem and the data, may require filtration, denoising, cropping ...
- Filter outliers (may be also after extracting features)



Feature Extraction

- again, problem dependent
- Number of features affect complexity of the problem which is also limited by Edge devices



Inference on The Edge

- Sensors?
- MUC selection (Memory & CLK requirements)
- Running experiments & getting edge results (how is the performance affected?)
- Inference time & memory requirements for each of the algorithms



Model Conversion

- Hand-crafted solution
- Commercial & Open Source frameworks for model-to-code generation



Training ML/DL models on PC

- Get results, judge the behaviour of the algorithms (under/over fitting)
- Judge the accuracies and other performance metrics, is it acceptable by defined standards?
- Regularization & Parameter tuning to boost performance



Choosing ML Models

- Problem dependency
- data-dimensionality dependency
- Edge devices sets constraints on feasible models



Prepare Train & Test Sets

- Shuffle data for better generalization in both train & test
- Transformation pipeline, what you can/can't do on Edge devices?