1. Detecting Cardiovascular disease. (Domain is medical, either single person or list of people go in. Prediction of cardiovascular disease comes out)
2. Categorization. Numeric inputs can be used to train weights.
3. Patient/hospitals trying to save money/governments distributing tests/insurance companies. Good model for telling who may be at higher risk for the disease. Tests for cardiovascular diseases can be expensive or scarce; we can save money by only testing the people at higher risk. Patient can try to eliminate own risk factors. (Leshan 1 – 3)
4. Noise in data, size of dataset, poor correlation of features with target, low number of features, generalization, model might overfit, data issues (nans, outliers, …) (Maram 4-6)
5. Kaggle <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>. 70k instances of data with 11 features (Getting data N\A because kaggle set)
6. Number of layers, learning rate, epochs, batch size, activation function, Error rate stop condition. Fine tune in cross fold validation.
7. Predicting at 80% threshold. Or identify highly indicative variables. (Noah 7-8)
8. Timeline. Have project meetings where we do the coding (possibly an effective use of pair programming)

* Oct 24: Meet to Record Proposal
* Nov 10: Proposal report
* Nov 10: Data preprocessing (probably minimal because kaggle normally has good data sets) (categorical one hot encoding, see if there are nans we need to deal with, potentially feature selection, train-test-validation split 70-20-10?)
* Nov 13: Implement MLP forward direction (don’t worry about hyperparamters)
* Nov 16: Implement backward propagation (make sure we have all hyperparameters)
* Nov 18: Evaluation/validation
* Nov 18-27: Thanksgiving break
* Nov 29: Final bugfixes
* Nov 29: Project Completed
* Dec 2: Project Report Rough Draft
* Dec 6: Project Report Submission
* Include names and emails of all team members
* Leshan coordinator