Project proposal

Students: Naor Maoz, Vadim Kachevski

Dataset description:

Our dataset is about the prediction of someone to have a heart failure based on different clinical features.

The dataset contains 11 features, interesting to see are unchangeable features such as: age or sex.

The rest of the features are generic features such as: resting blood pressure or serum cholesterol.

https://www.kaggle.com/fedesoriano/heart-failure-prediction

Questions we want to answer:

1. What will be the differences between different Machine Learning algorithms from different groups: Ensemble, Neural network, Trees, Cluster. On our data and why?
2. Will feature selection lower our error rate for our algorithms?
3. What will be the impact on our prediction when we will change the size of the Data we use in our training?
4. Will the Adaboost always perform better than the decision tree algorithm when it uses the decision tree as its classifier?
5. Will the features that are selected in the features selection are the same in all the algorithms.

5 Techniques we would want to use: Adaboost, Neural network, Decision Tree, K nearest neighbors, SVM.