Sandia National Laboratories

Exceptional service in the national interest

Managing Randomness to Enable Reproducible Machine Learning

Hana Ahmed, Jay Lofstead

June 3, 2022

SAND: 1539351

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

U.S. DEPARTMENT OF ENERGY

NNSA

About Me

1. Senior @ Scripps College
2. Math+Computer Science major, Writing minor
3. Year-Round Intern @ Sandia National Labs w/ Dr. Jay Lofstead

AGENDA

1. Randomness and reproducibility
2. Pseudo-random number generators
3. Experiments with randomness in ML
4. Conclusions and future works

Reproducibility in science and ML

Reproducibility:

"obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis" (National Information Standards Organization)

Why this is important for ML:

- Ensuring that models can be regenerated
- Understanding the variance in accuracy caused by randomness

Randomness in ML

Examples:

- Shuffling training data
- Randomized subsets of input features
- Random initial weights

Pseudo-random number generators (PRNGs): Algorithms that generate sequences of pseudo-random values.

Approach

- **3 ML algorithms:** Neural network, K-means, Naïve Bayes
- **6 datasets:** Heart Disease, Wine, Iris, Breast Tissue, Wisconsin Breast Cancer (WBC), Somerville Happiness
- **4 variables**: Random seed, Train/test ratio, Train data set, Test data set
- **PRNG:** mt19937 (Mersenne Twister)

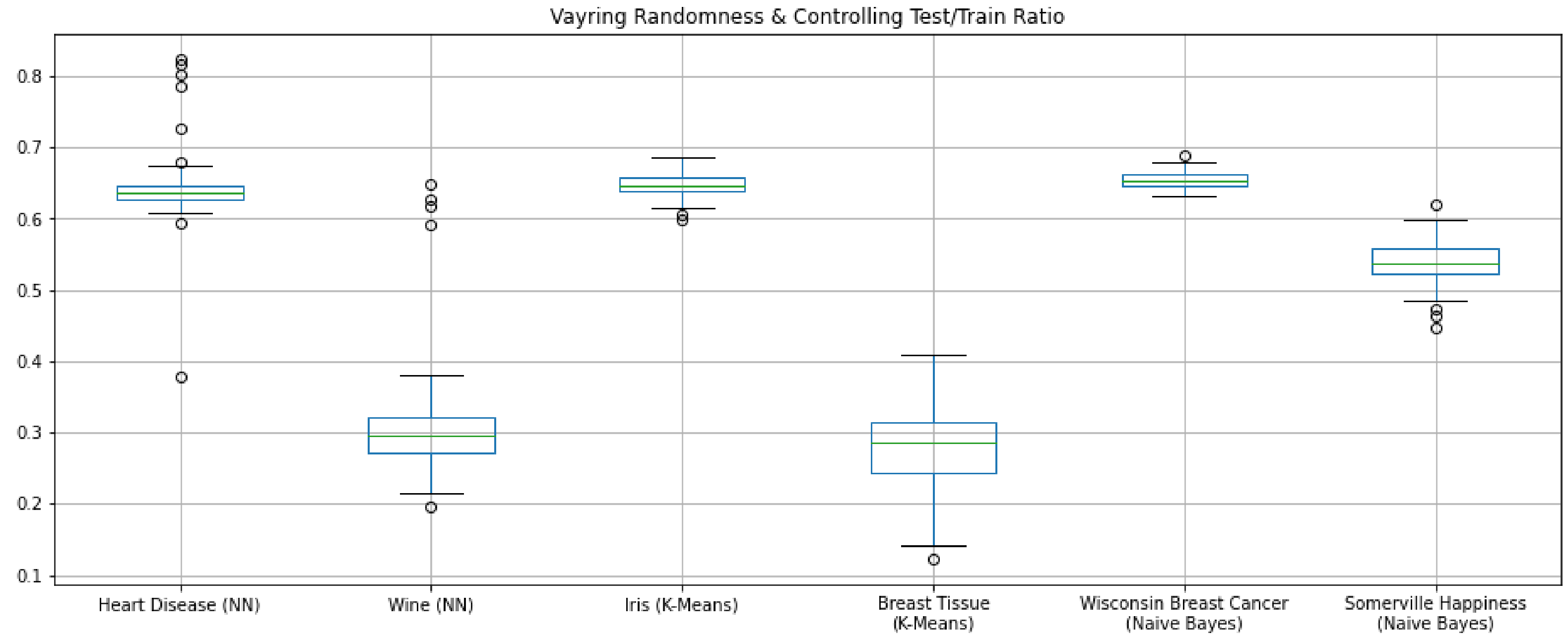
Experiment 1

Seed: **Varied**

Train/test ratio: **Fixed**

Train data: **Varied**

Test data: **Varied**



- NN: **44.53%** difference on Heart Disease, **45.17%** difference on Wine.
- K-Means: **8.88%** difference on Iris, **28.62%** difference on Breast Tissue.
- Naïve Bayes: **5.64%** difference on WBC, **17.3%** difference on Somerville Happiness.

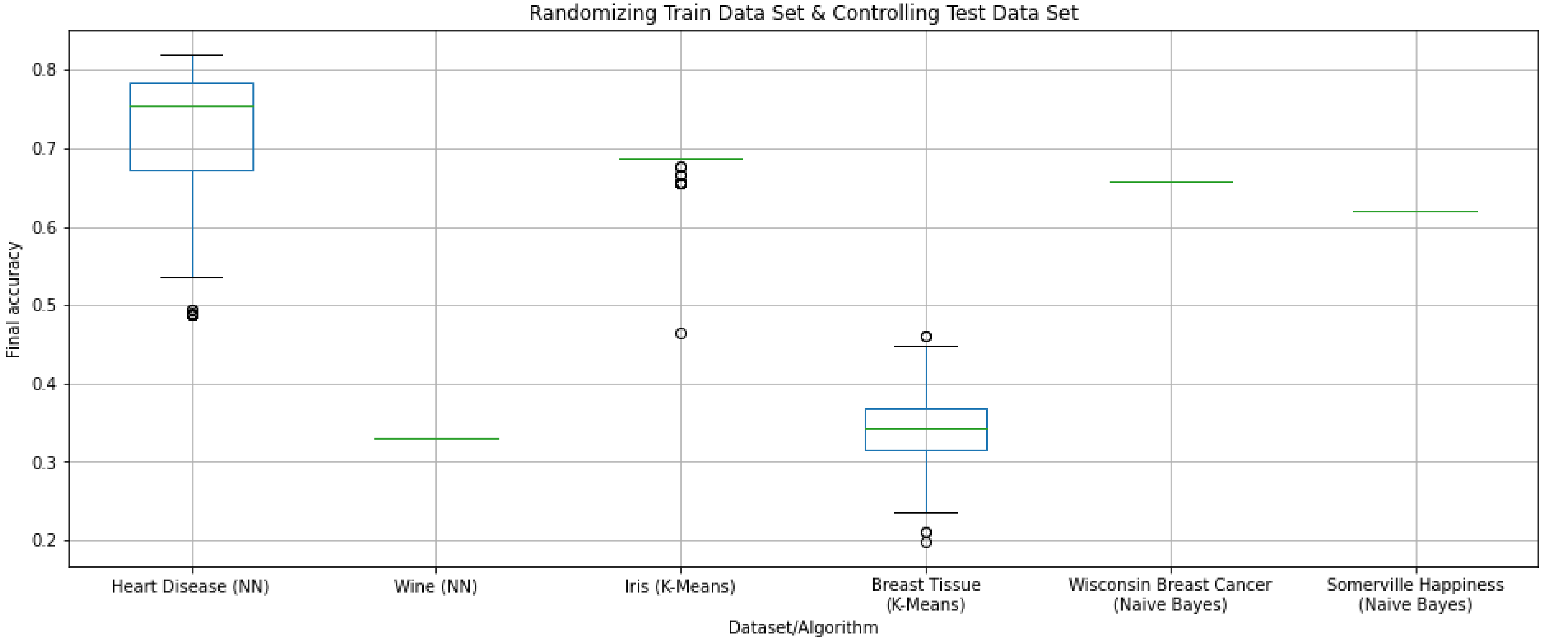
Experiment 2

Seed: **Varied**

Train/test ratio: **Fixed**

Train data: **Varied**

Test data: **Fixed**



- NN: **33.23%** difference on Heart Disease, **0%** difference on Wine.
- K-Means: **22.22%** difference on Iris, **26.31%** difference on Breast Tissue.
- Naïve Bayes: **0%** difference on WBC, **0%** difference on Somerville Happiness.

Experiment 3

Seed: **Fixed**

Train/test ratio: **Varied**

Train data set: **Varied**

Test data set: **Fixed**

Model performance varied widely for each algorithm on all data sets despite controlling the seed.

Train/Test	Heart Disease (NN)	Wine (NN)	Iris (K-Means)	Breast Tissue (K-Means)	Wisconsin Breast Cancer (Naive Bayes)	Somerville Happiness (Naive Bayes)
10/90	0.4873	0	0.6296	0.2737	0.6455	0.5703
20/80	0.4873	0	0.5833	0.3095	0.6279	0.5701
30/70	0.7476	0.3296	0.523	0.4324	0.6155	0.56
40/60	0.4873	0.3296	0.5556	0.4127	0.5776	0.5294
50/50	0.4873	0.3296	0.6667	0.0566	0.5673	0.5211
60/40	0.5663	0.3296	0.8333	0.3571	0.5878	0.5263
70/30	0.5205	0.3296	1.0	0.4688	0.5789	0.5
80/20	0.8372	0.3296	1.0	0.3810	0.5971	0.4286
90/10	0.4873	0.3296	0	0.5455	0.7101	0.5

Conclusions

- Random number seeds significantly impact the quality of ML models.
- **Current work**: understanding the potential variance range for ML algorithms for any given data set
- **Future work**:
  - Experimenting with different types of ML algorithms and data sets.
  - Applying same techniques to parallel ML algorithms
  - Running identical experiments using a GPU

Thank you.

Contact: hahmed@sandia.gov