

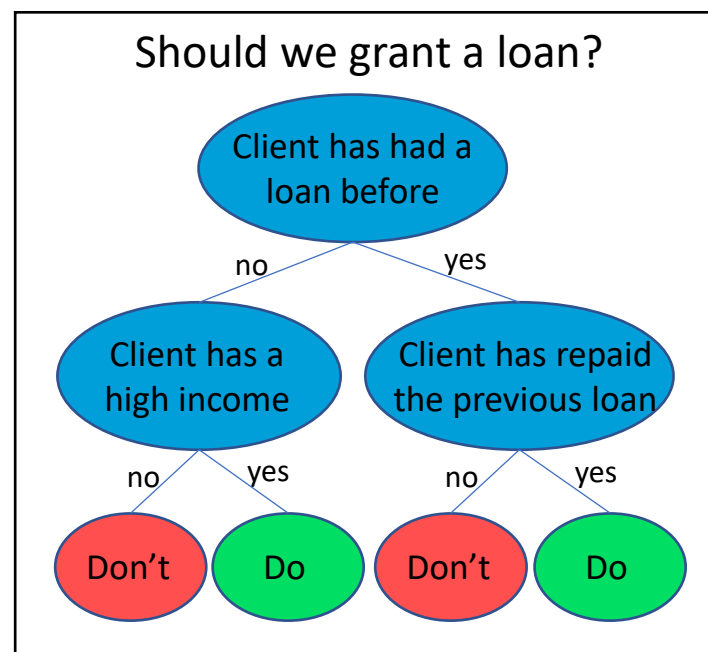
# Building random forests with optimal decision trees

**Research question:** Can optimal decision trees be used to produce a better forest, compared to the forest obtained using heuristic algorithms?

**Author:** Jord Molhoek; **Supervisor:** dr. Emir Demirović

## 1. (Optimal) Decision Trees

- Series of simple decisions
- Model is easy to interpret
- Often built using heuristics
- Globally optimize the tree
- NP-Hard
- MurTree uses dynamic programming<sup>1</sup>



## 2. Random Forests

- Simplicity of the tree can limit the performance
- Make an ensemble of trees vote on the outcome
- Train these trees with randomness:
  - Randomize the tree generation procedure<sup>2</sup>
  - Sample rows (bagging)<sup>3</sup>
  - Sample columns (random subspace method)<sup>4</sup>

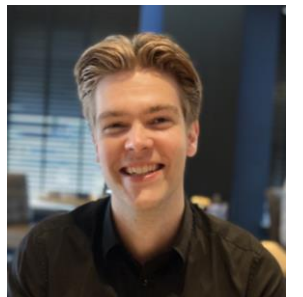
## 3. Forests of Optimal Trees

- Random root
- Sample data
- Different weights
- Decrease correlation
- Algorithm that returns all optimal decision trees, instead of one
- Performance varies from dataset to dataset

Full Text

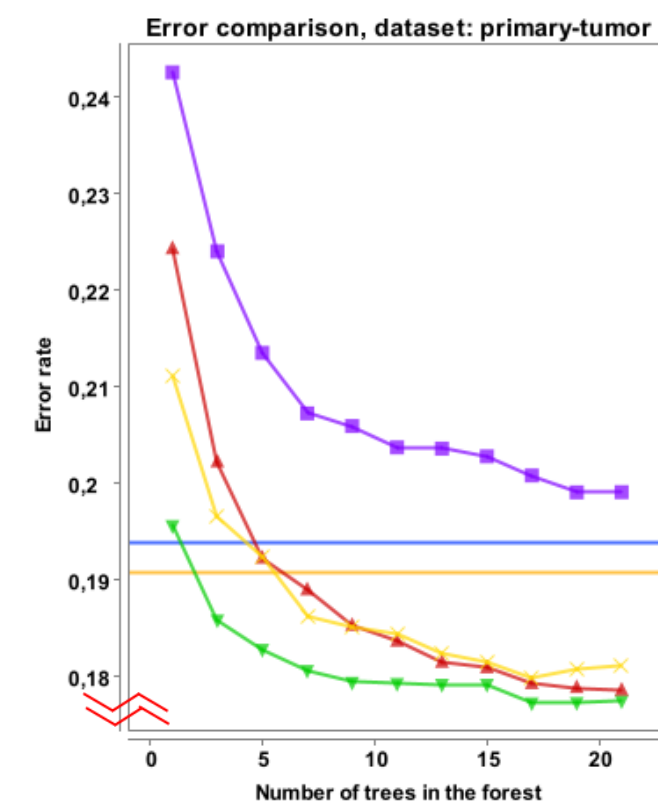
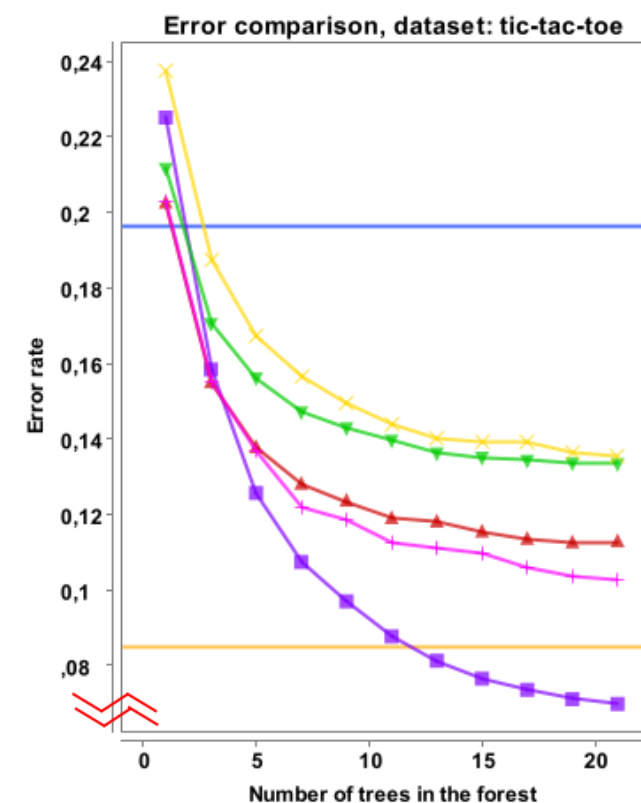
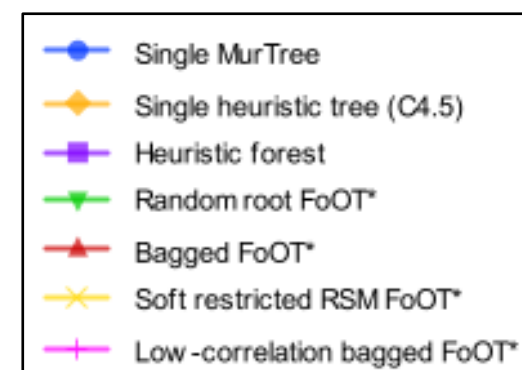


Author



Jord Molhoek  
j.molhoek@student.tudelft.nl

## 4. Results



\*FoOT = Forest of Optimal Trees  
max depth of optimal trees = 4

## References

1. Demirović et al., 2021: arXiv: 2007.12652 [cs.LG]
2. Breiman, 2001: *ML*, doi.org/10.1023/A:1010933404324
3. Breiman, 1996: *ML*, doi.org/10.1007/BF00058655
4. Ho, 1998: *TPAMI*, doi.org/10.1109/34.709601