Rashomon Ambiguity Averse Active Learning

December 5 Update

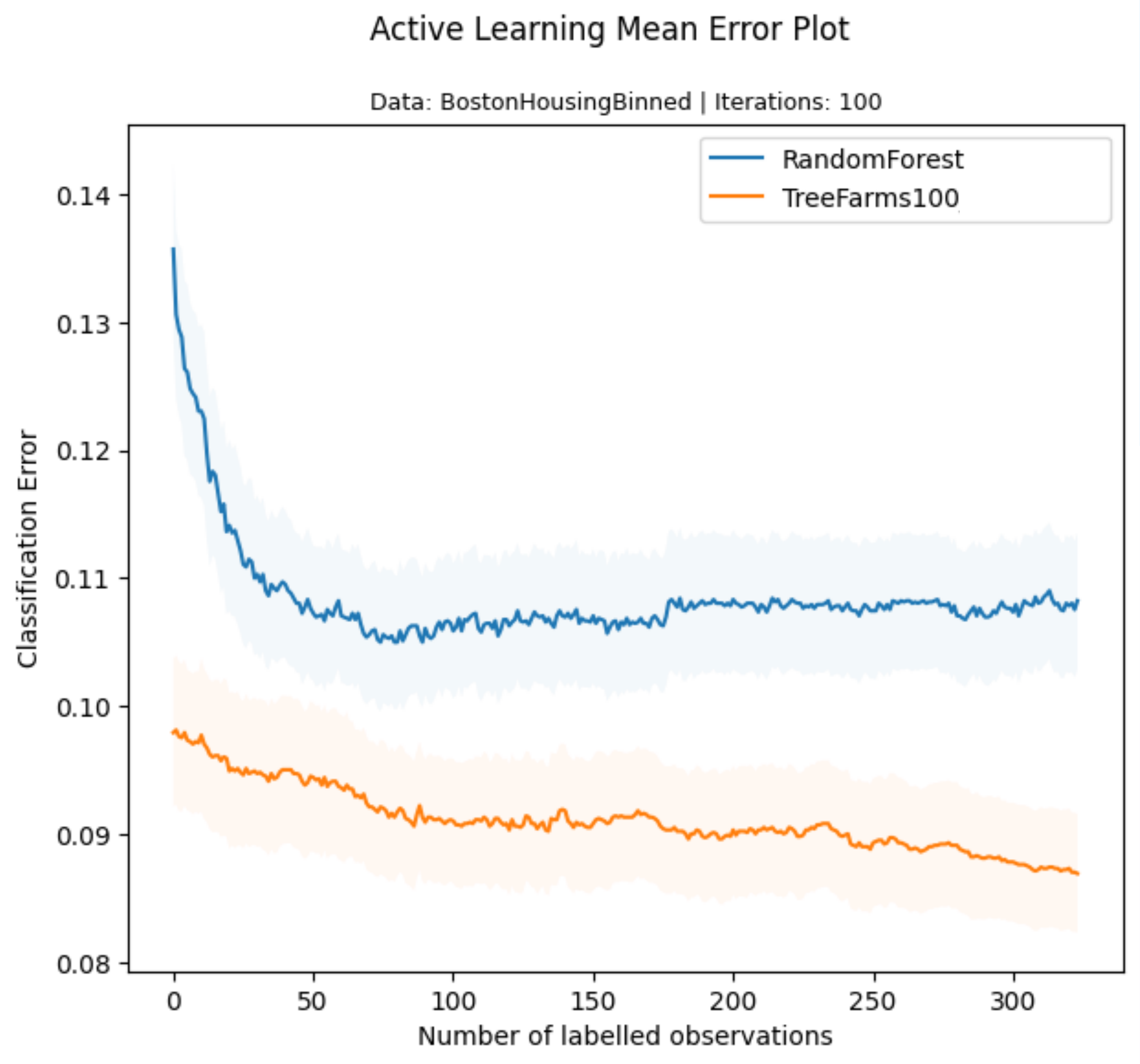
Simon Nguyen

Bad Trees in Random Forests

1. People often use Random Forests for active learning classification problems.
2. However, Random Forests ensemble a random selection of data and covariates, potentially incorporating bad decision trees.
3. This motivates the use of only good decision trees in ensemble methods.
4. The Rashomon Set of good decision trees: TreeFarms!

Active learning

- Simulation:
  - Yellow line: random forests.
  - Blue line TreeFarms with the best 100 decision trees.
- Clearly, using the best 100 decision trees is much better than ensembling all the decision trees in random forests.
- Problem solved, right?

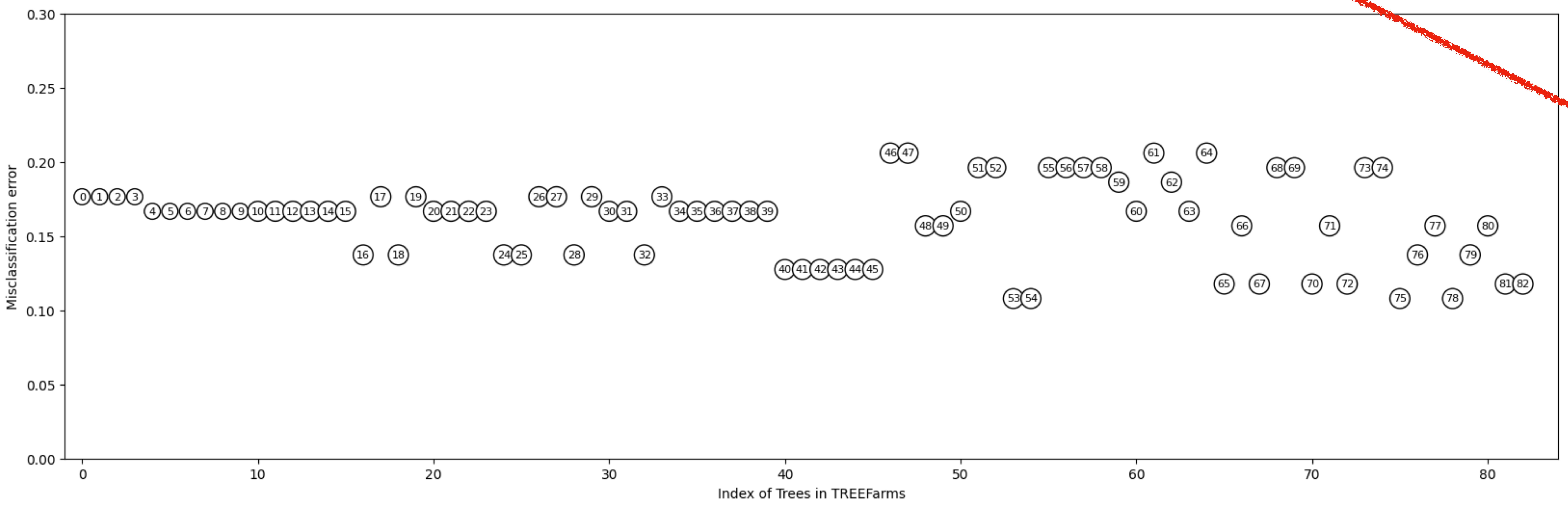


Multiplicity of Explanations in Tree Farms

1. However, TreeFarms suffers from multiplicity of explanations.
    - ie. many trees repeat the same explanation!
2. Redundancy in explanations leads redundancy in predictions.
3. This redundancy skews our notion of uncertainty.
    1. Redundancy in decision trees may overinflate agreement amongst models, leading to an underestimation of a predicted observation’s uncertainty.
    2. This will be more clear in the next couple slide.
4. The Rashomon set of decision trees from TreeFarms does not give us a good measure of predictive uncertainty.
5. Let’s take a look at what this means!

Multiplicity of Explanations in TreeFarms

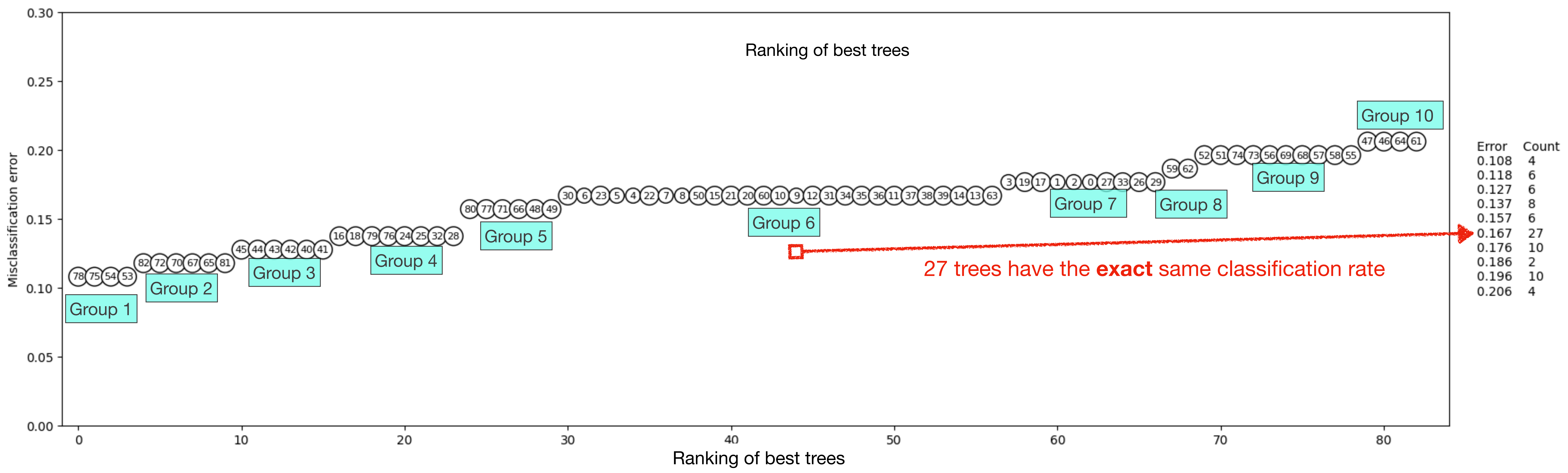
- The following contains the misclassification error rate by the index of the tree for one fitting of TreeFarms.
- Note how many trees have the same **exact** misclassification rate!
- This is indicative of trees sharing the same explanation of the data, albeit ordering covariates differently.



Error	Count
0.176	10
0.167	27
0.137	8
0.127	6
0.206	4
0.157	6
0.196	10
0.108	4
0.186	2
0.118	6

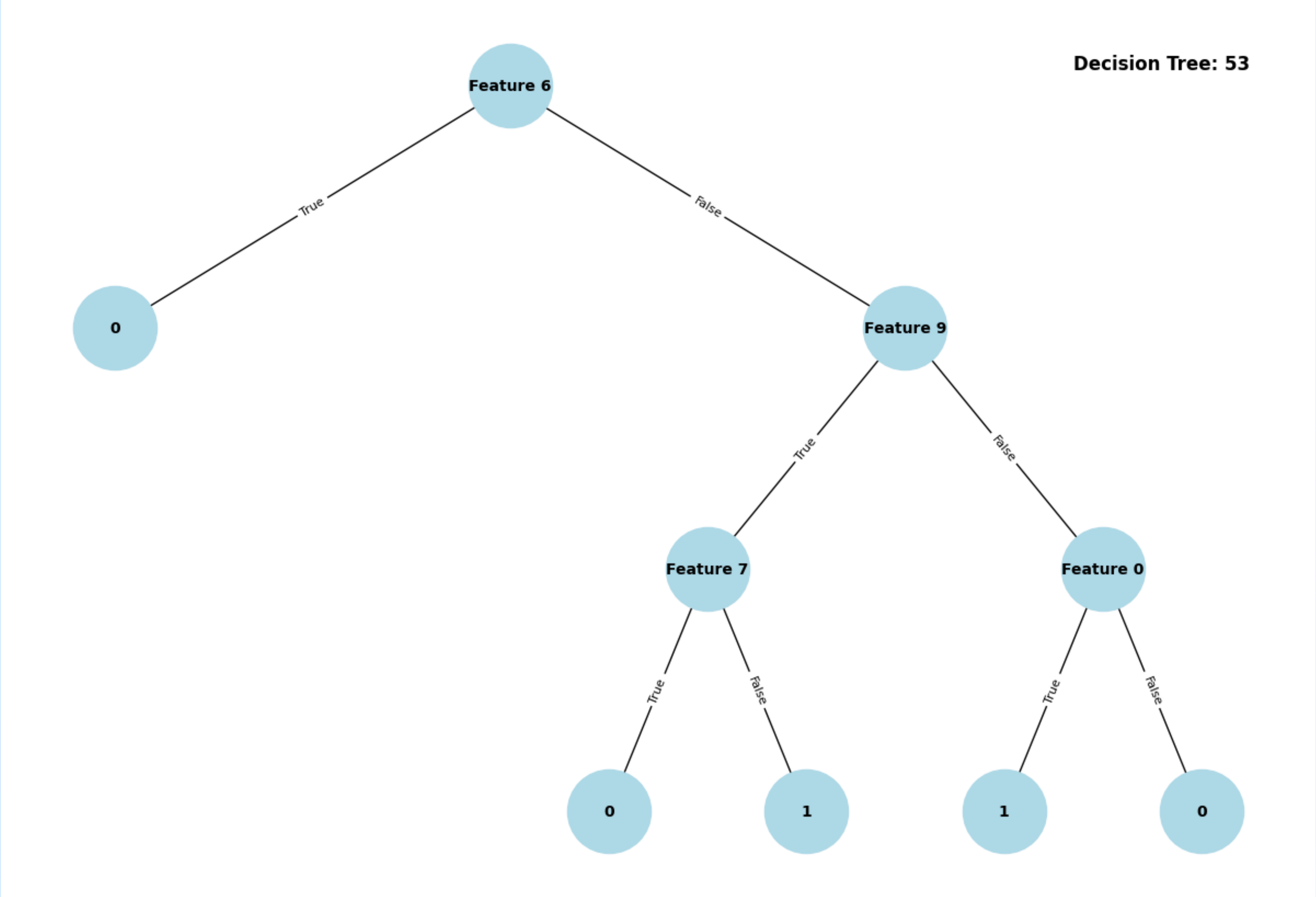
Multiplicity of Explanations in TreeFarms (Grouped)

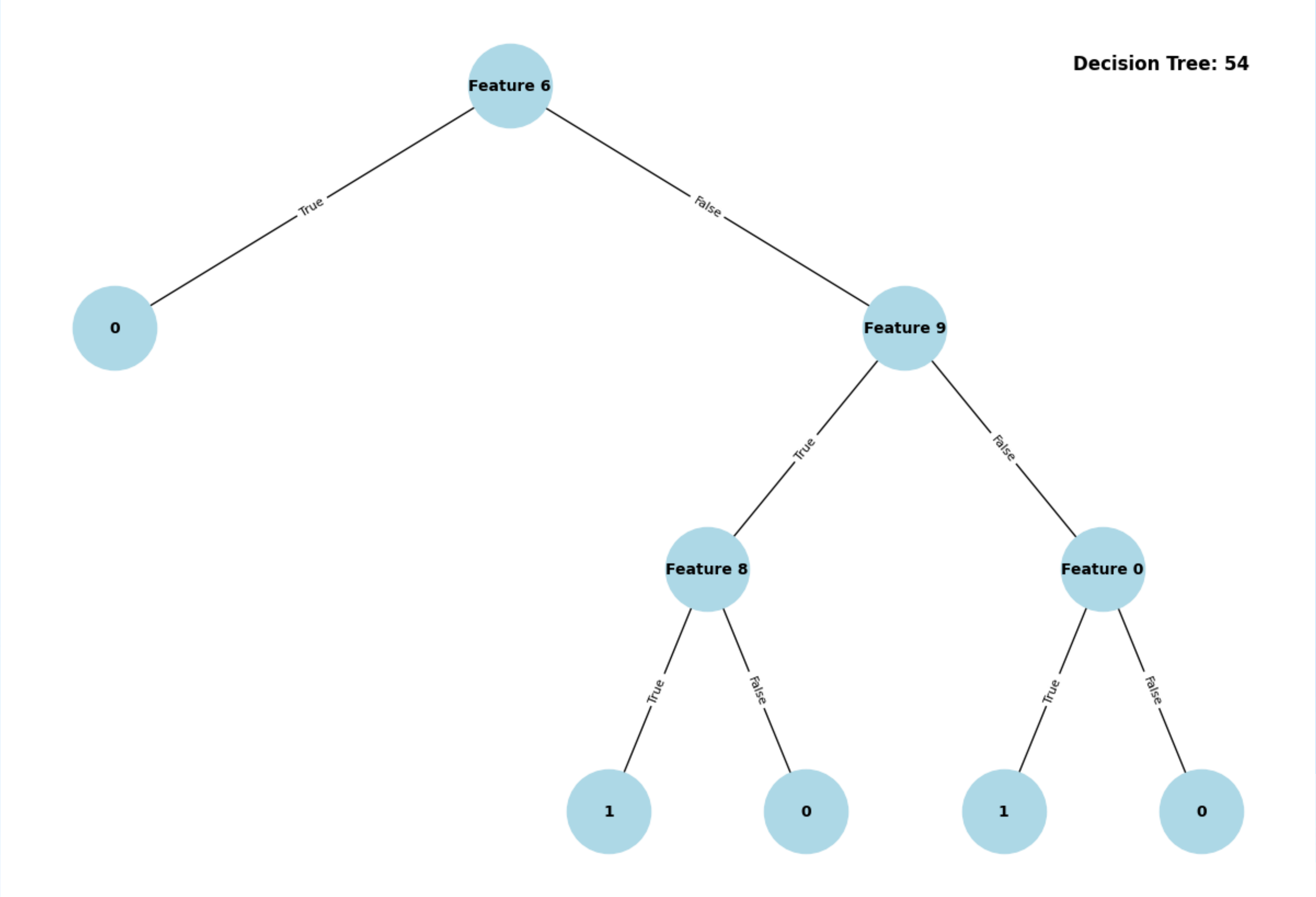
- I now group the trees by their misclassification error (y-axis).
- The tree index are noted by the number in the circle.
- Let’s take a look at what the trees in the first four groups look like!

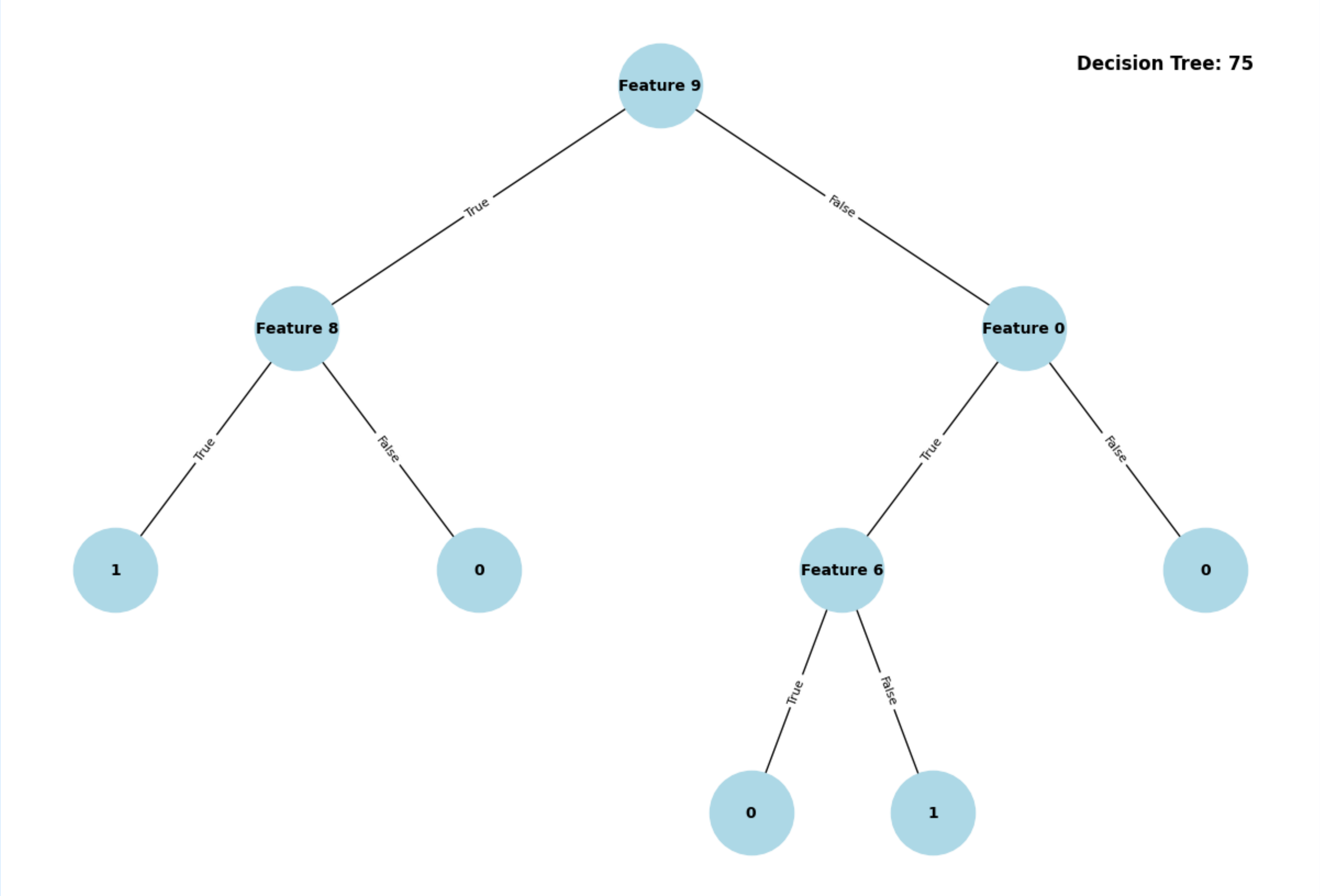


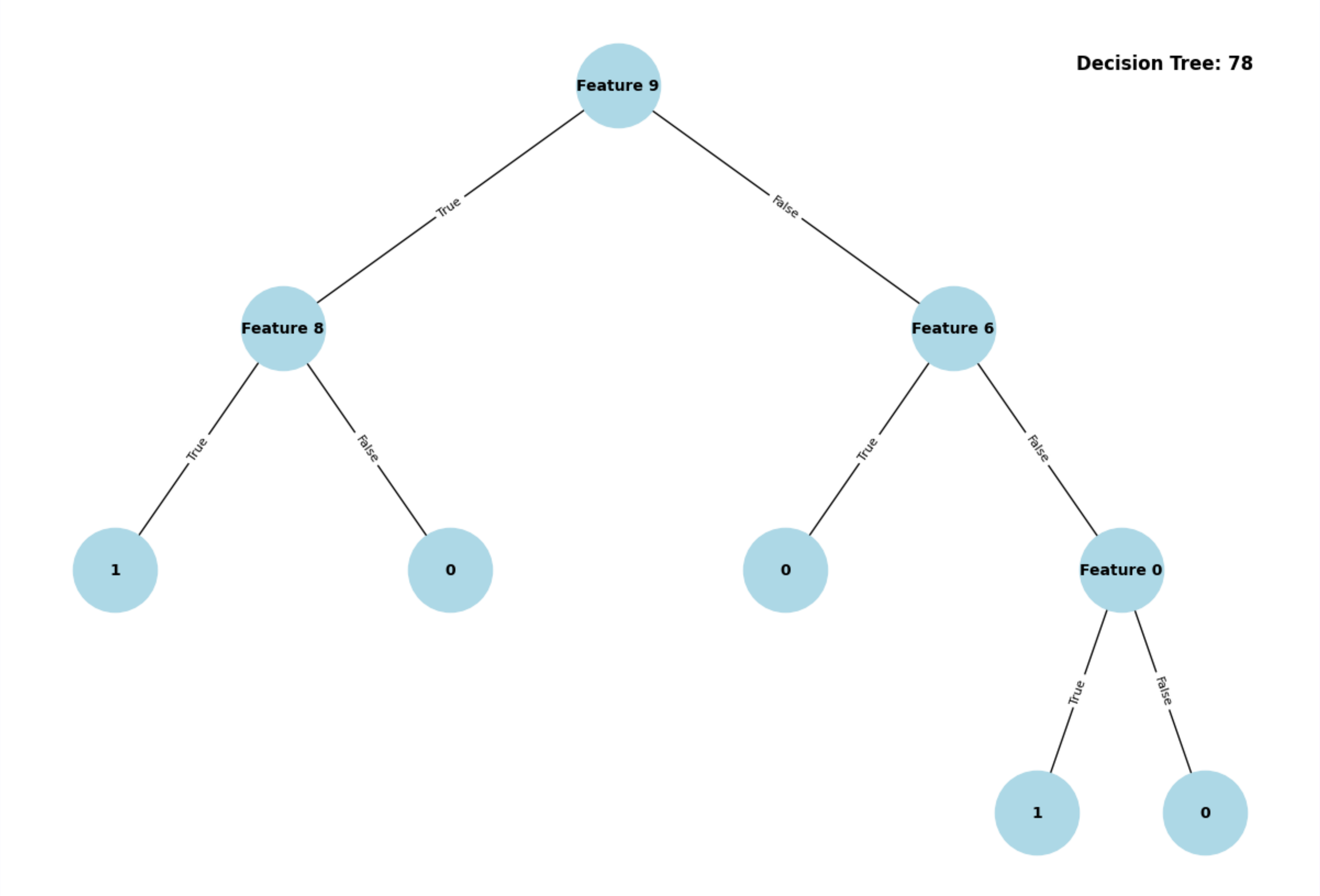
Group 1

Apologies on the lack of image quality/size

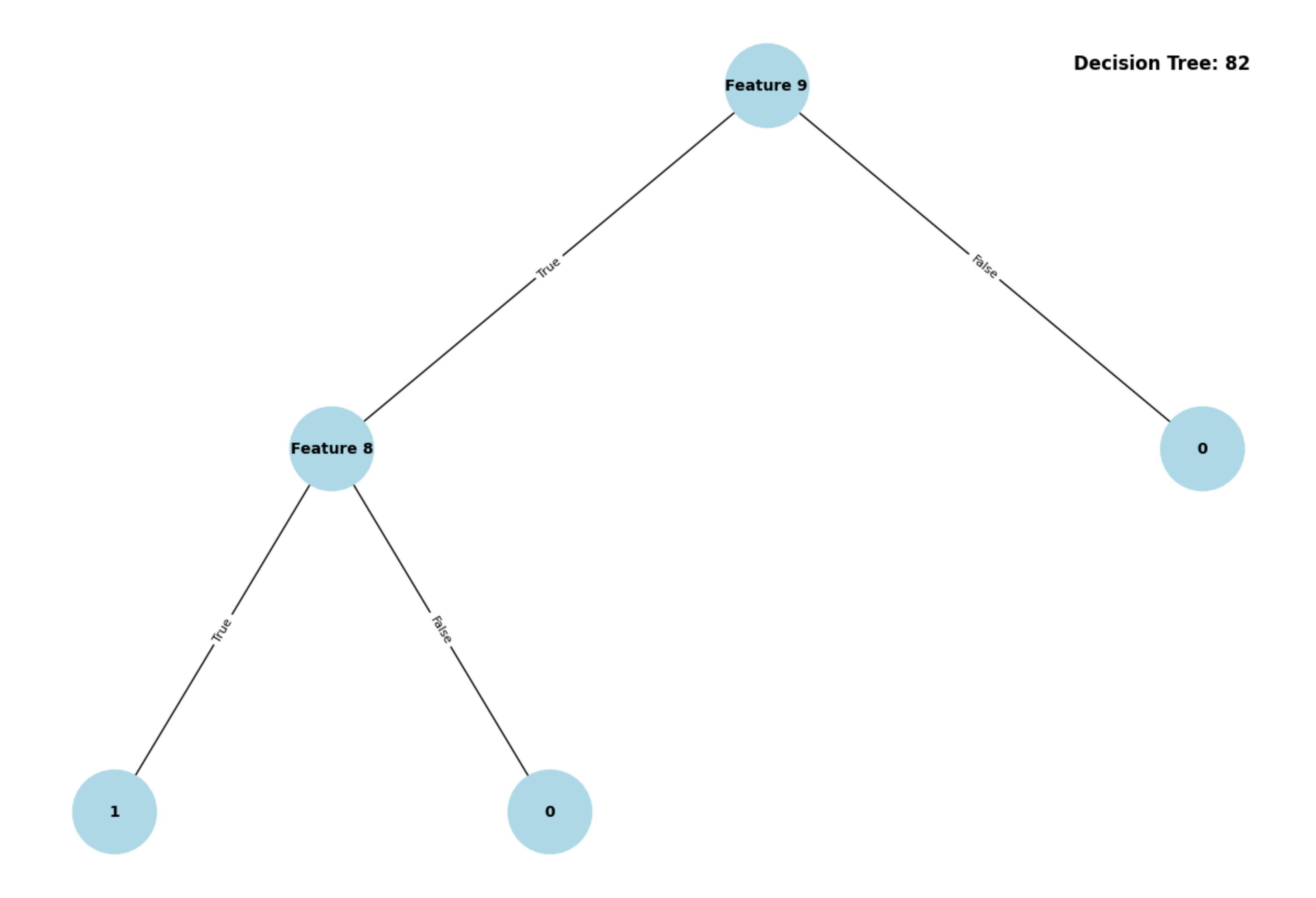


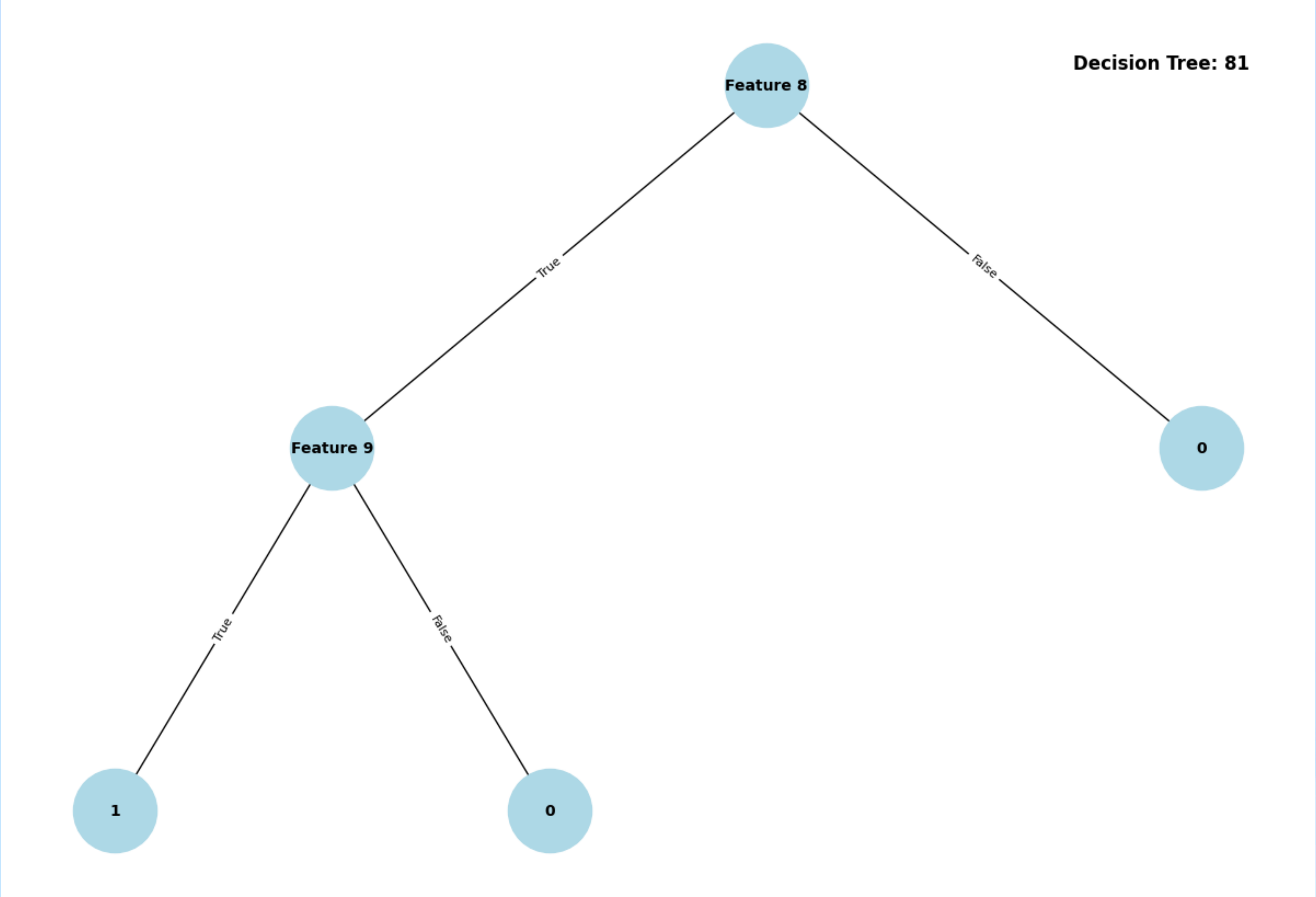


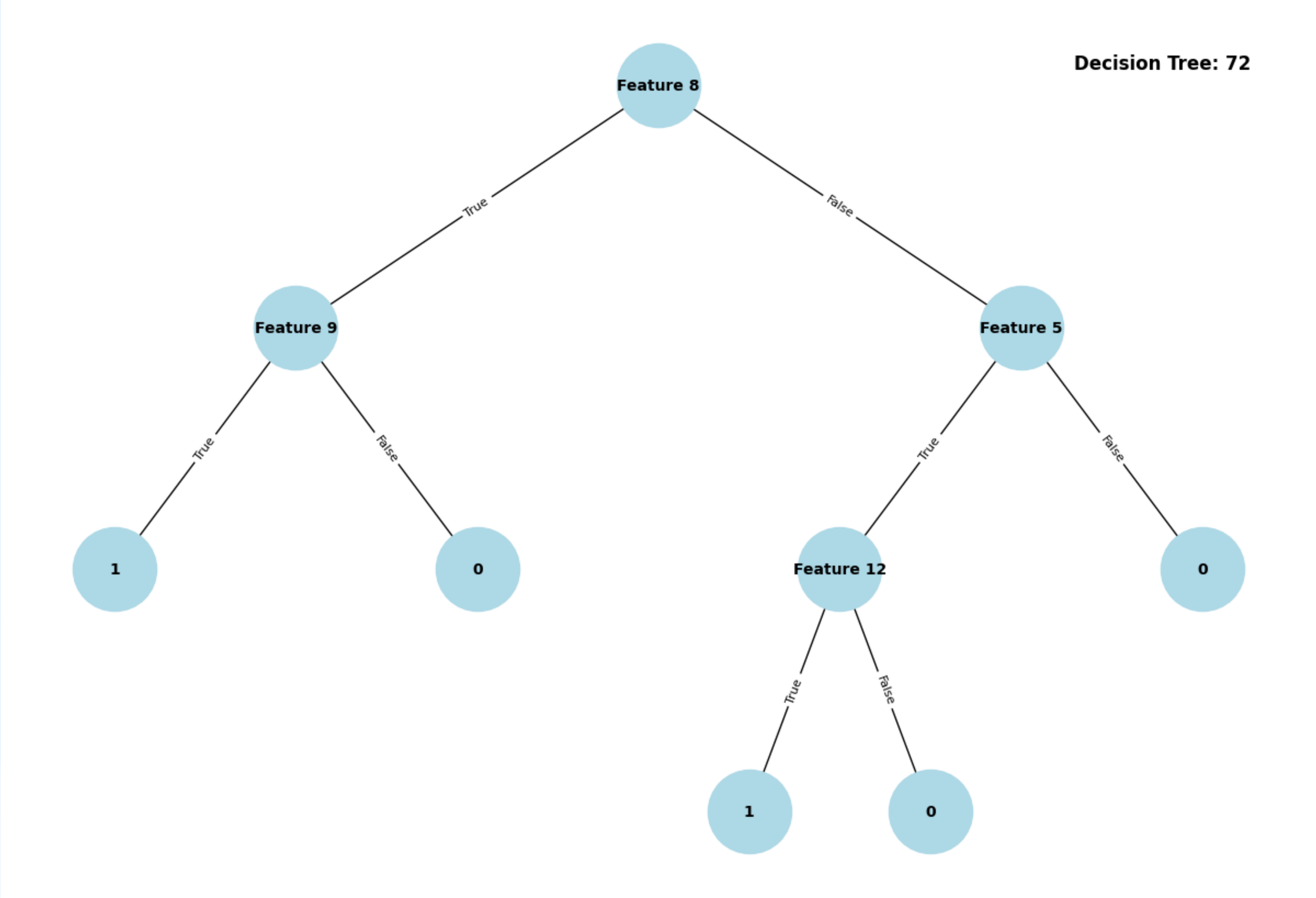


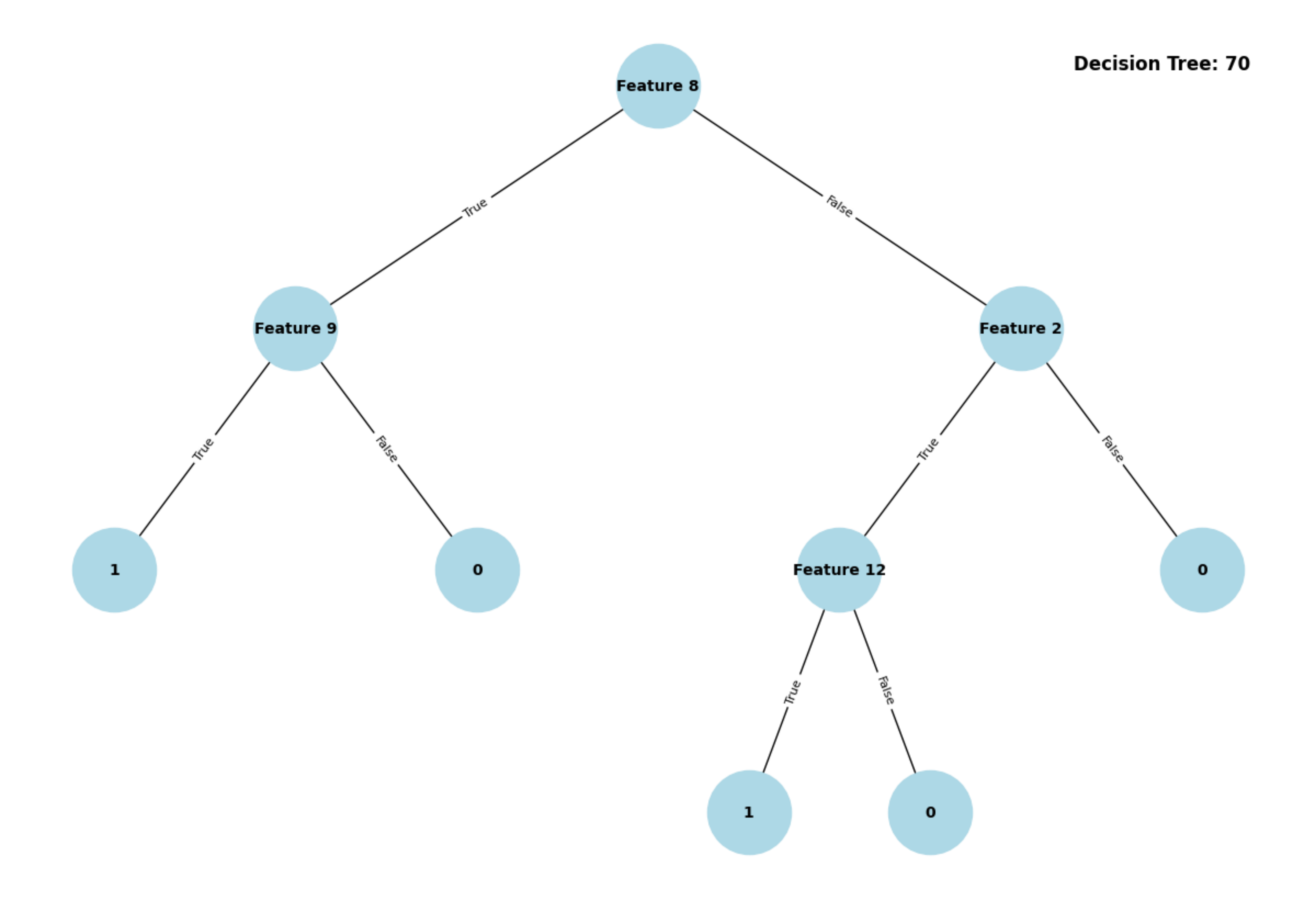


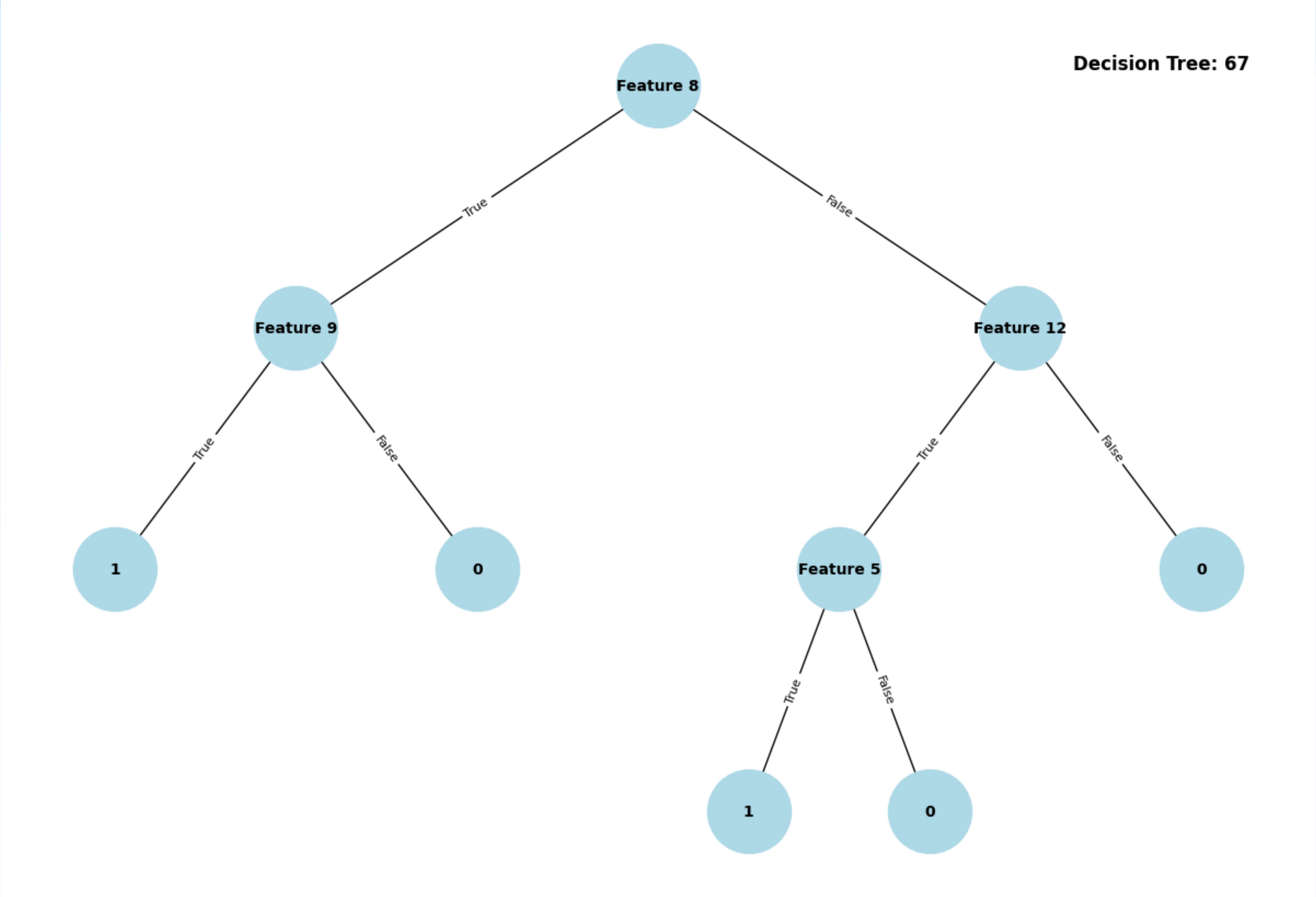
Group 2

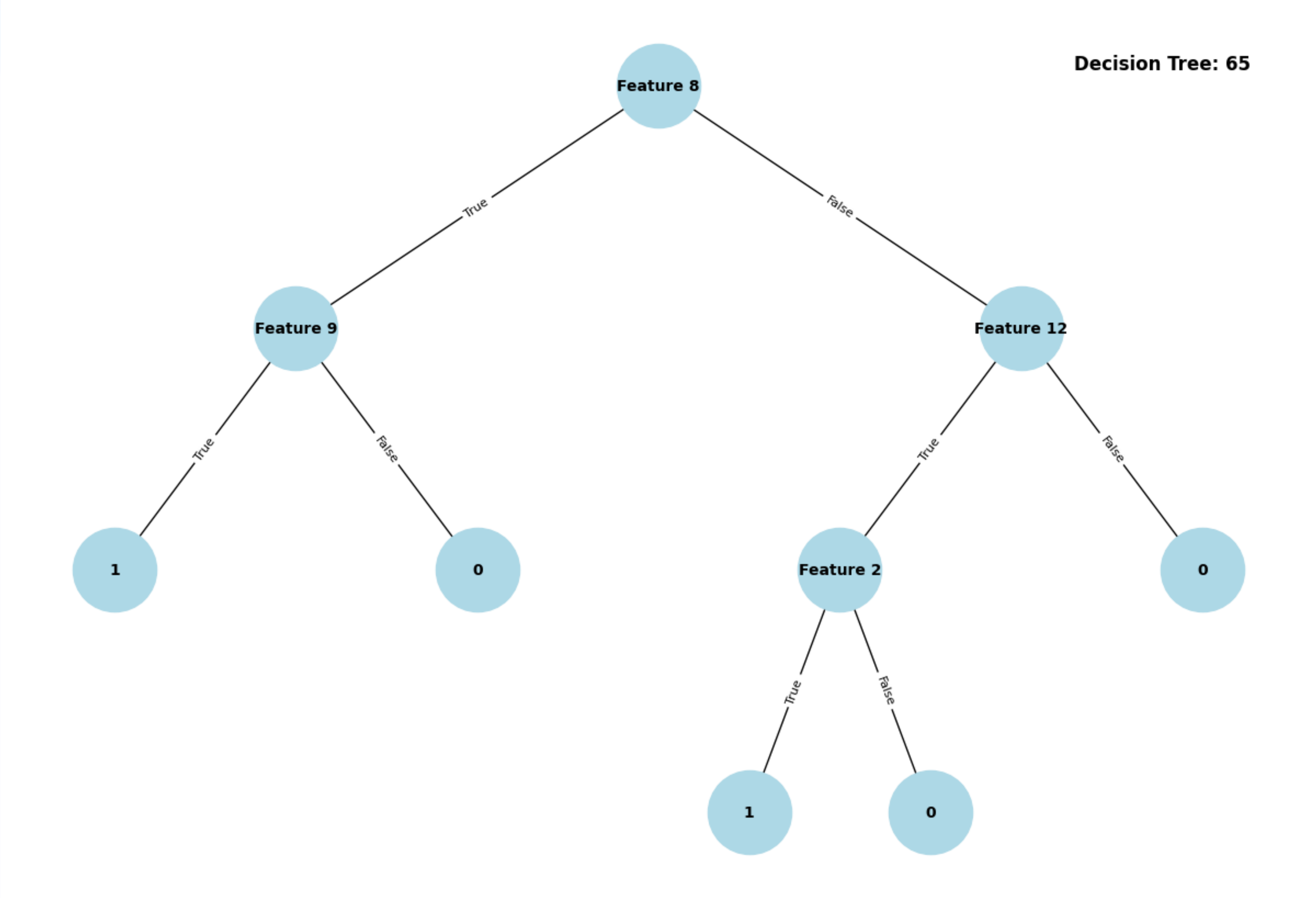

Decision Tree: 82
Feature 9
True
False
Feature 8
0
True
False
1
0


Decision Tree: 81
Feature 8
True
False
Feature 9
0
True
False
1
0

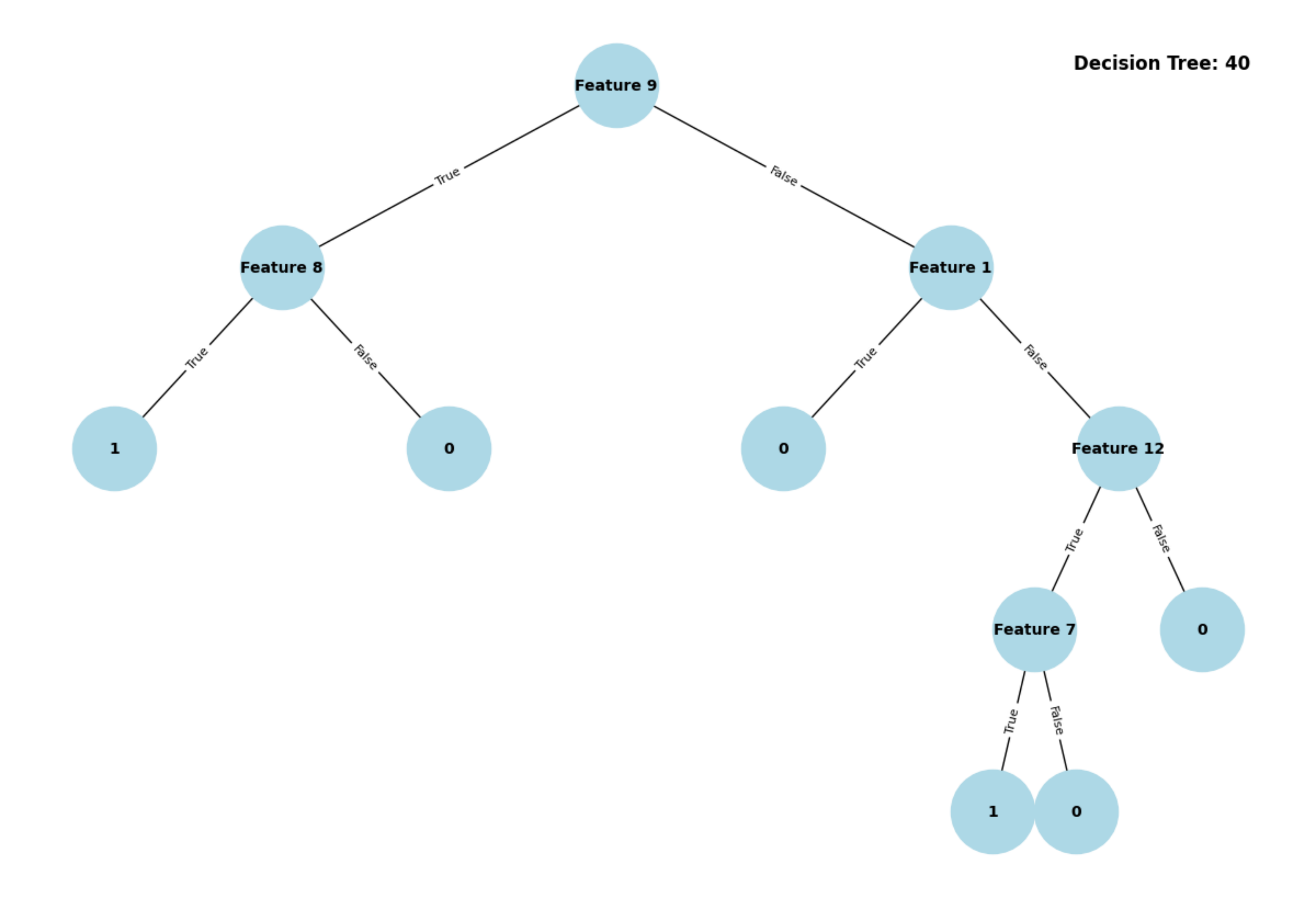

Decision Tree: 72
Feature 8
True
False
Feature 9
Feature 5
True
False
True
False
1
0
Feature 12
0
True
False
1
0

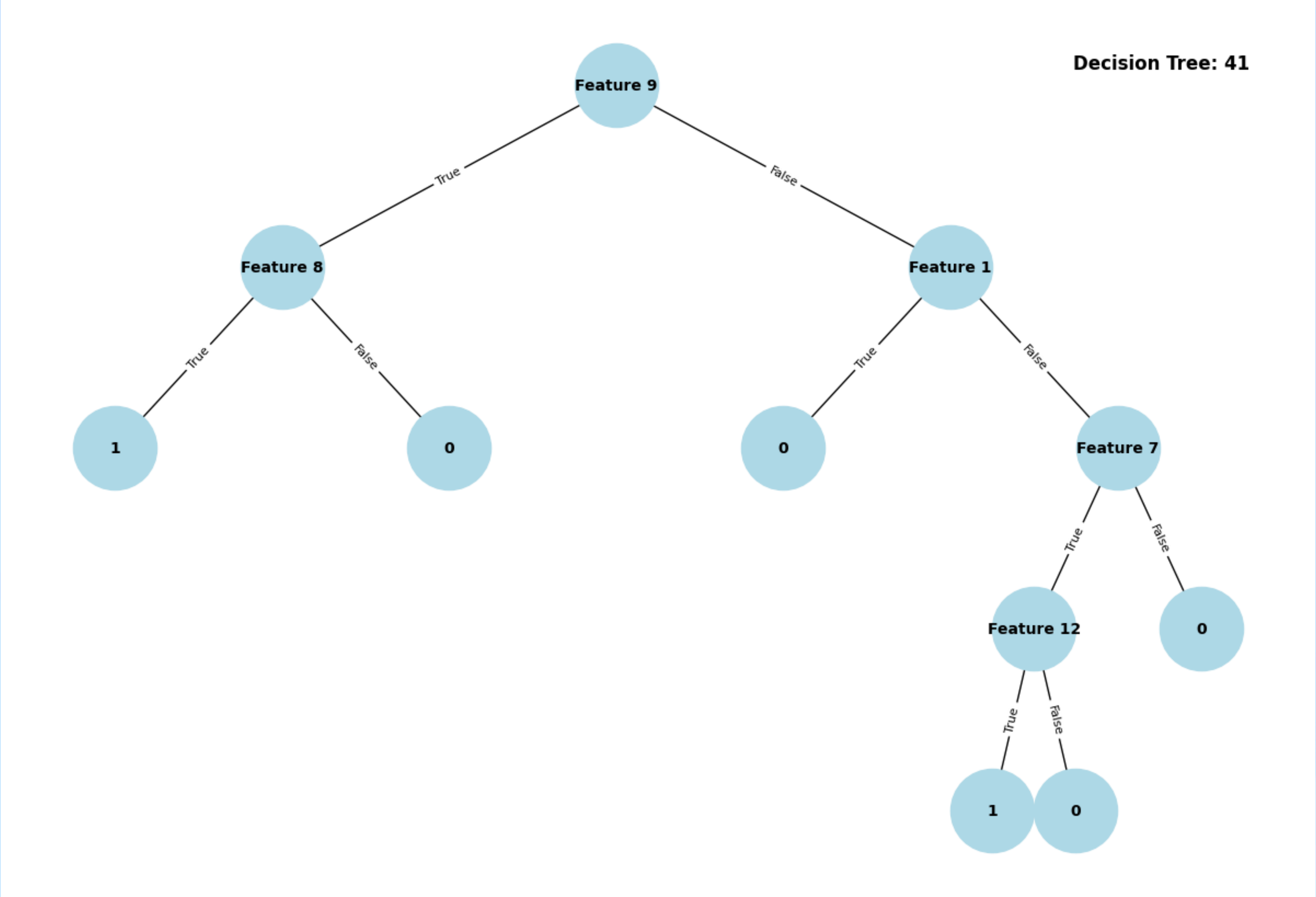

Decision Tree: 70
Feature 8
True
False
Feature 9
Feature 2
True
False
True
False
1
0
Feature 12
0
True
False
1
0

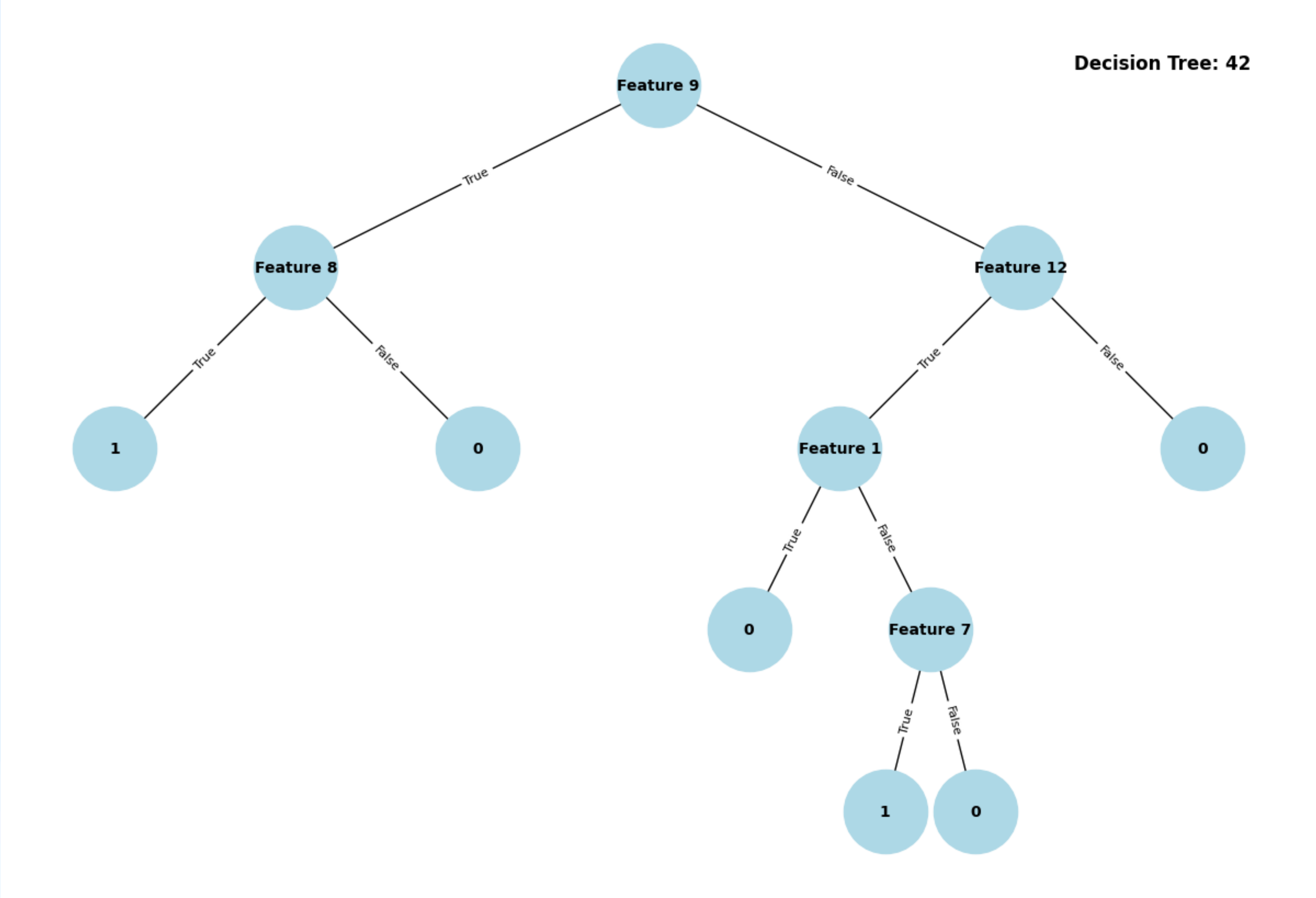

Decision Tree: 67
Feature 8
True
False
Feature 9
Feature 12
True
False
True
False
1
0
Feature 5
0
True
False
1
0

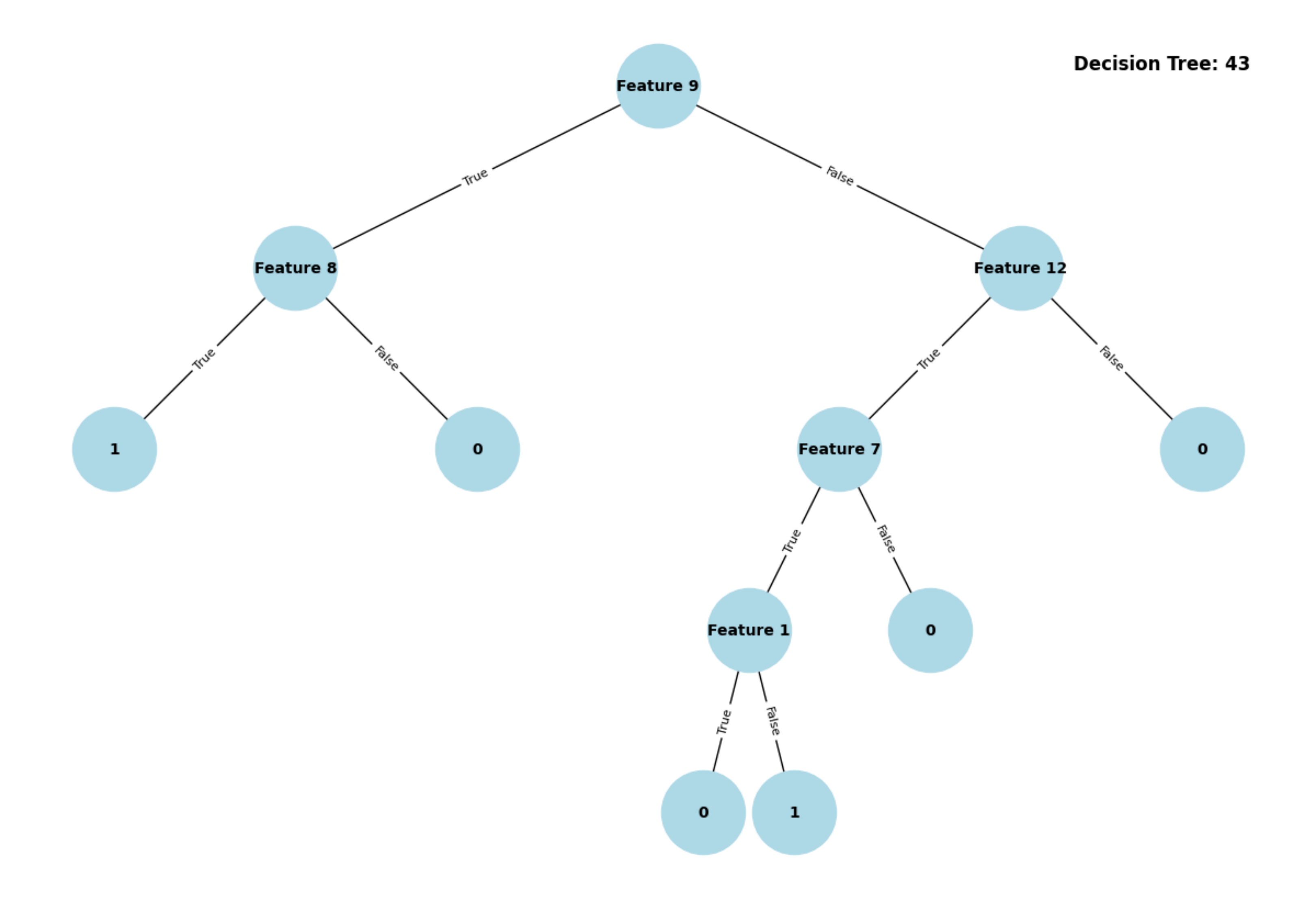

Decision Tree: 65
Feature 8
True
False
Feature 9
Feature 12
True
False
True
False
1
0
Feature 2
0
True
False
1
0

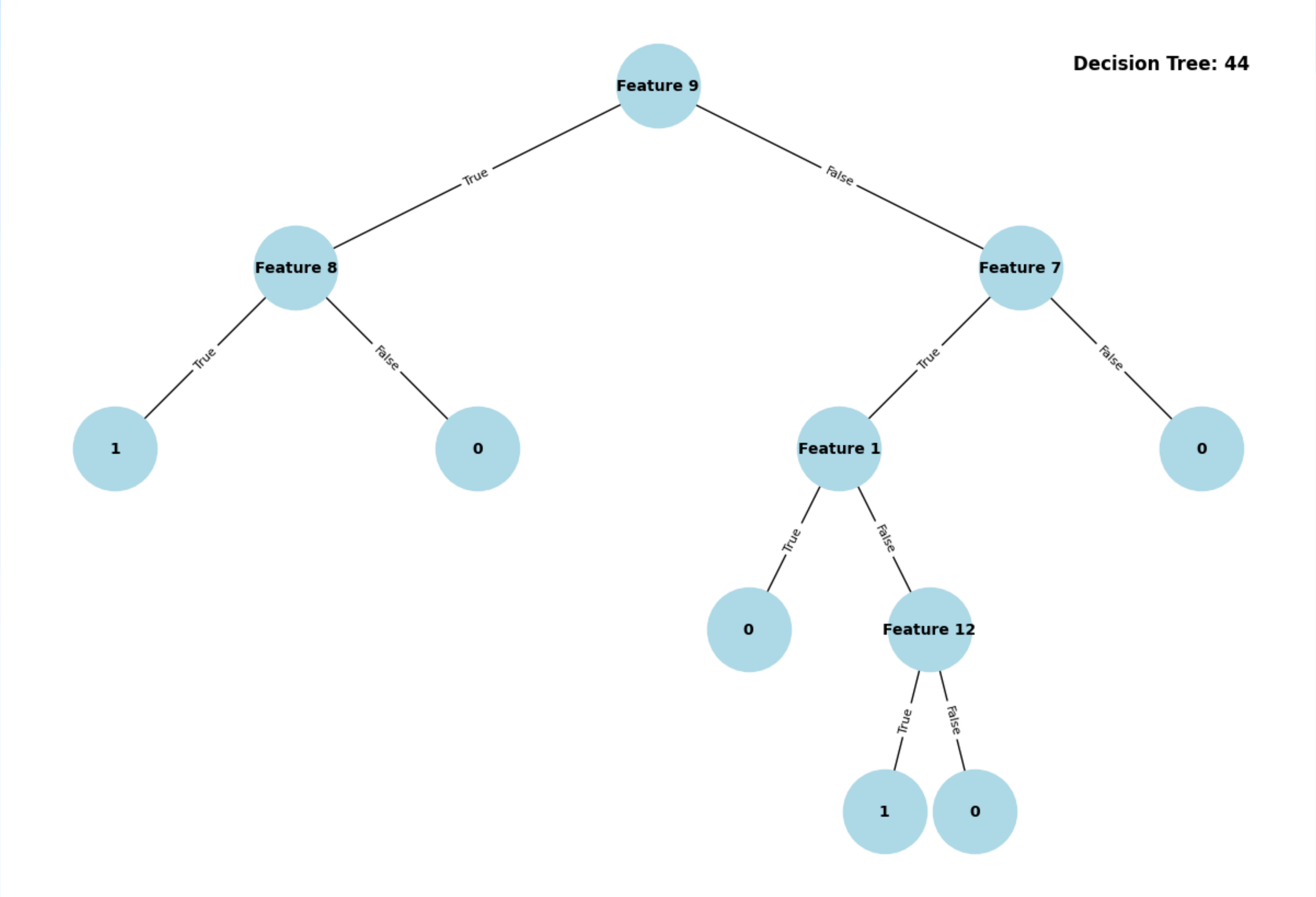
Group 3

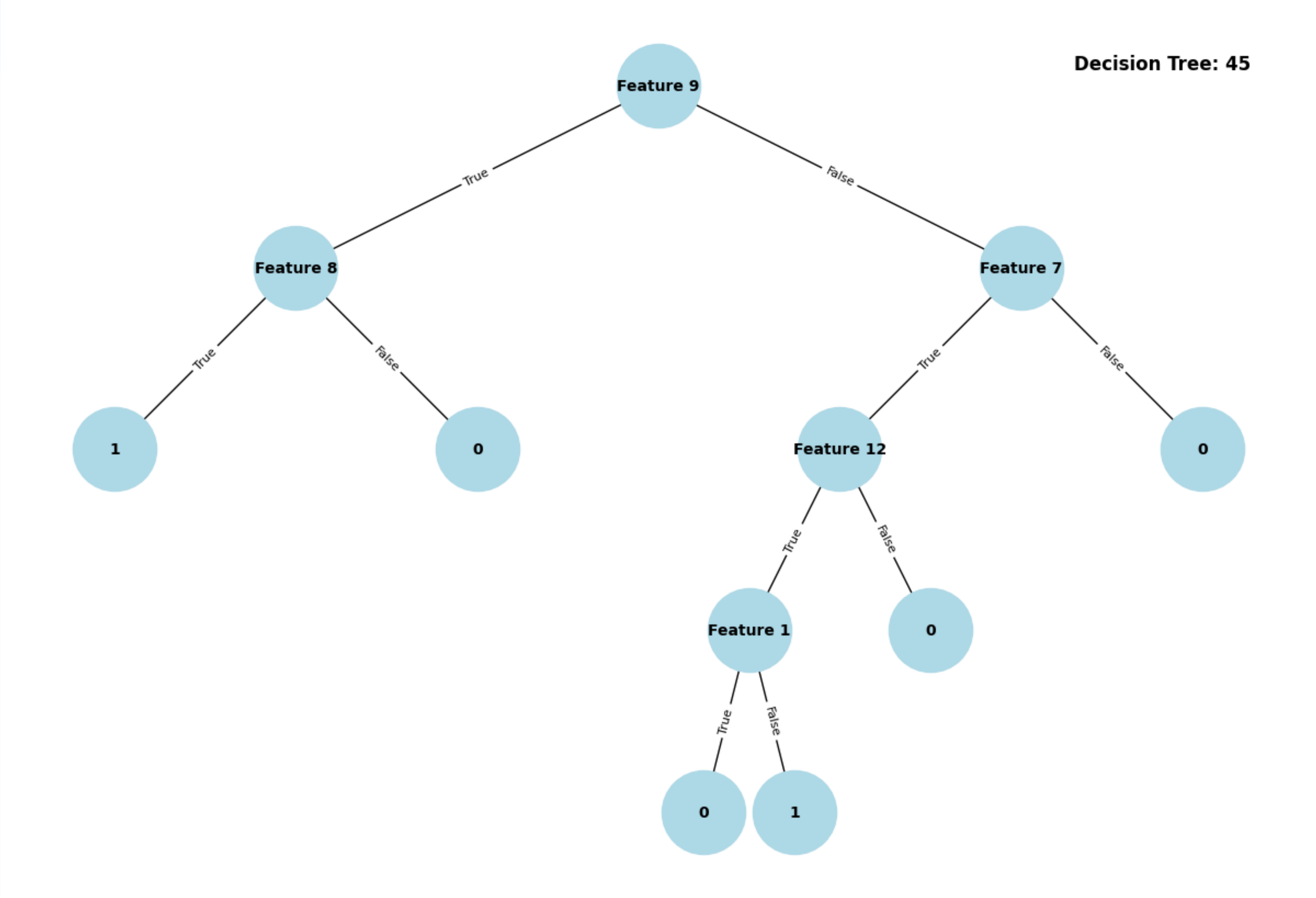

Decision Tree: 40
Feature 9
Feature 8
Feature 1
Feature 12
Feature 7


Decision Tree: 41
Feature 9
Feature 8
Feature 1
Feature 7
Feature 12

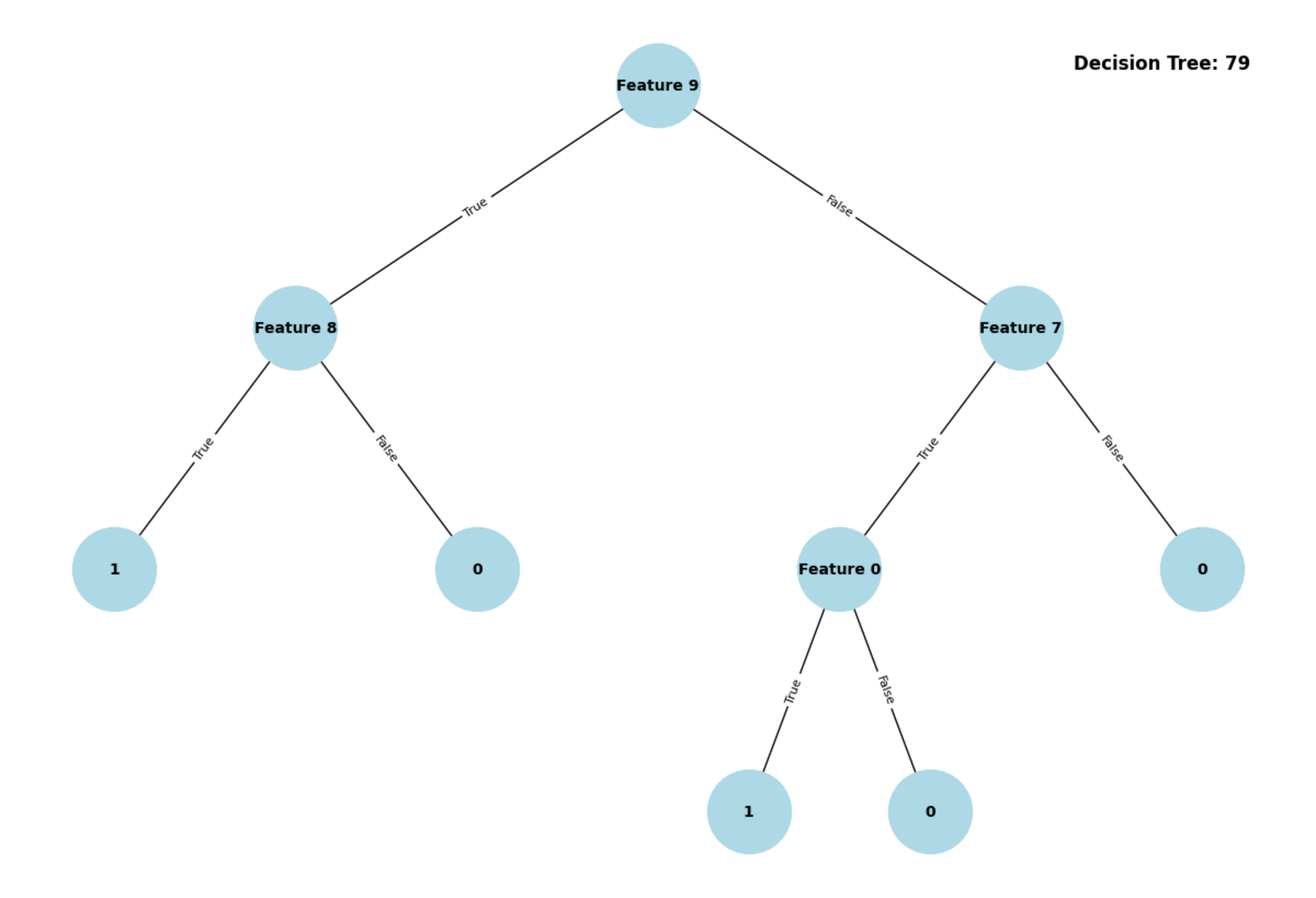

Decision Tree: 42
Feature 9
Feature 8
Feature 12
Feature 1
Feature 7

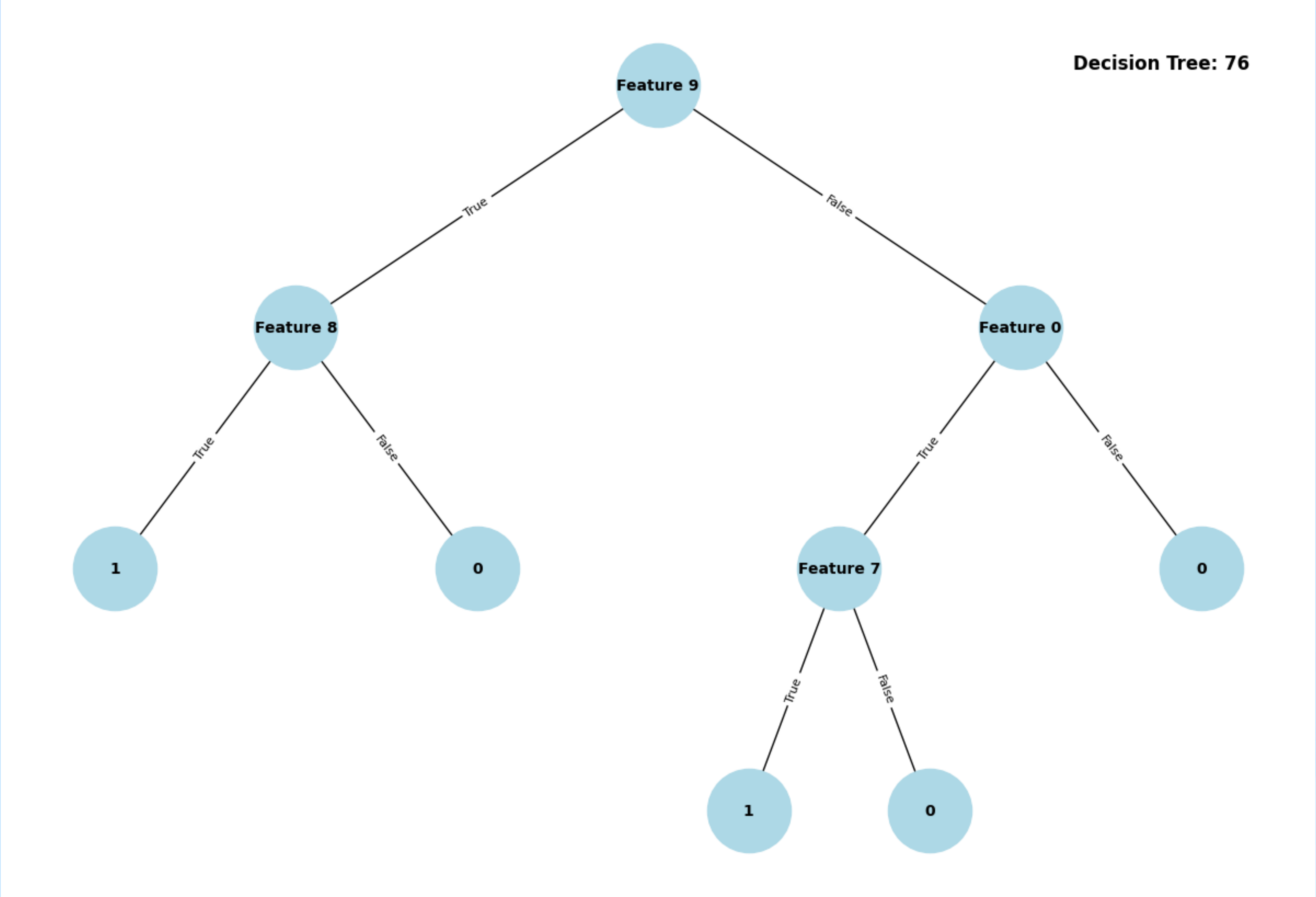

Decision Tree: 43
Feature 9
Feature 8
Feature 12
Feature 7
Feature 1

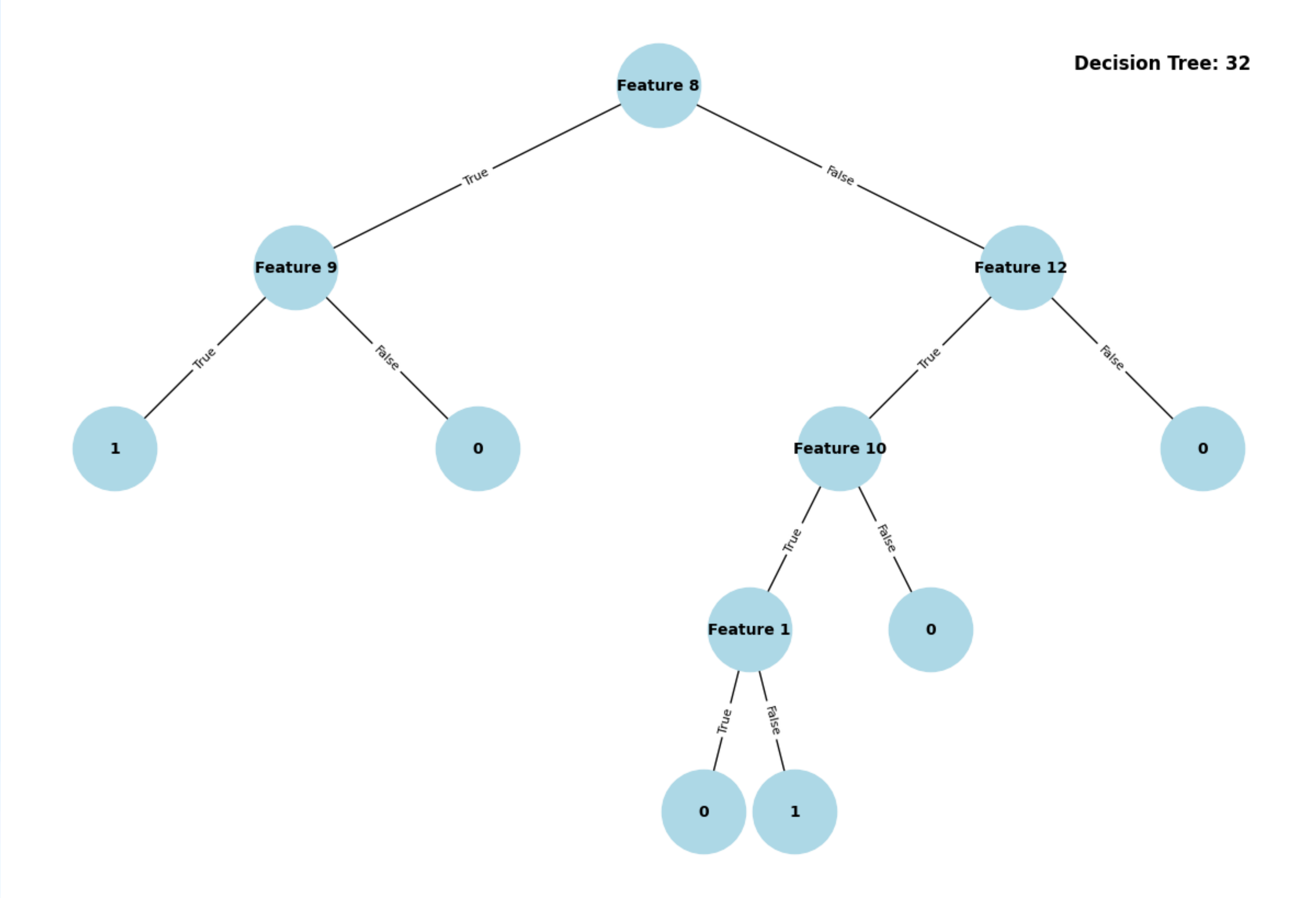

Decision Tree: 44
Feature 9
Feature 8
Feature 7
Feature 1
Feature 12

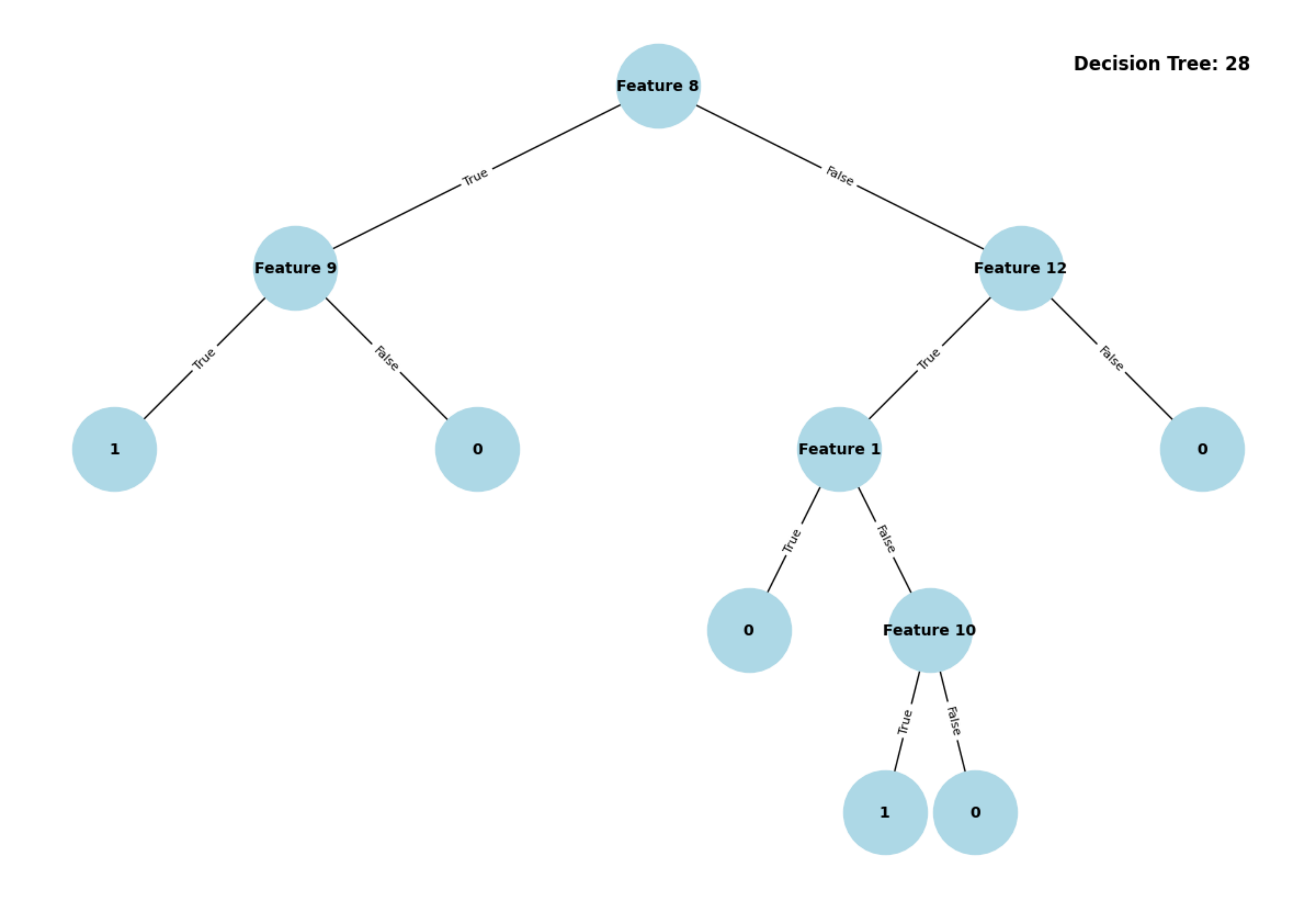

Decision Tree: 45
Feature 9
Feature 8
Feature 7
Feature 12
Feature 1

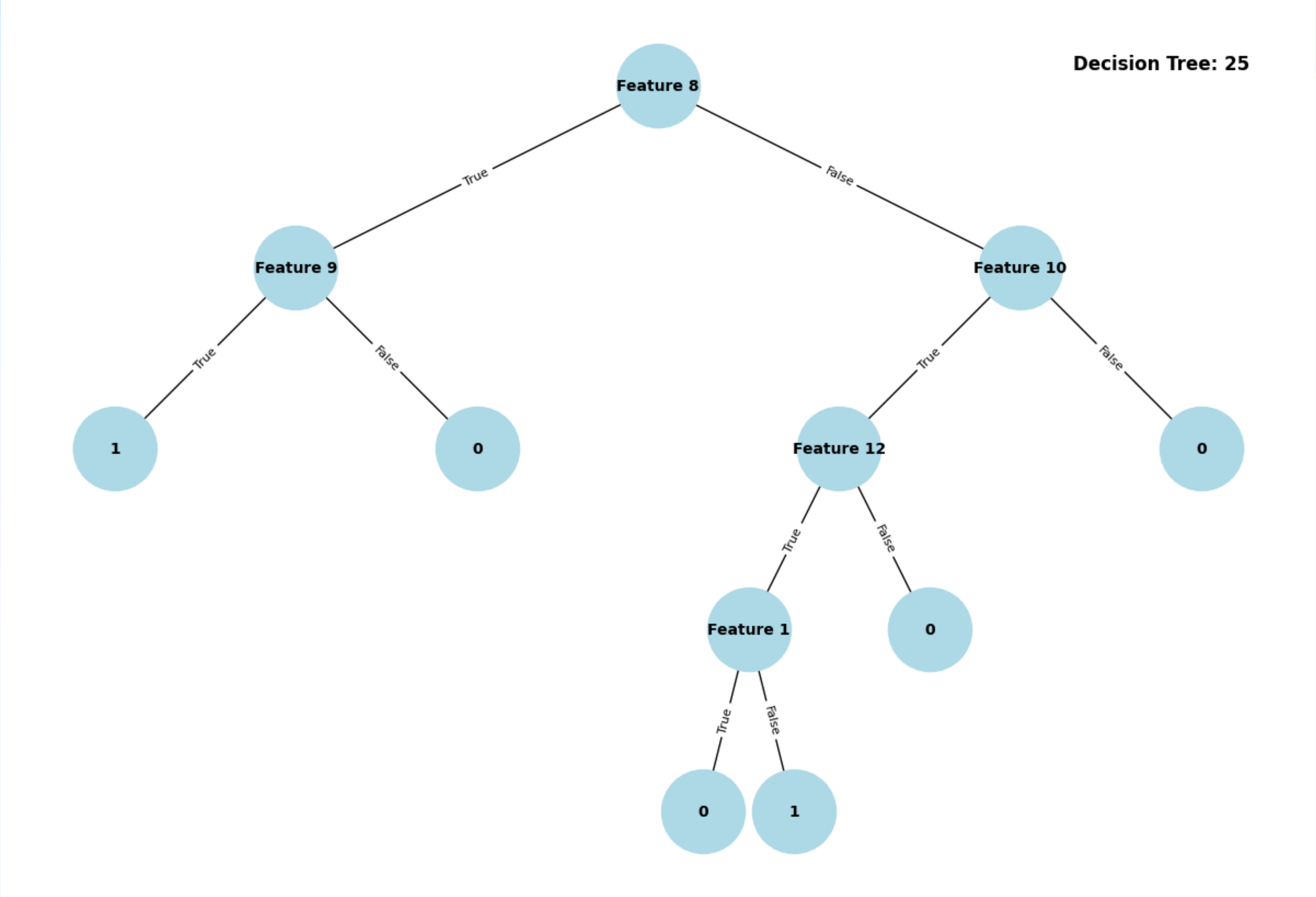
Group 4

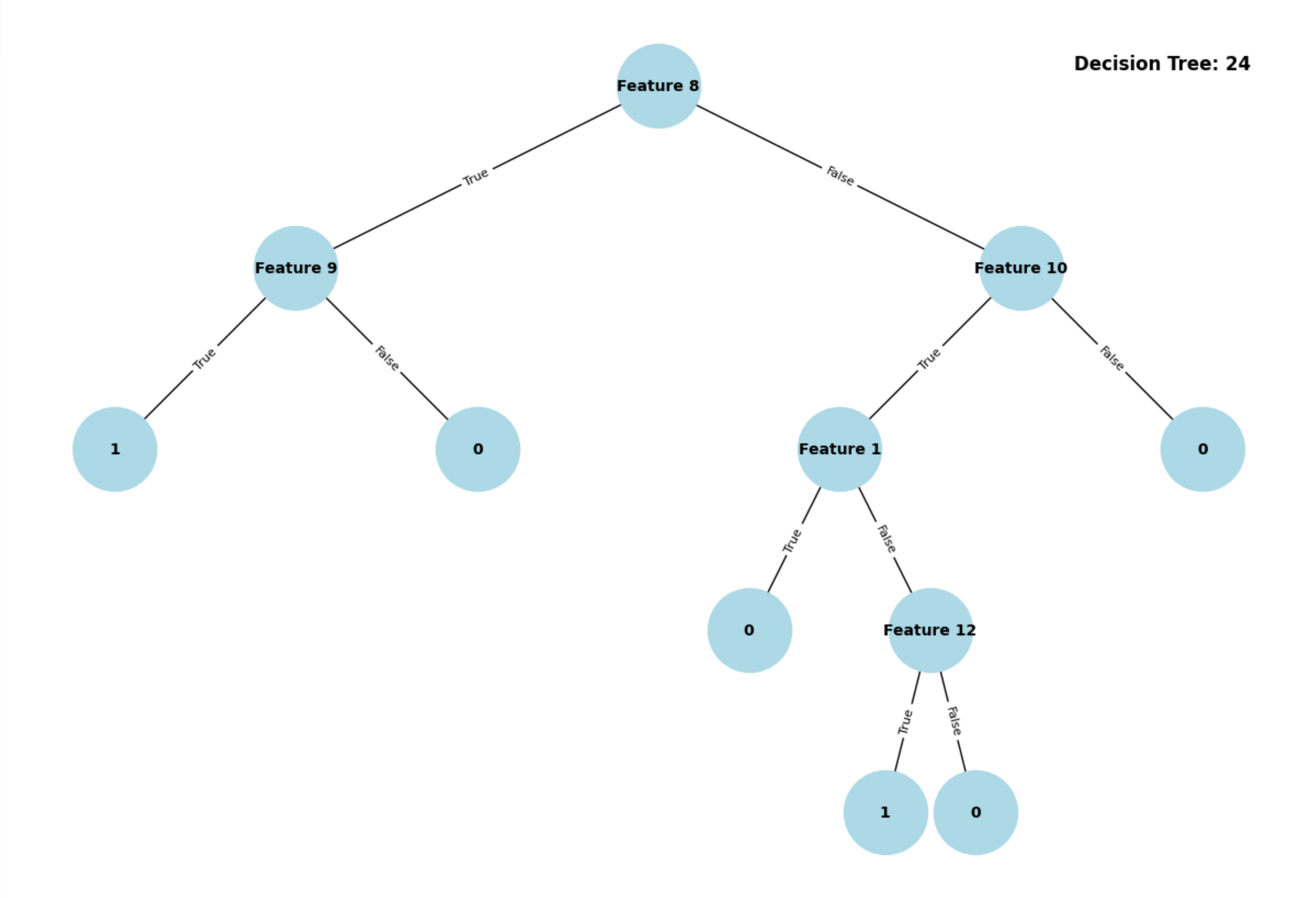












(Excluding two trees for space)

Why does this matter?

- Take away: **All these trees within groups look the same!!!**
- This means that predictions from each tree group will also be exactly the same
  - This is seen when examining the data.
  - (This is difficult to visually present).
- If there are multiple redundant copies of explanations in TreeFarms' decision trees, how does this affect our query-selection criteria?

How does this affect uncertainty?

- In active learning, we measure uncertainty by counting disagreement among ensemble model predictions.

$$\text{VoteEntropy} = \arg\max_x - \sum_{y \in \mathcal{Y}} \frac{\text{vote}_{\mathcal{R}}(y, x)}{|\mathcal{R}|} \log \frac{\text{vote}_{\mathcal{R}}(y, x)}{|\mathcal{R}|}$$

$$\text{such that } \text{vote}_{\mathcal{R}} = \sum_{r \in \mathcal{R}} \mathbb{1}\{r(x) = y\}$$

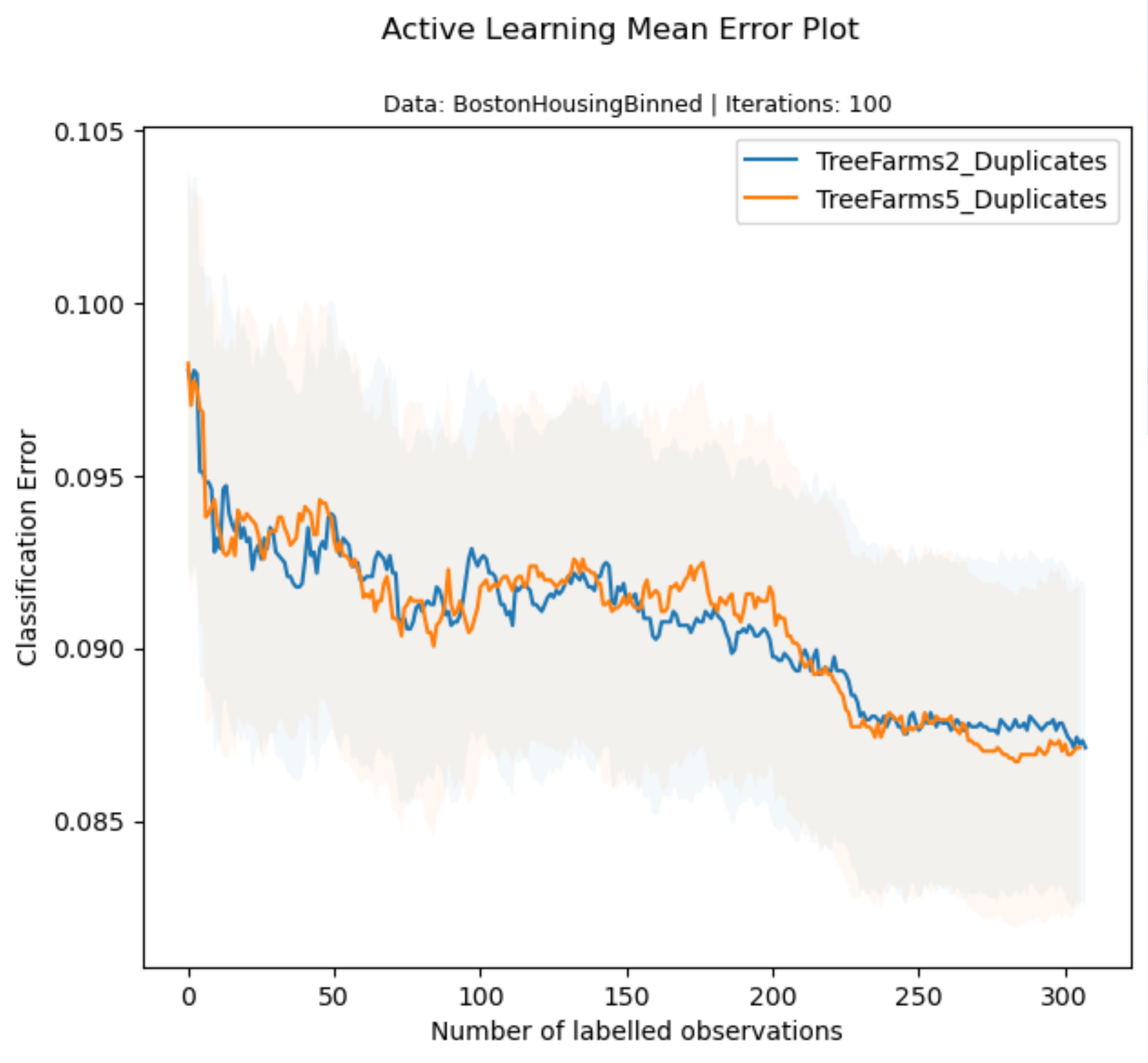
- where $\text{vote}_{\mathcal{R}}$ is the number of "votes" that label y receives for x amongst the trees in Rashomon set $\mathcal{R}$.

- Duplicate trees in the ensemble method can skew vote entropy by **artificially inflating agreement in the vote**!

- Underestimating uncertainty will lead to suboptimal query selection!

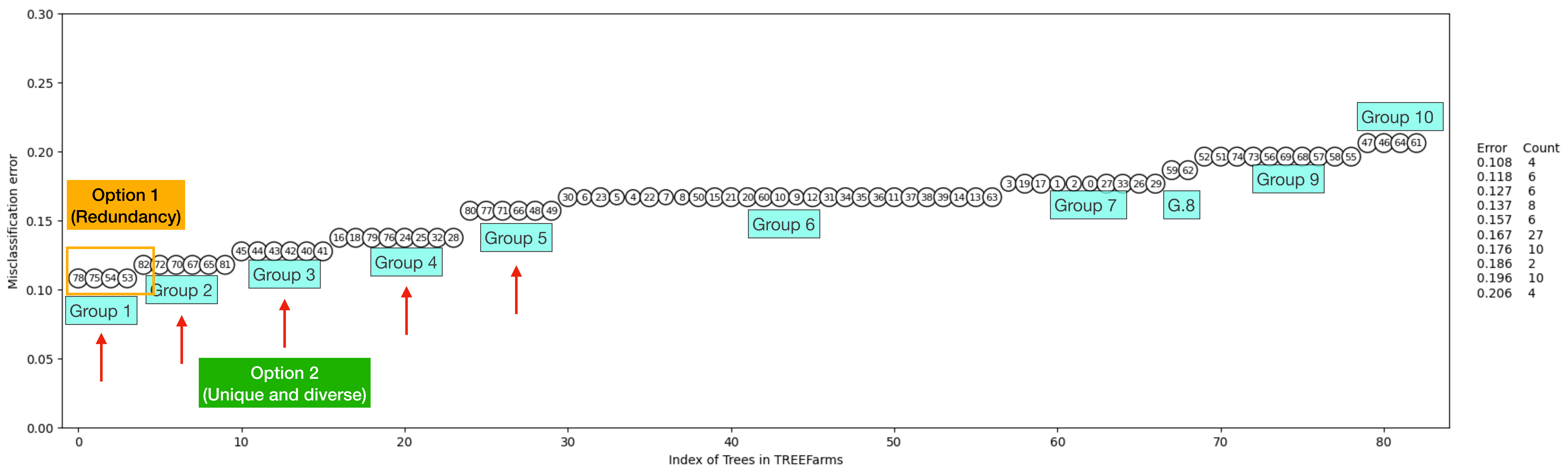
Active learning with Redundant Trees

- In the simulation to the right, active learning is performed with only the best 2 tree!
- Note how similar it is to the active learning method using the top **[5 and 10]** trees!
- Using the best two decision tree vs. the top **[5 and 10]** trees doesn’t make a difference, as they tend to be all the same!
    - We saw this for one iteration in the grouped tree plots before.
    - We now see how it affects query-selection in active learning.
- This deficiency comes from TreeFarms’ multiplicity of explanation and the geometry of trees.



How do we fix this?

- **Possible solution:** We only extract one tree from each group.
- Let's say we measure uncertainty by ensembling the top 4 models.
- Approaches:
  1. Due to redundancy in TreeFarms, we would choose trees [78, 75, 54, 53] (this is what we have been doing - ignoring the redundancy).
  2. Accounting for this redundancy, we would instead choose trees [78, 82, 45, 16, 80] (or any arbitrary tree across the top 5 groups).
- Rather than relying on redundant trees, I attempt to ensure the ensemble reflects multiple explanations of the data across different groups.

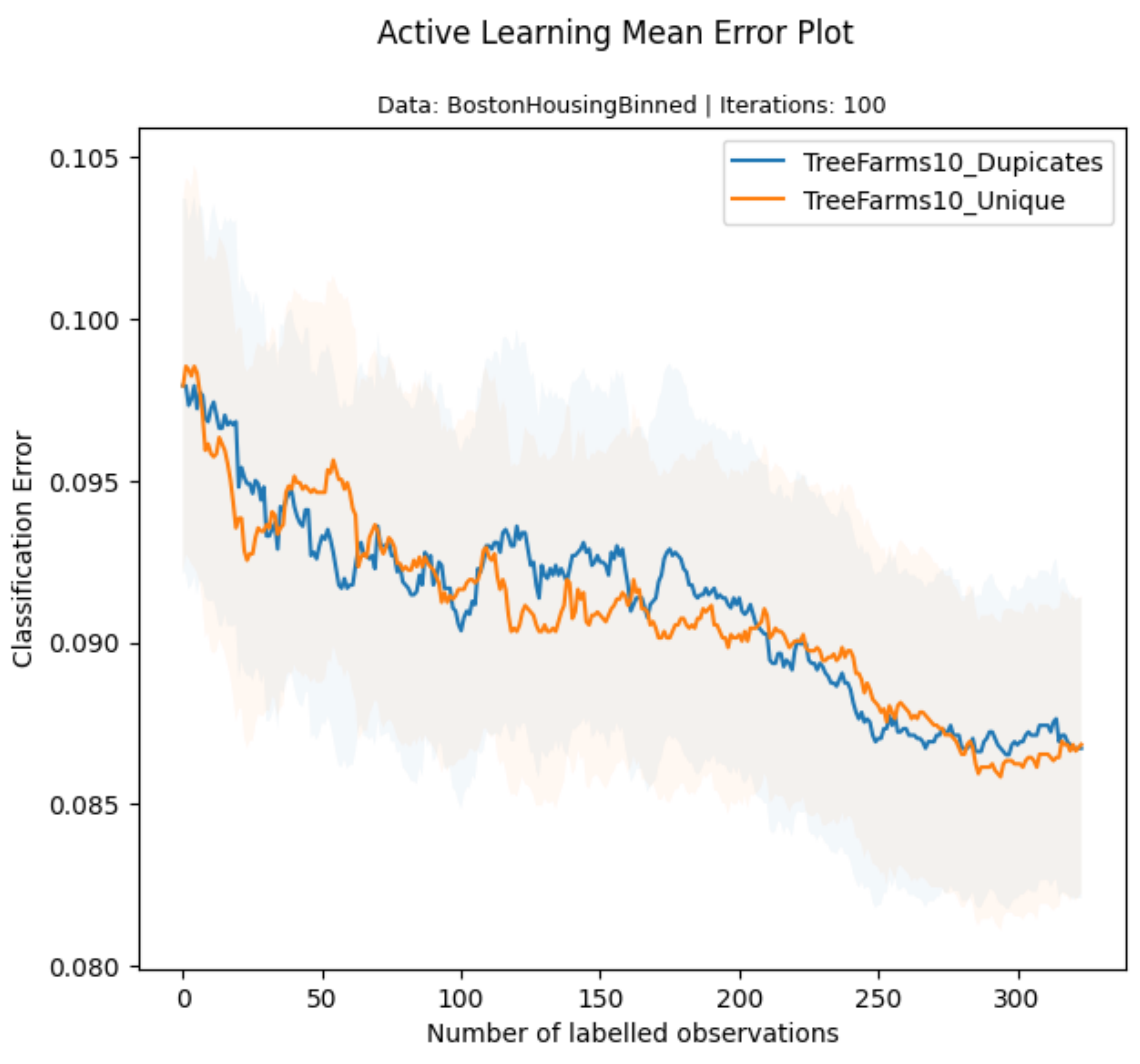


How well does selecting unique decision trees from TreeFarms work?

- The drawing of unique trees **does not work that well**!

- The active learning methods are still very similar!
  - Can also be seen formally with a Wilcoxon Ranked Sign Test.

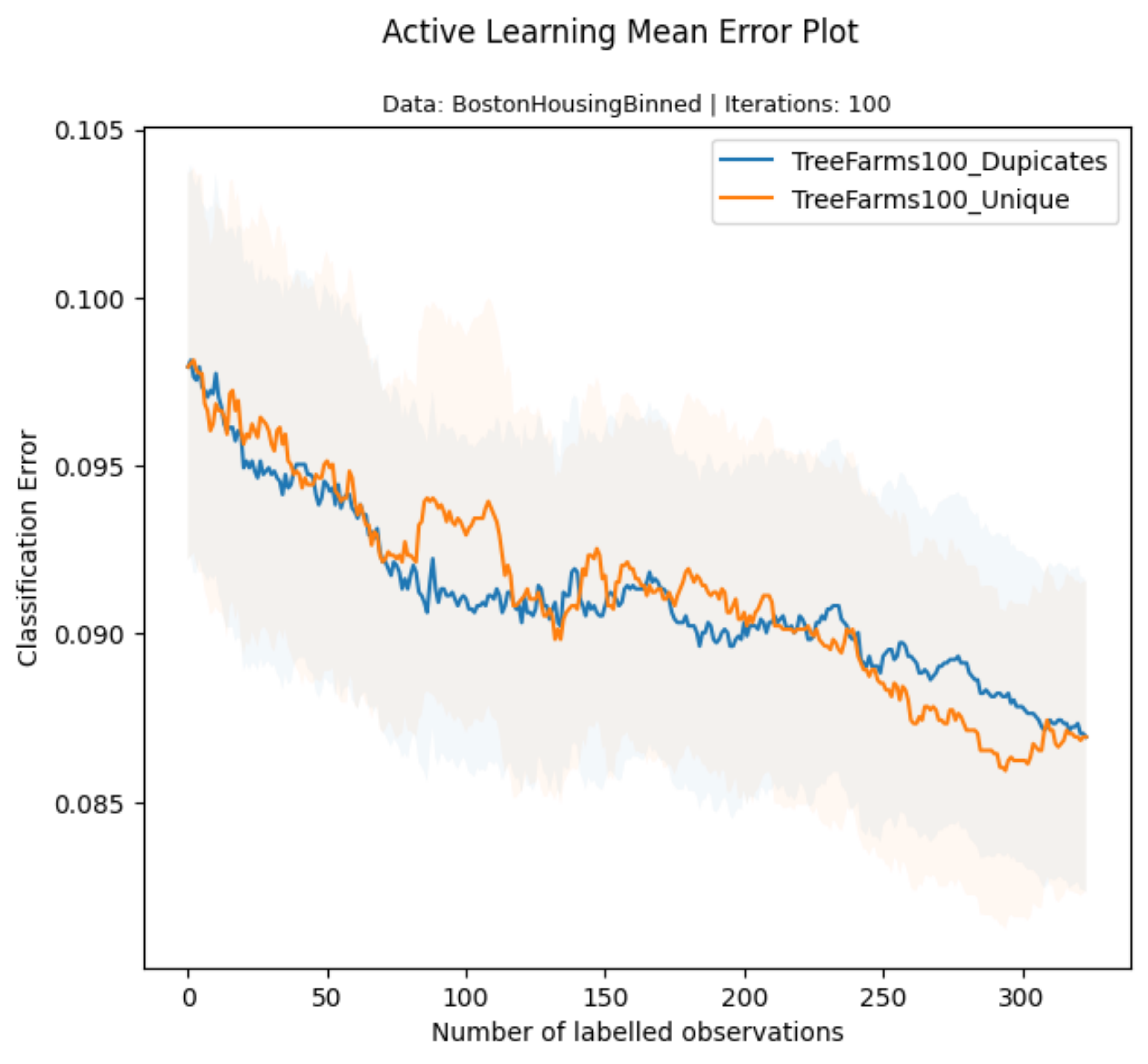
- This image gives the classification error between random forests,TreeFarms with the top 10 models **(not accounting for redundancy)**, and TreeFarms with the top 10 **unique** models.



(A relative plot is provided on Slide 19)

Unfortunately (again)

- The problem does not go away if we reduce our ensemble method to only the top 10 trees :(
- This can be seen as both a pro and a con
  - Con: doesn't work sad
  - Pro: Sets up the story nicely for RPS!
- I will run more simulations
  - I can try taking one tree from each and every of the explanation groups (as opposed to just the top k groups)
  - simon likes to run simulations, but it's probably b/c they're fun and he's scared of dealing with bigger problems hehe
- **If you have insight on this, pls lmk!**
  - Could be due to the retraining/memory issue Tyler and I talked about on Tuesday



(A relative plot is provided on Slide 20)

Option 2

- Since this did not work, we can resort to Rashomon Partition Sets
  - Need to work out for classification
  - RPS intrinsically does not duplicate explanations
  - And still comprehensively enumerates the Rashomon Set like TreeFarms
- This outline of using Rashomon Sets to measure uncertainty in active learning with
  1. Random Forests
  2. TreeFarms
  3. Rashomon Partition Sets

  would make a good story to show the benefits of RPS.

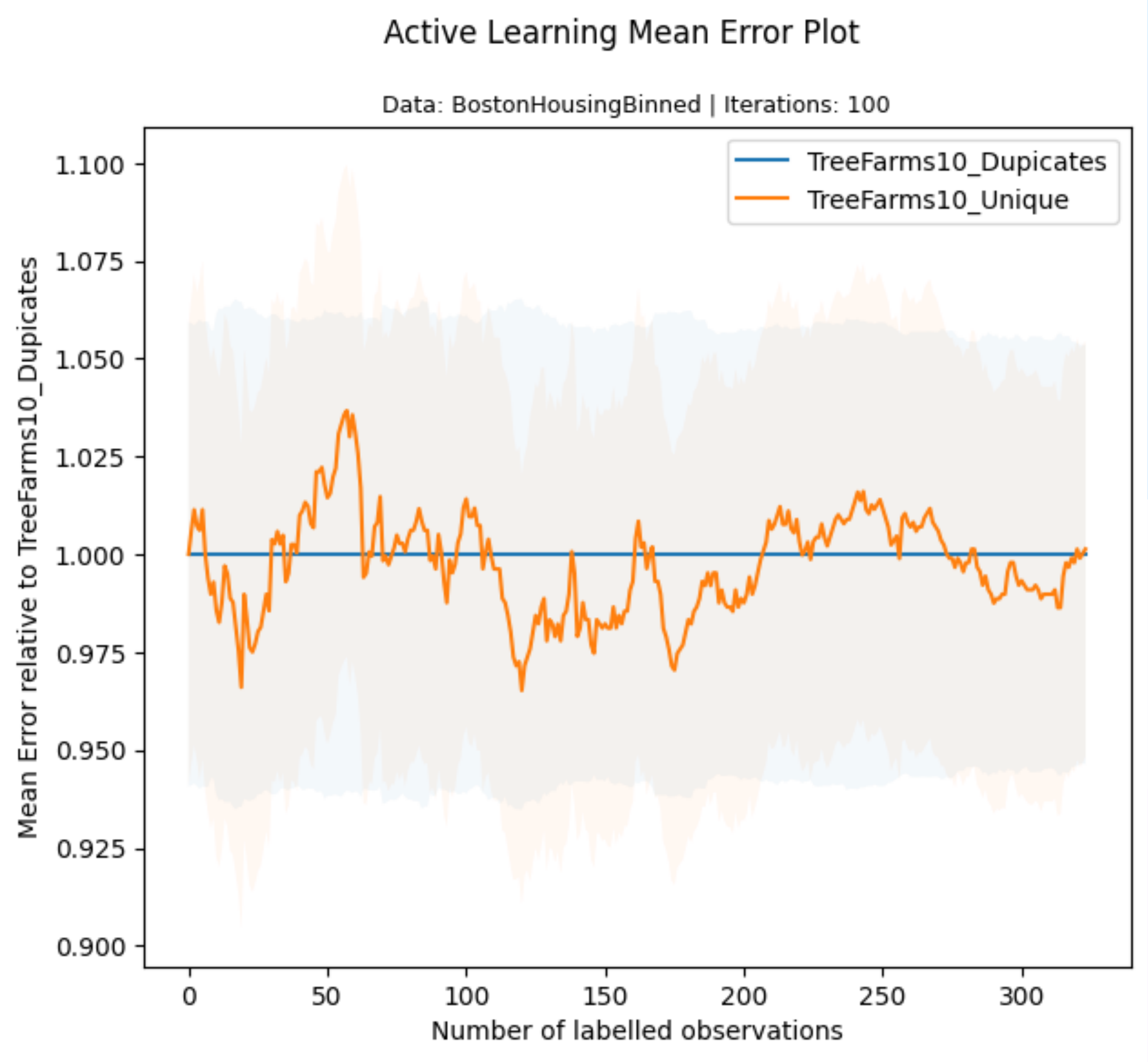
Appendix

TreeFarms Unique vs. Redundant/Duplicate Plot

- The following image is the active learning plot comparing the the unique vs. redundant strategies from Slide 15

- The blue line represents the baseline model, TreeFarms without accounting for the redundancy in decision trees.

- The blue line represents the unique model, TreeFarms accounting for the redundancy by only selecting one decision tree from the best 10 unique explanation groups.



TreeFarms Unique vs. Redundant/Duplicate Plot

- The following image is the active learning plot comparing the the unique vs. redundant strategies from Slide 16

- The blue line represents the baseline model, TreeFarms without accounting for the redundancy in decision trees.

- The blue line represents the unique model, TreeFarms accounting for the redundancy by only selecting one decision tree from the best 100 unique explanation groups.