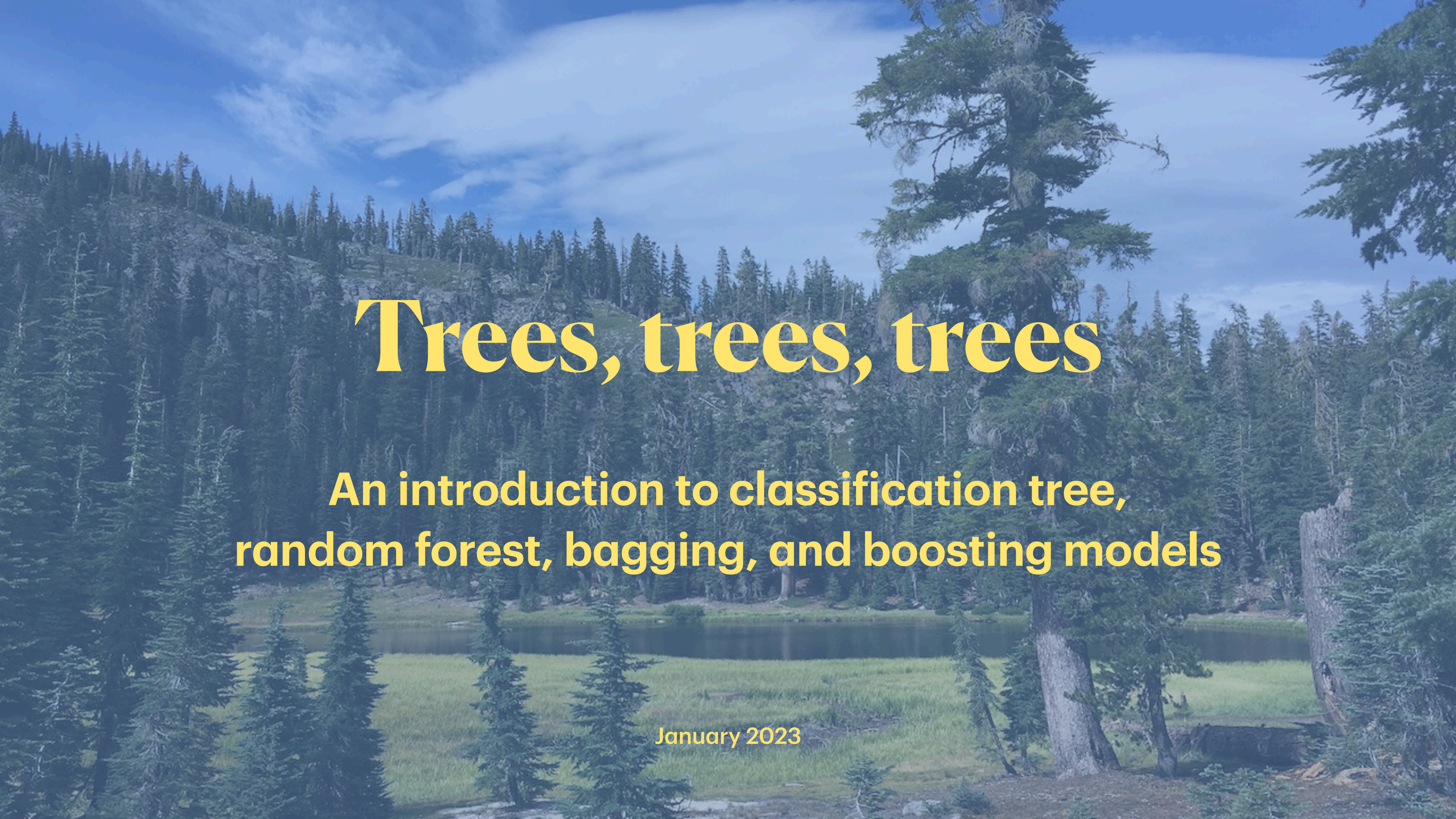


Trees, trees, trees

An introduction to classification tree, random forest, bagging, and boosting models

January 2023

Dataset and problem

Why does someone decide to buy hearing aids?

	Age	Sex	Hearing test	Reported handicap	Stigma	x = 28	Purchased aids ⅕
Case 1	76	M	65	32	3	...	**1**
Case 2	61	M	45	26	4	...	**0**
Case 3	68	F	50	24	4	...	**0**
n = 753	...	...	...	...	...	...	**...**

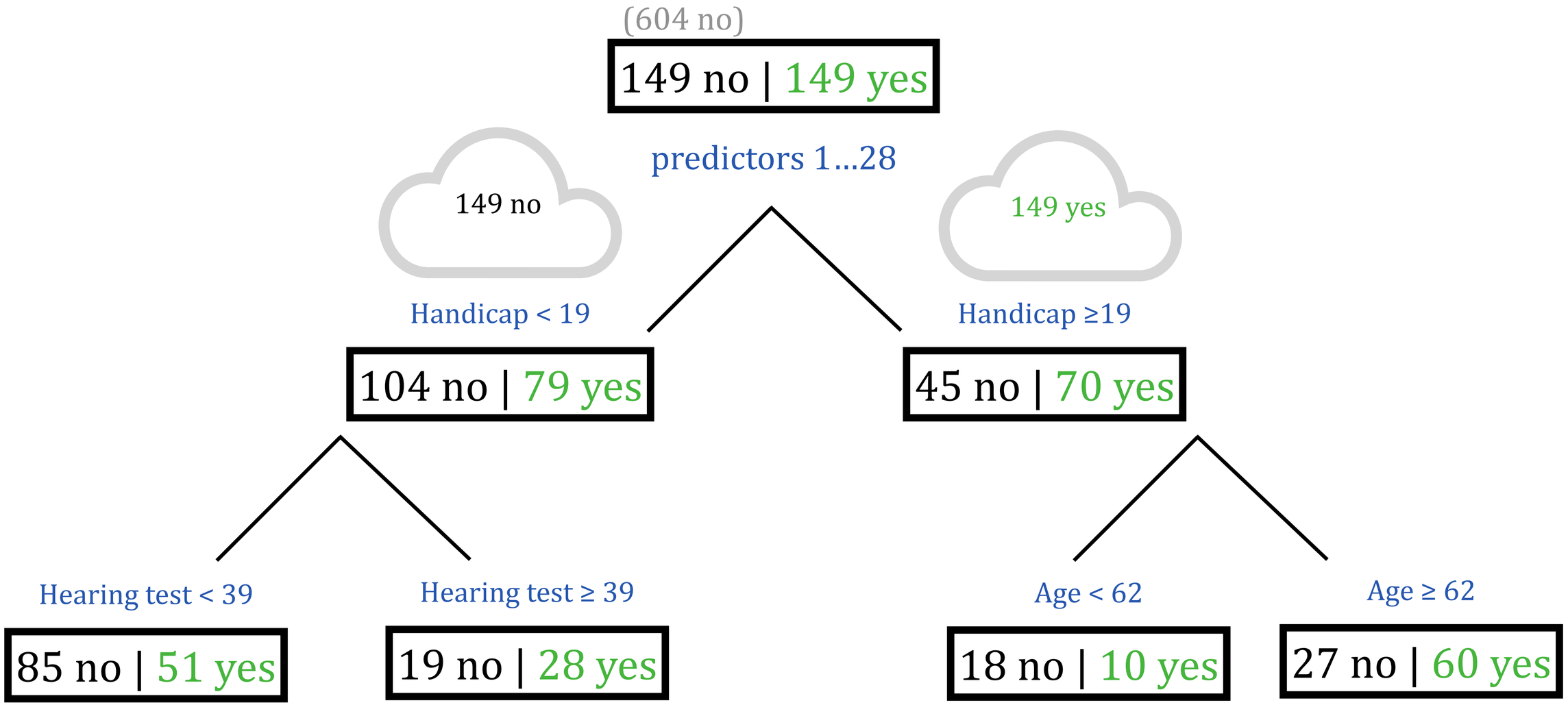
Initial attempt: Logistic regression

> glm (Purchased ~ Age + Sex + Hearing + Handicap… , data = df, family = “binomial”)

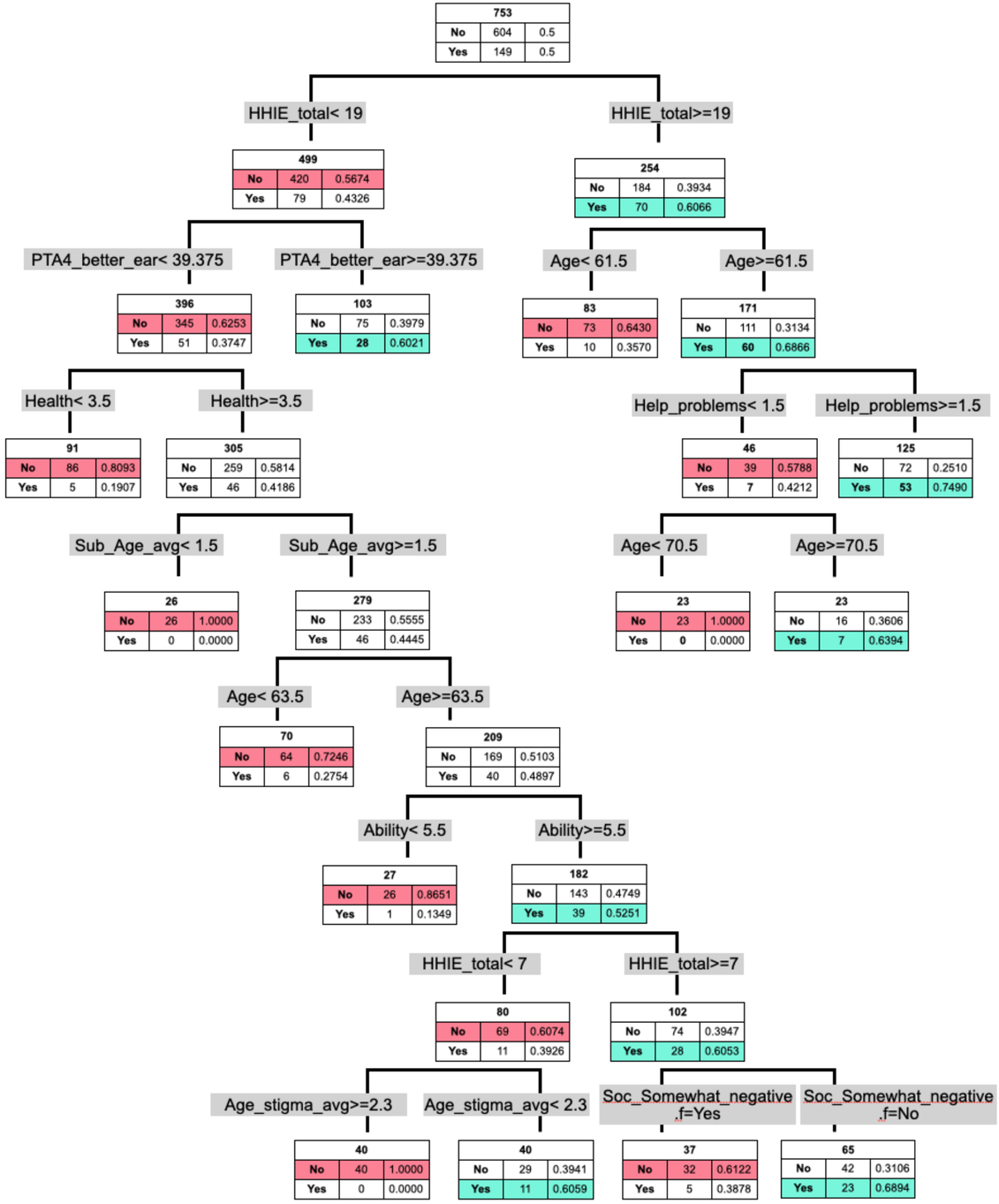
	Odds ratio	**CI lower**	**CI upper**	**p-value**
Age	1.046	1.024	1.069	<0.0001
Handicap	1.047	1.027	1.068	<0.0001
Stigma	0.85	0.71	1.01	0.065
Know someone	2.10	1.12	4.30	0.029

	LR
Accuracy	**63.5**
Sensitivity	**59.7**
Specificity	**64.4**

Classification tree (CART)



Complexity Parameter	Number of splits in tree	Overall accuracy	Sensitivity	Specificity	Area Under Curve
0.2	0	0.50	0.000	1.000	0.5000
0.1	1	0.6507	0.46980	0.69536	0.5826
0.05	3	0.672	0.5906	0.6921	0.6413
0.03	3	0.672	0.5906	0.6921	0.6413
0.025	3	0.672	0.5906	0.6921	0.6413
0.02	3	0.672	0.5906	0.6921	0.6413
0.015	5	0.7025	0.5906	0.7301	0.6604
0.013	12	0.6534	0.8188	0.6126	0.7157
0.012	12	0.6534	0.8188	0.6126	0.7157
0.011	15	0.6454	0.8792	0.5877	0.7335
0.01	22	0.7211	0.8725	0.6838	0.7781
0.005	31	0.745	0.9262	0.7003	0.8133
0	41	0.7822	0.9195	0.7483	0.8339



	LR	**Tree**
Accuracy	63.5	**65.3**
Sensitivity	59.7	**81.2**
Specificity	64.4	**61.3**

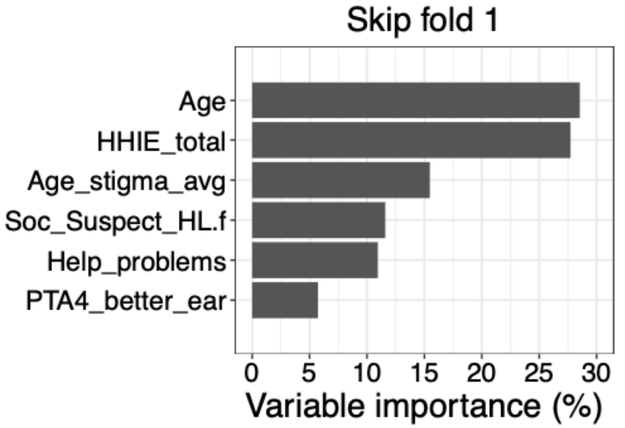
Model stability

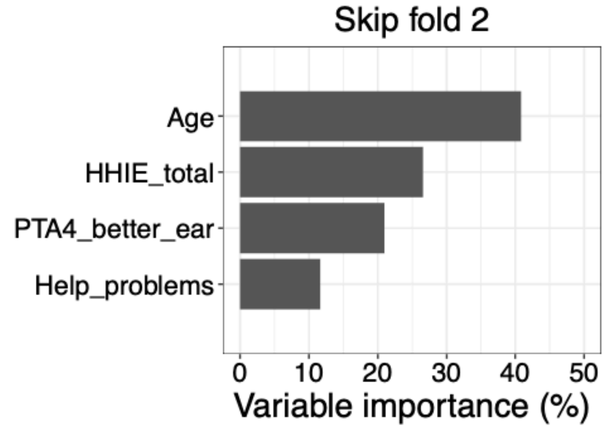
Method

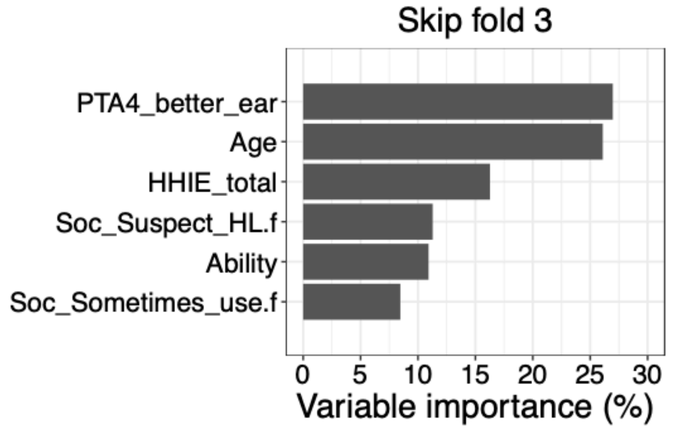
Drop out 1 of 5 folds, keeping the proportion of Yes/No the same

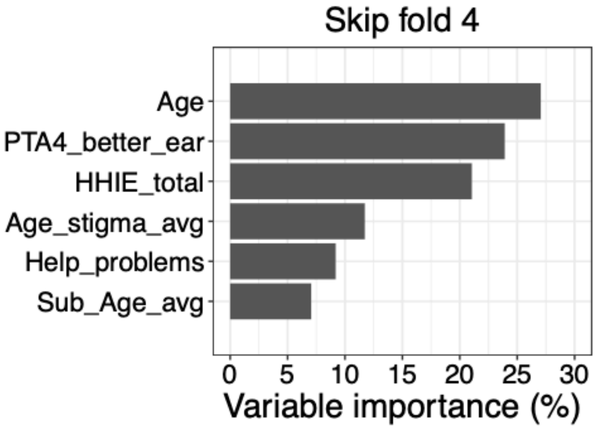
Variable importance

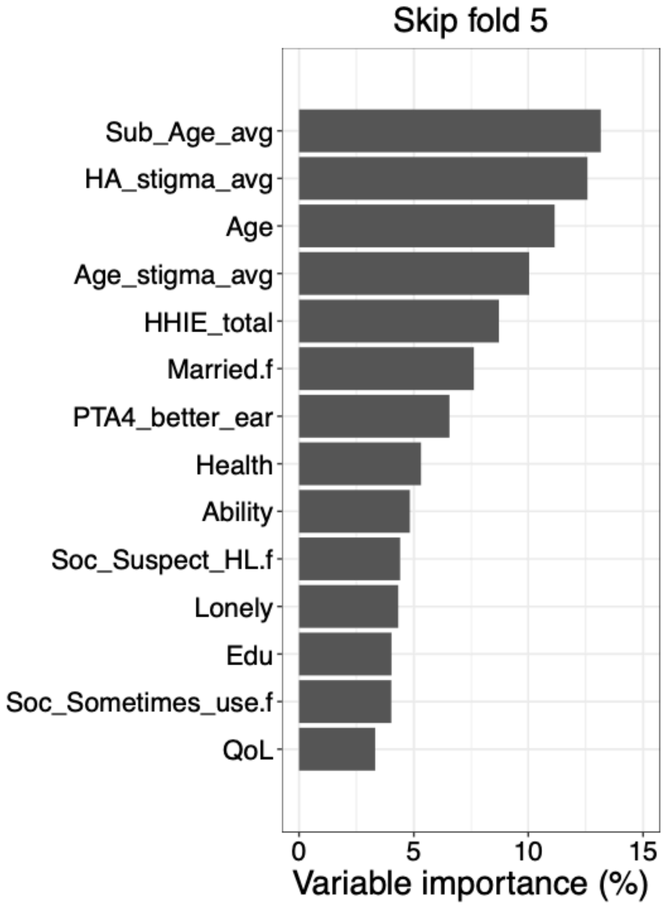
How many cases each predictor switched from a wrong to correct classification







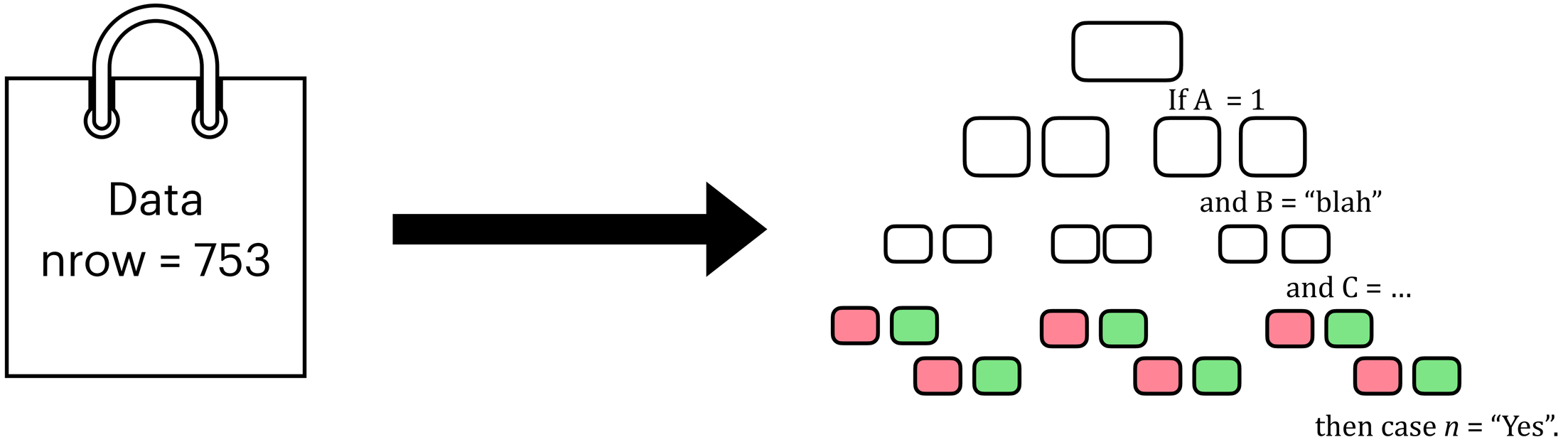




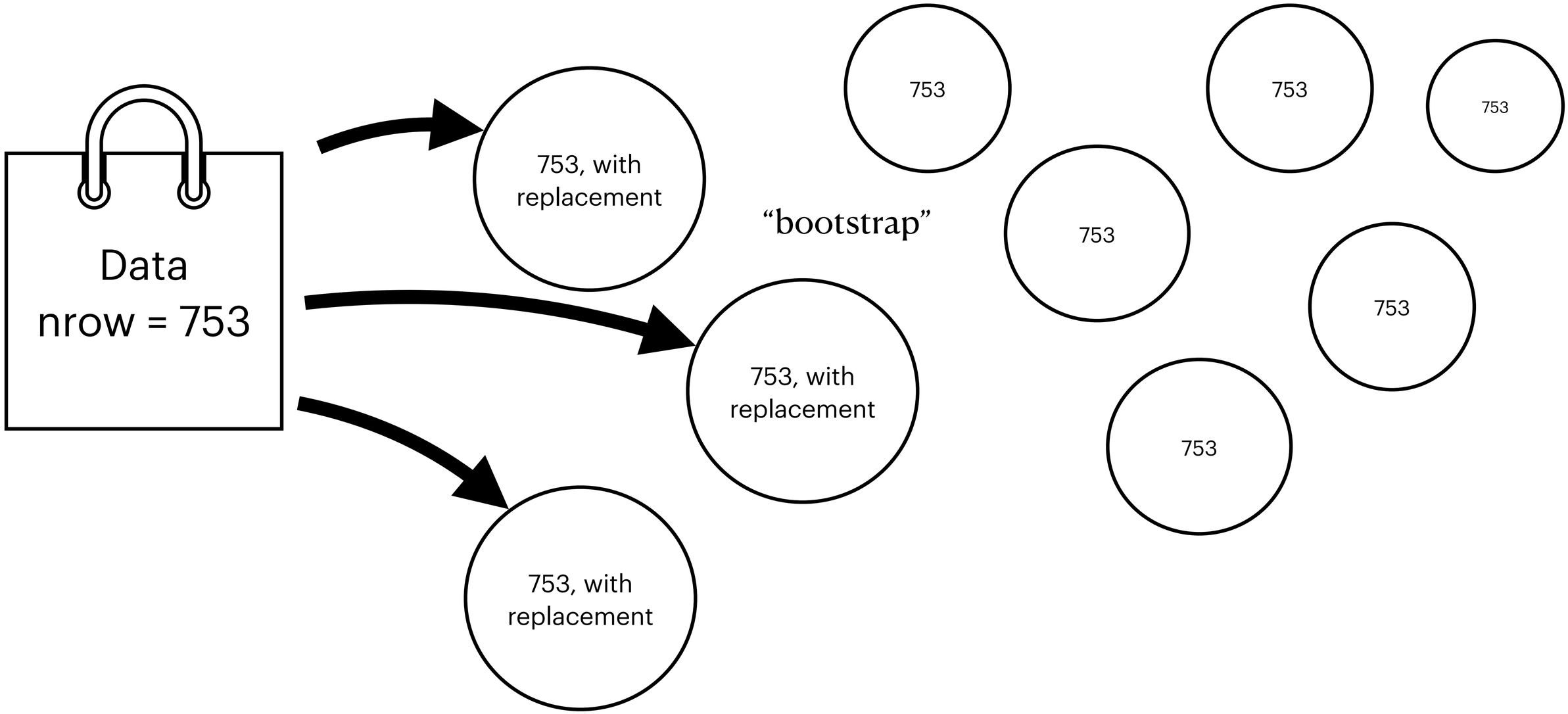
Pros & cons

LR	Single tree				
Interpretation?	Easy to interpret High variance				

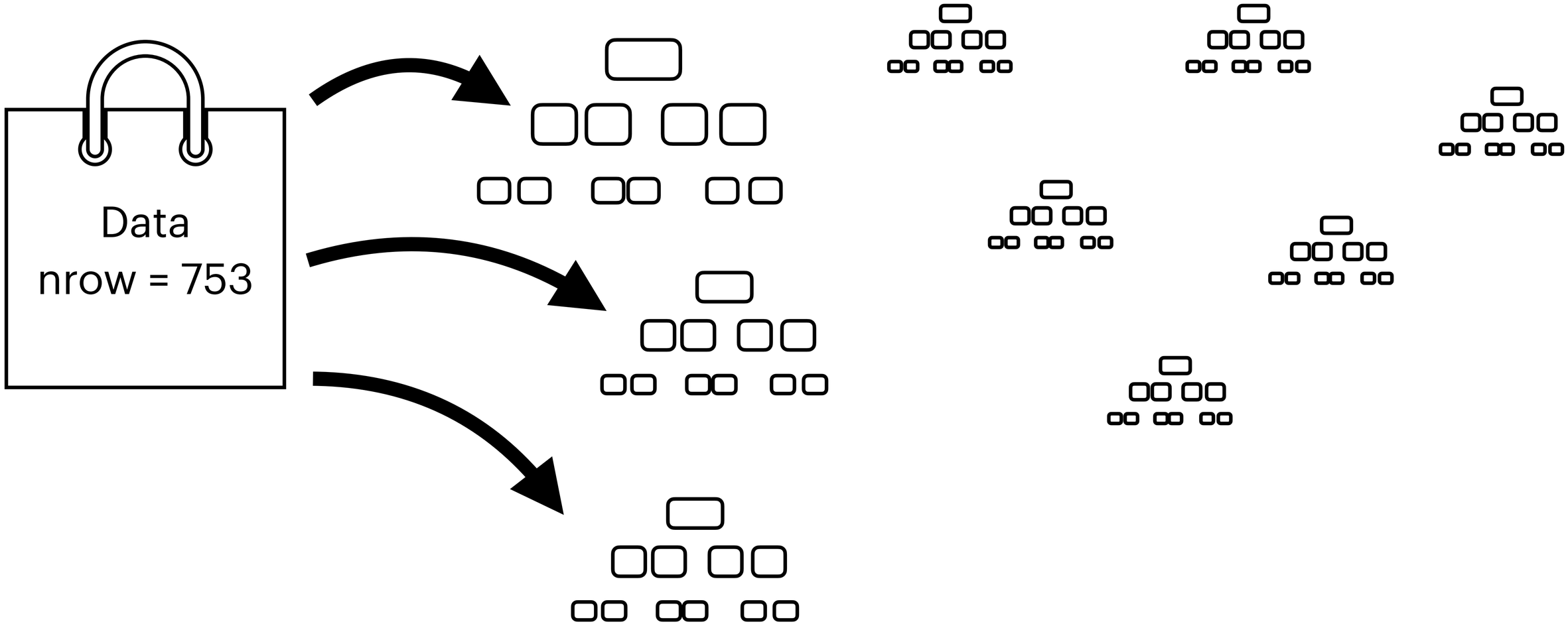
Classification tree



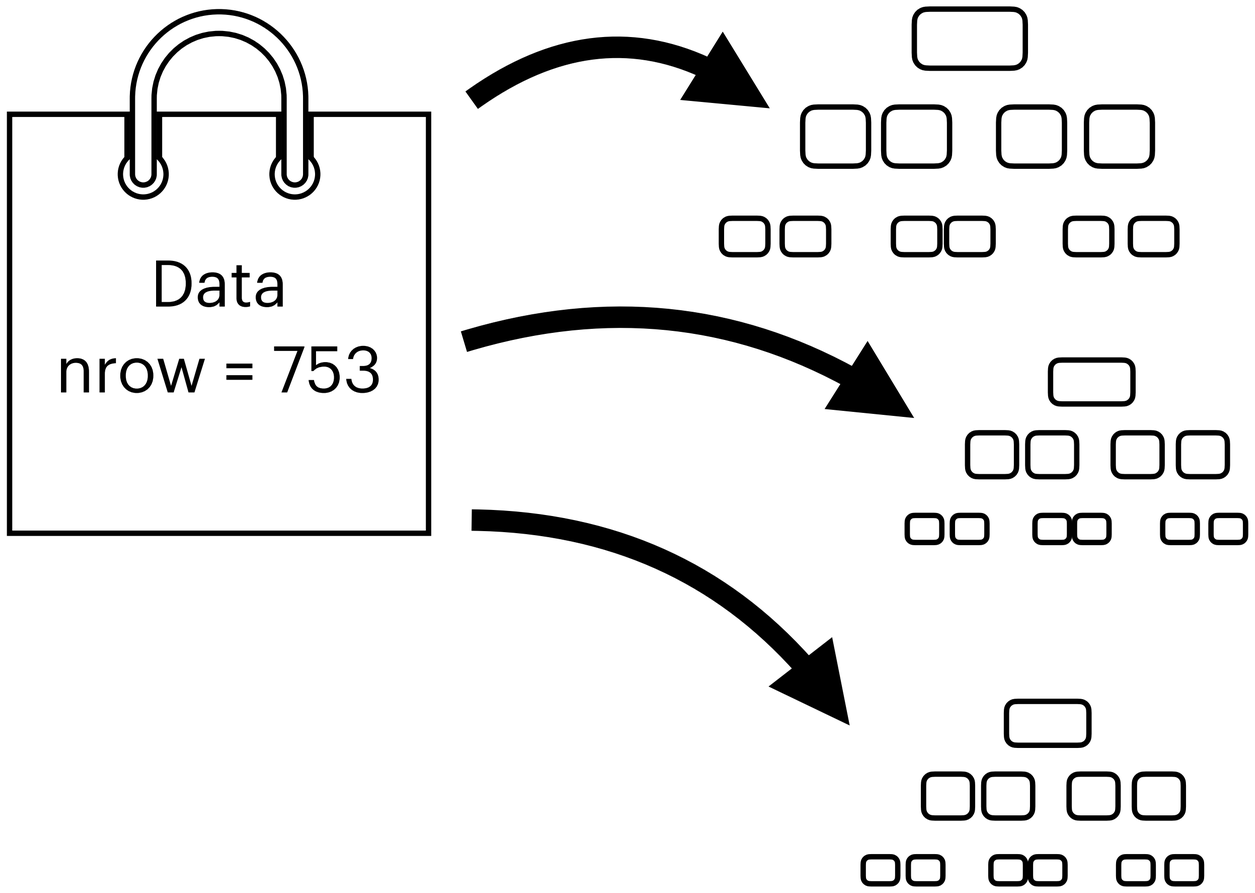
Bagging: Bootstrap aggregating



Bagging: Bootstrap aggregating



Bagging: Bootstrap aggregating



"aggregating"

	Tree 1	Tree 2	Tree ...	**Majority vote**
Case 1	Yes	No	No	**No**
Case 2	No	No	Yes	**No**
Case 3	Yes	No	Yes	**Yes**
...	No	Yes	No	**No**
Case 753	No	No	No	**No**

Bagging: Bootstrap aggregating

ntrees = 200

		Predicted	
		No	**Yes**
Actual	**No**	604	0
	Yes	100	49

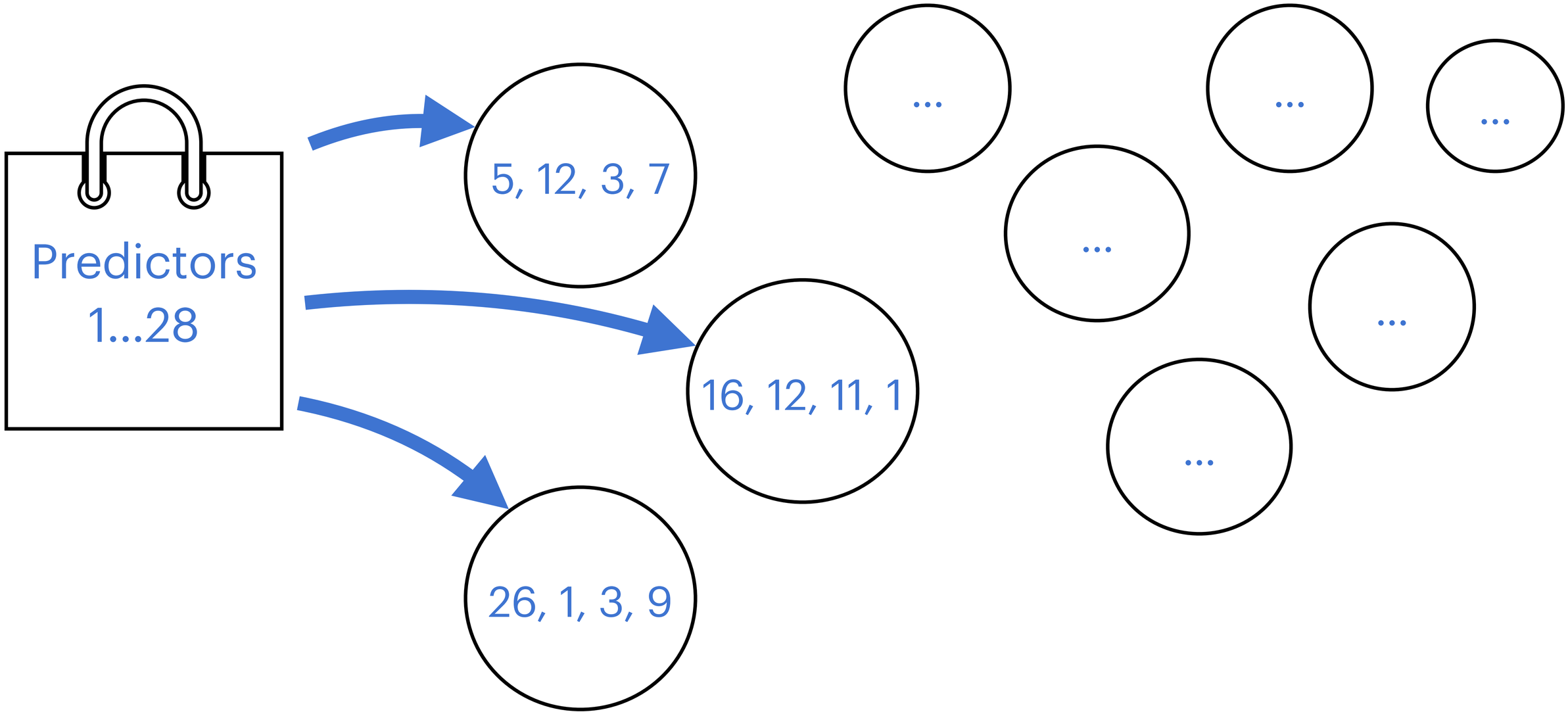
	LR	Tree	**Bag**
Accuracy	64	65	**87**
Sensitivity	59	81	**33**
Specificity	64	61	**100**

*Unable to correct for imbalanced classes

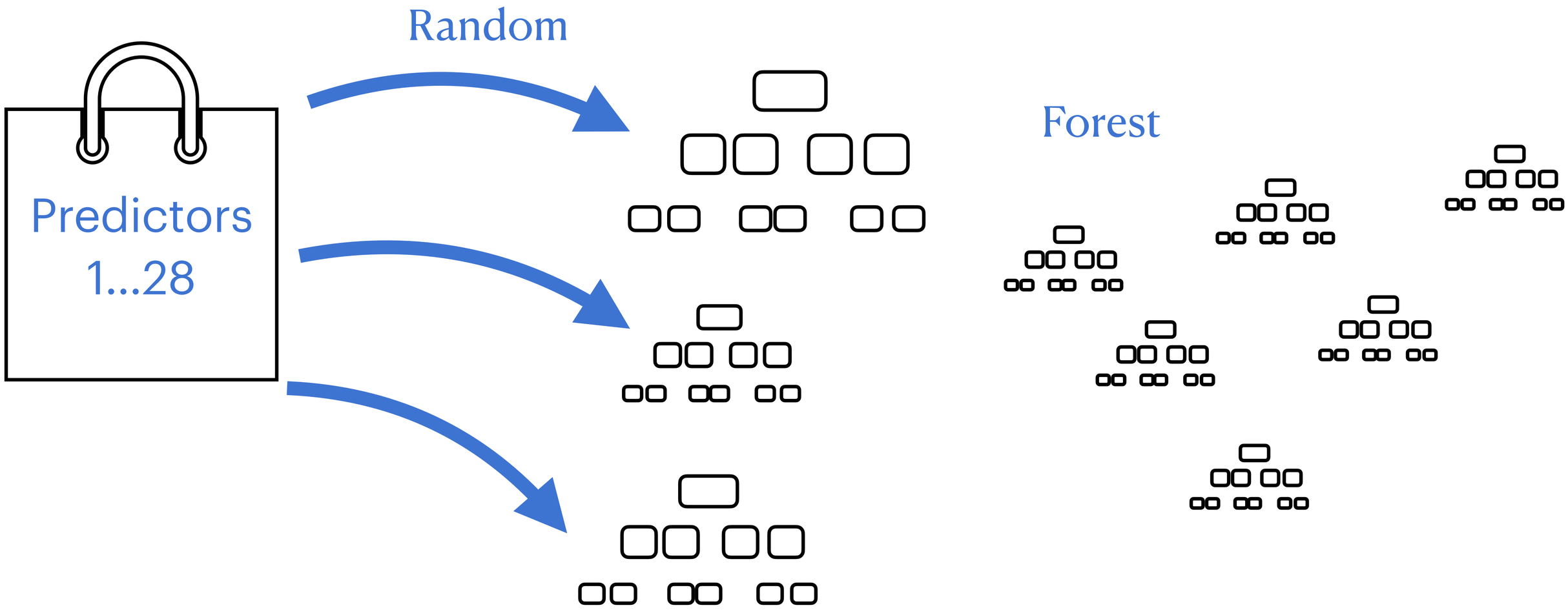
Pros & cons

LR	Single tree	Bagging		
Interpretation?	Easy to interpret	Less easy to interpret		
	High variance	Less variance		
		All trees look alike... less accurate		

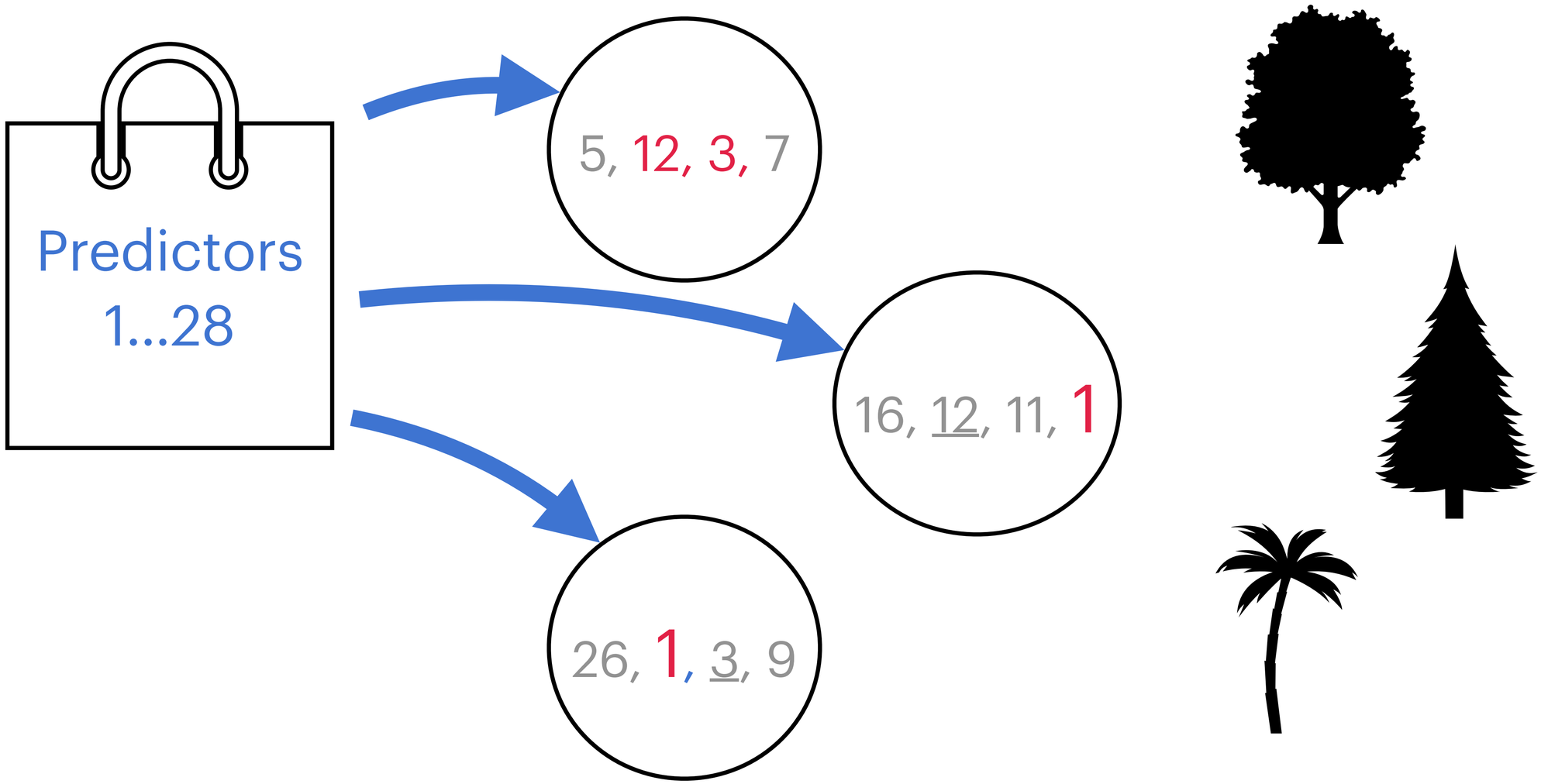
Random forest



Random forest



Random forest



Random forest

$ntrees = 200$

$m = 4$

		Predicted	
		No	**Yes**
Actual	**No**	519	85
	Yes	99	50

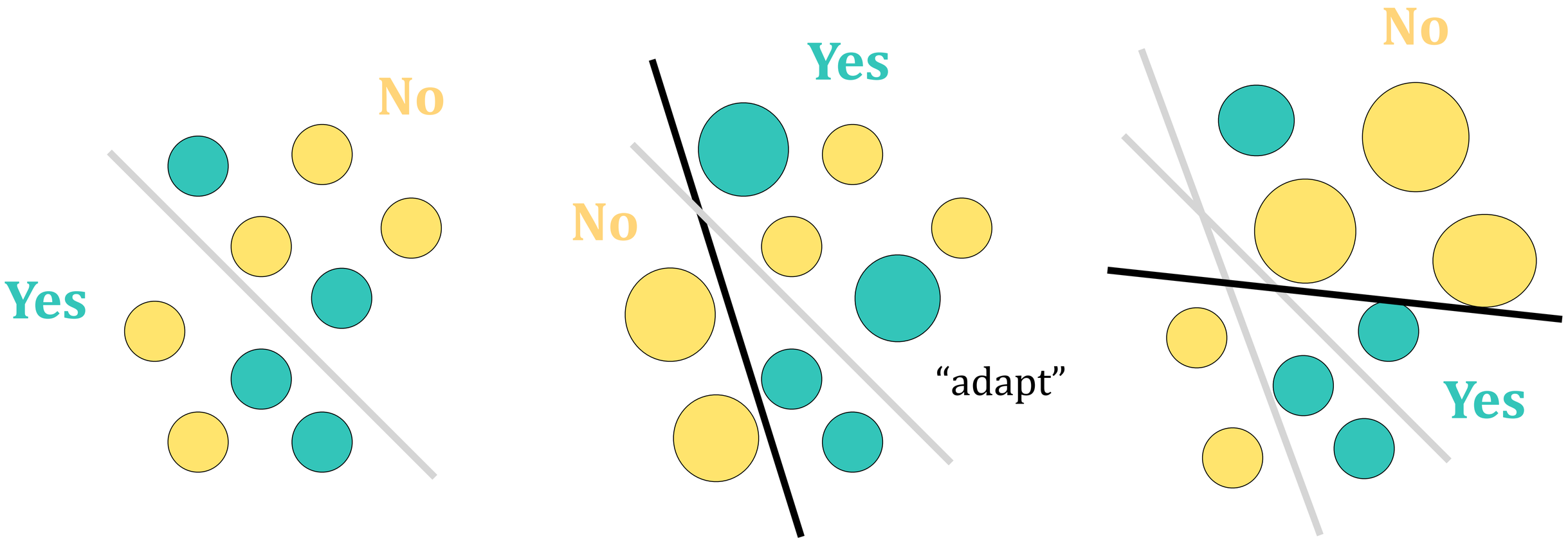
	LR	Tree	Bag	**RF**
Accuracy	64	65	87	**77**
Sensitivity	60	81	33	**34**
Specificity	64	61	100	**86**

Even with oversampling of “Yes” cases?!?

Pros & cons

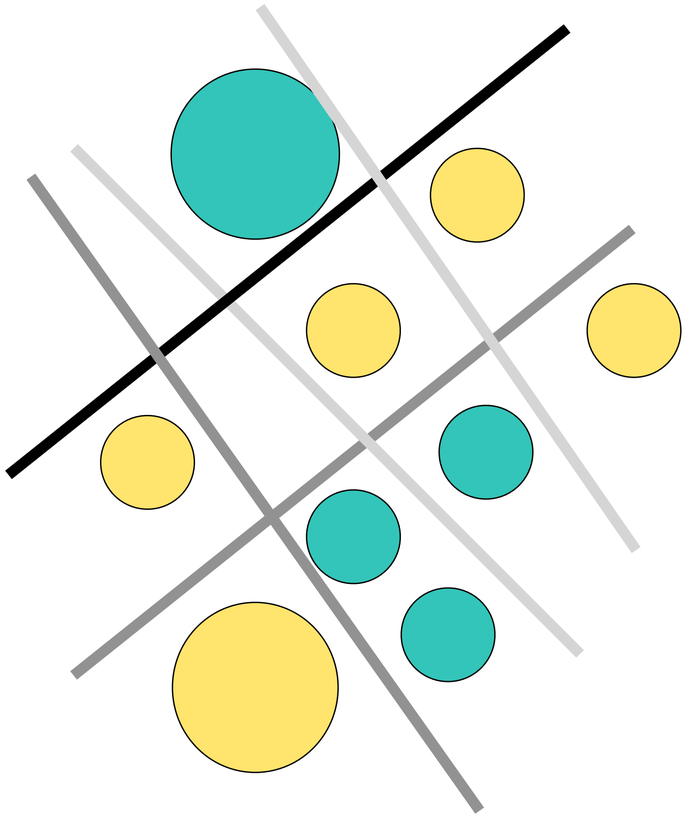
LR	Single tree	Bagging	Random forest	
Interpretation?	Easy to interpret	Less easy to interpret	Less easy to interpret	
	High variance	Less variance	Less variance	
		All trees look alike… less accurate	De-correlated trees; more accurate ???	

AdaBoost: Adaptive boosting



Tree 1 Tree 2 Tree 3

AdaBoost: Adaptive boosting



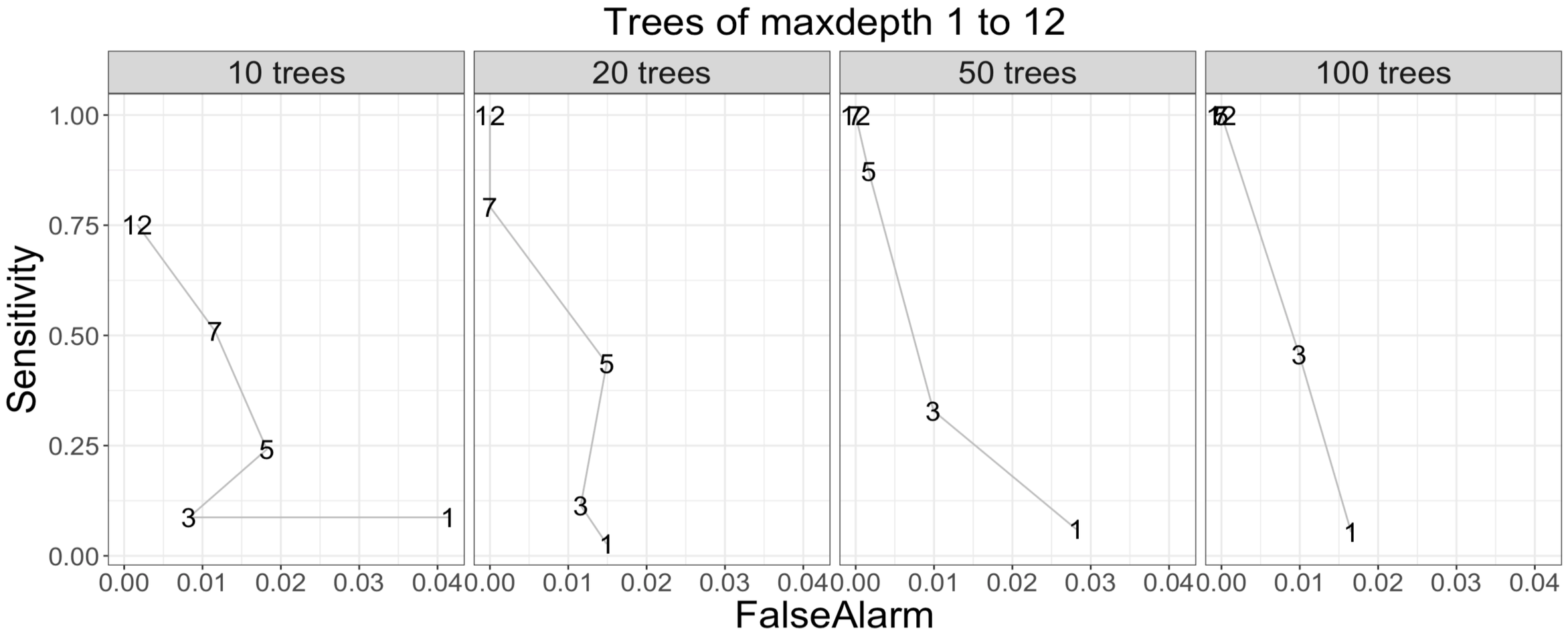
Tree *Gazillion!*

Over-fitting

Final model = (W1)Tree 1 + (W2)Tree 2 + (W3)Tree 3

Others: Gradient Boosting, XGBoost, LightGBM

AdaBoost: Adaptive boosting



AdaBoost: Adaptive boosting

mfinal = 10 trees
maxdepth = 12

5-fold CV,
stratified sampling

	LR	Tree	Bag	RF	**Ada(1)**	**Ada(cv)**
Accuracy	64	65	87	77	**98**	**77**
Sensitivity	60	81	33	34	**88**	**13**
Specificity	64	61	100	86	**100**	**92**

Pros & cons

LR	Single tree	Bagging	Random forest	Boosting
Interpretation?	Easy to interpret	Less easy to interpret	Less easy to interpret	Less easy to interpret
	High variance	Less variance	Less variance	Less variance
		All trees look alike… less accurate	De-correlated trees; more accurate	Build on previous trees; very accurate??
			Can’t overfit	Possible to overfit

Trees, trees, trees

- Ensembles are usually better than a single tree; “wisdom of the crowd”
- Quirks of the data
- How useful are the results from that method?
- How open is the field to that method?
- Each method has parameters to tune; for excruciating details: GitHub