SBESC XIII Brazilian Symposium on Computing Systems Engineering
November 21-24, 2023 - Porto Alegre - RS
sponsor SBC

RDSF: Everything at Same Place All at Once - A Random Decision Single Forest

Olavo A. B. Silva, Alysson K. C. Silva, Ícaro G. S. Moreira, José A. M. Nacif, Ricardo S. Ferreira
Universidade Federal de Viçosa

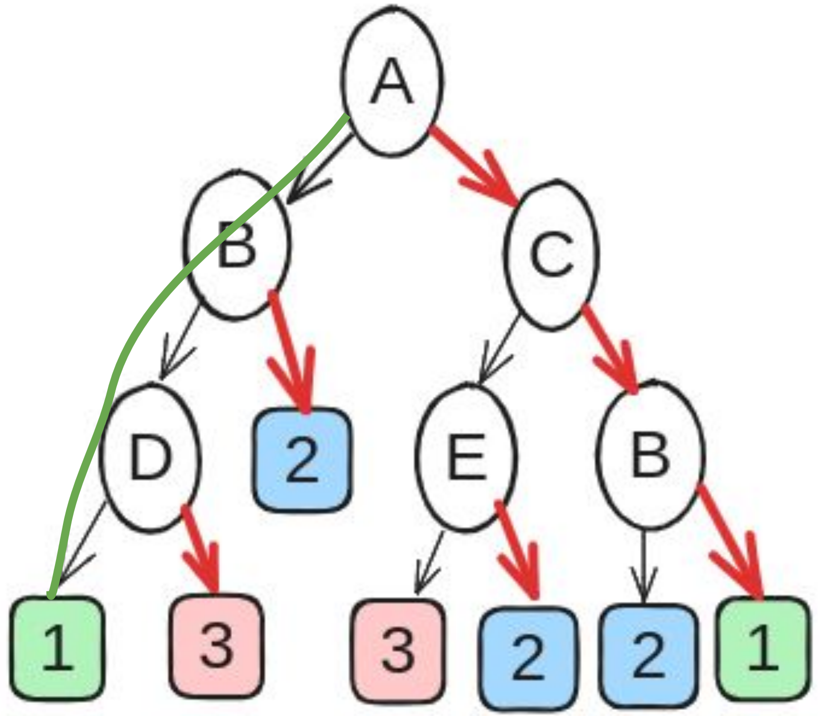
CNPq FAPEMIG CAPES

Summary

- Decision Trees and Random Forest
- Mapping a Random Forest in a single Decision Graph
    - Random Decision Single Forest (RDSF)
    - BIt adder Function, Majority and Priority Encoder
    - Binary Decision Diagram BDD
- Experimental Results
- Conclusion

SBESC

Decision Tree and Boolean Equations: Three Classes



3 Classes: 1, 2, and 3

$$\boxed{1} = \bar{A}\bar{B}\bar{D}$$

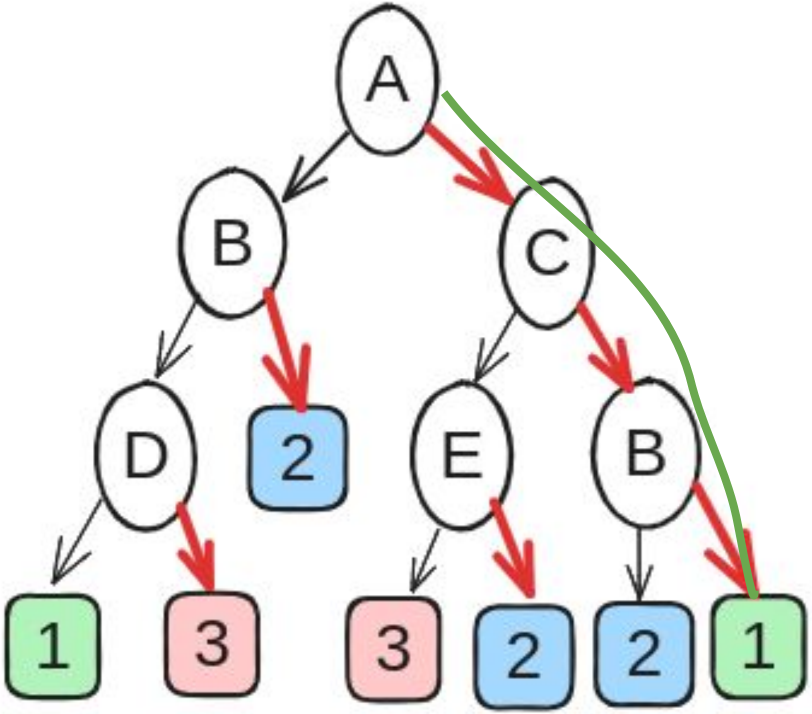
A= Length > 15.2 ?

B= Depth > 3.8 ?

…

SBESC

Each path generates one product term

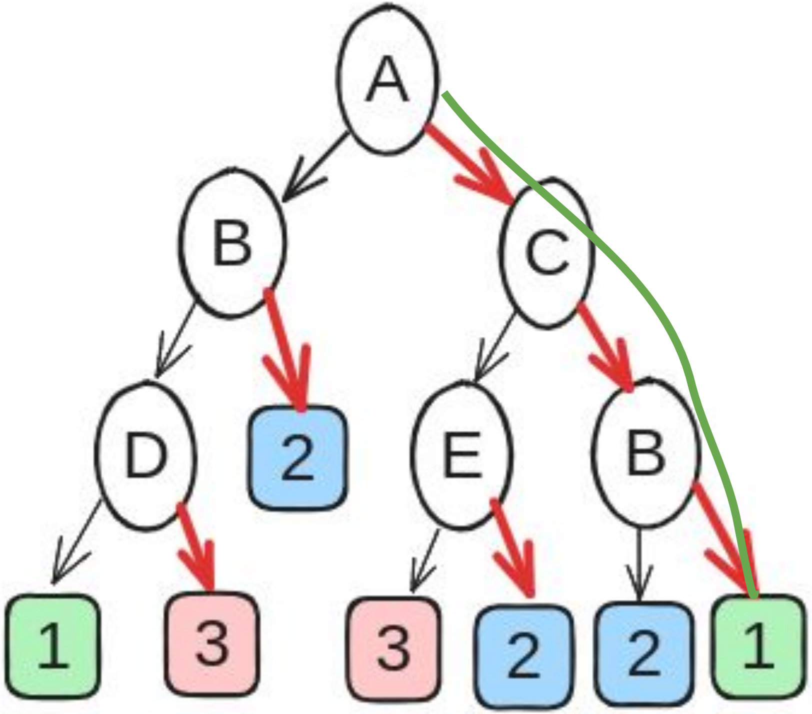


3 Classes: 1, 2, and 3

$\boxed{1} = \overline{A}\,\overline{B}\,\overline{D}$ or ABC

SBESC

Each class has a conjunctive logic equation



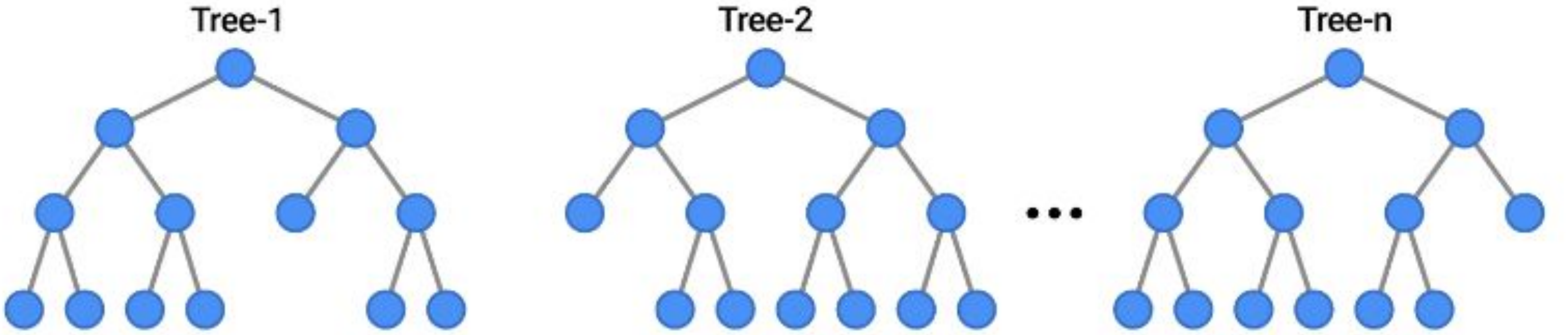
3 Classes: 1, 2, and 3

$\boxed{1}$ $= \bar{A}\bar{B}\bar{D}$ or ABC

$\boxed{2}$ $= \bar{A}B$ or $A\bar{C}E$ or $AC\bar{B}$

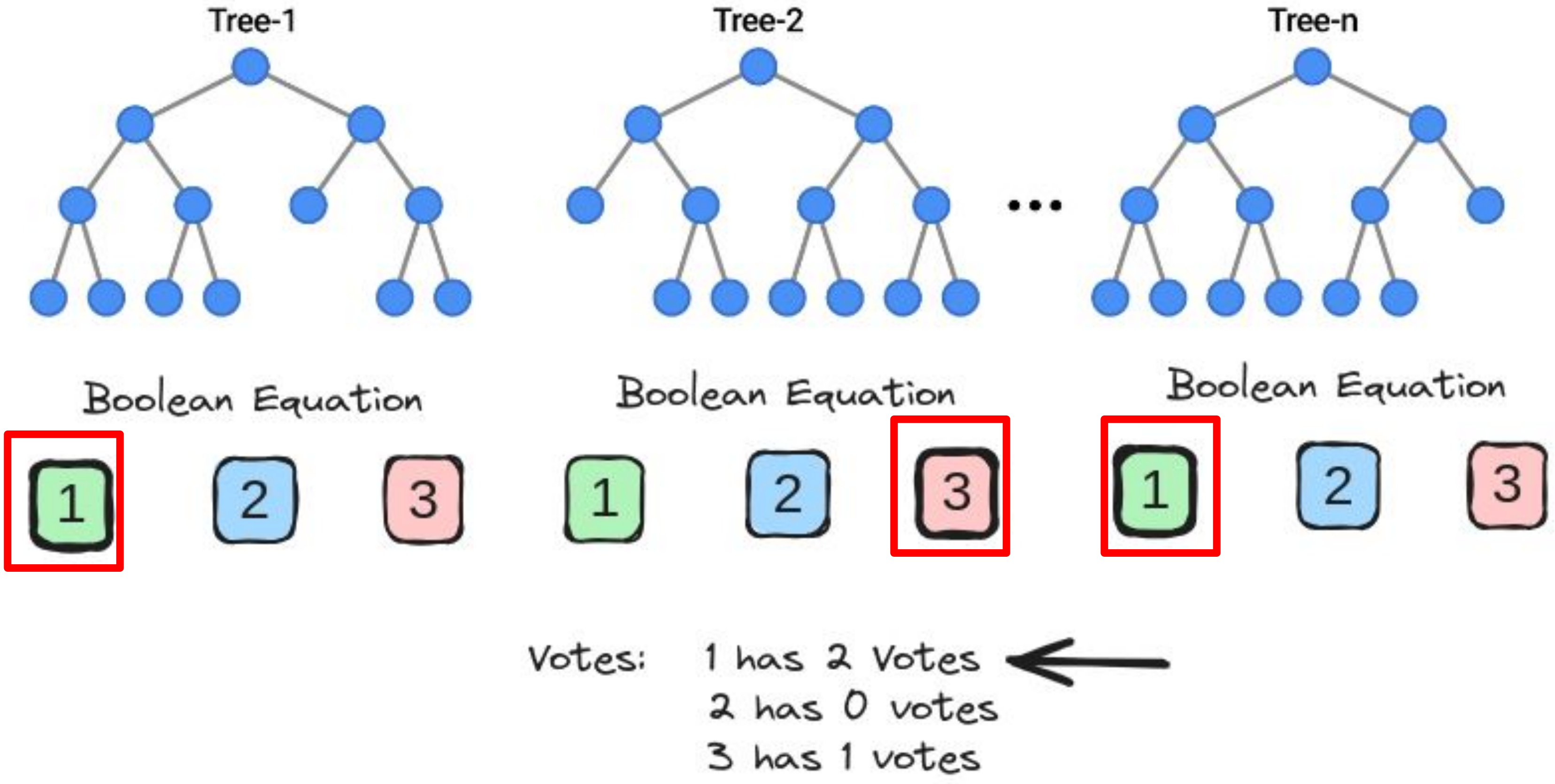
$\boxed{3}$ $= \bar{A}\bar{B}D$ or $A\bar{C}\bar{E}$

Random Forest

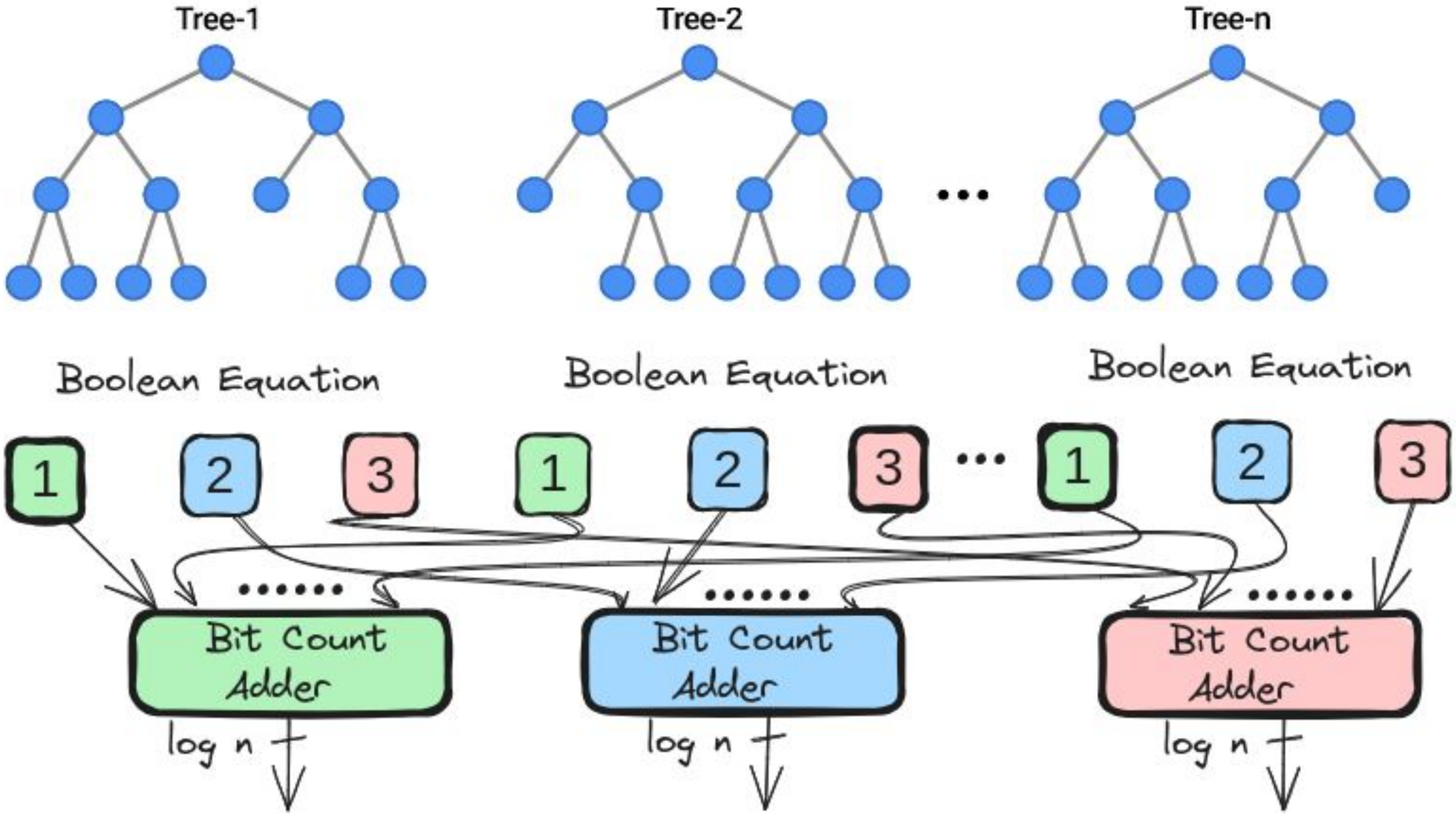


- Resilience against outliers and noisy data.
- By employing an ensemble of decision trees, Random Forest effectively reduces the vulnerability of individual trees to overfitting
- Enhancing its prediction accuracy.

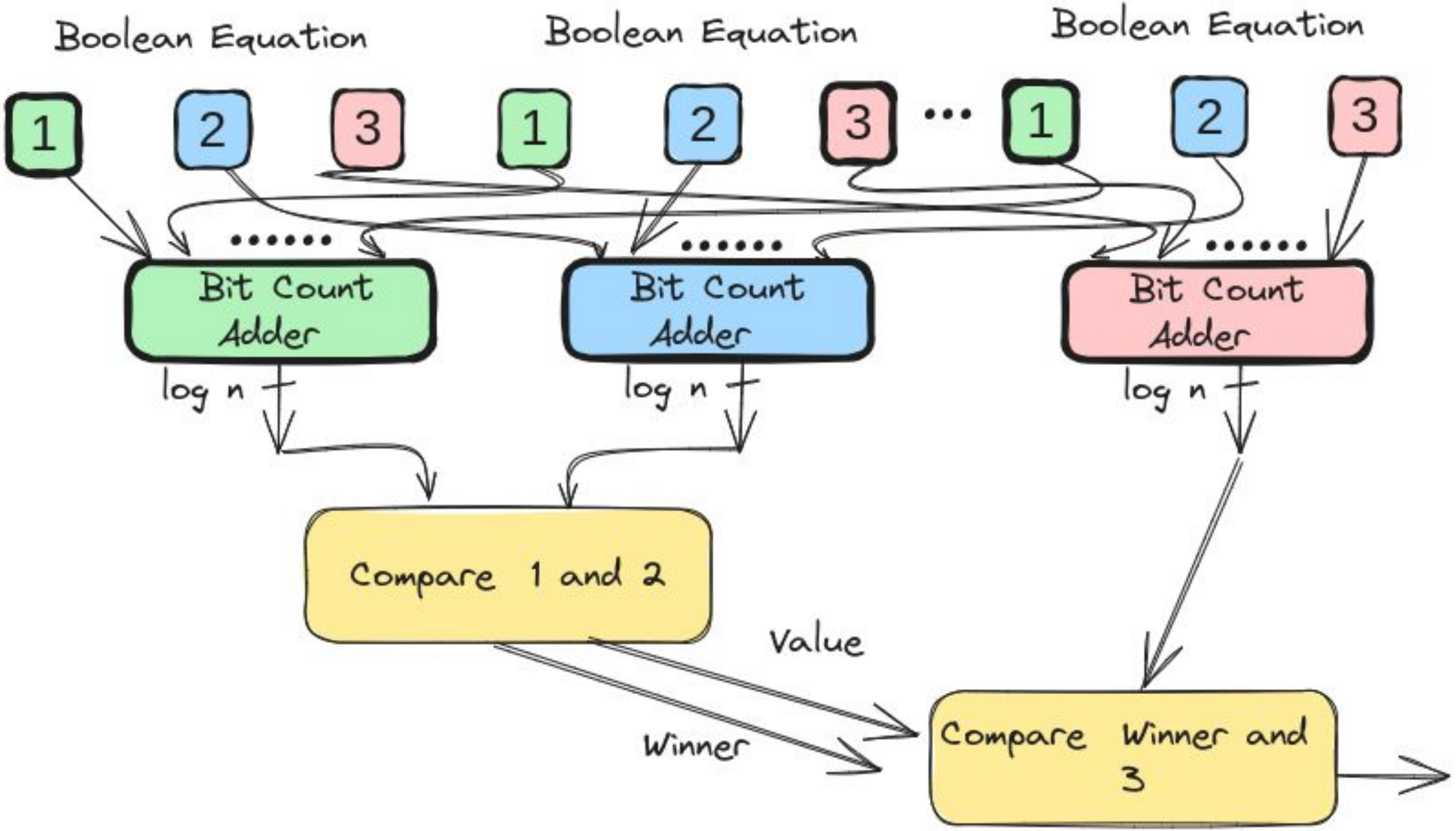
Random Forest



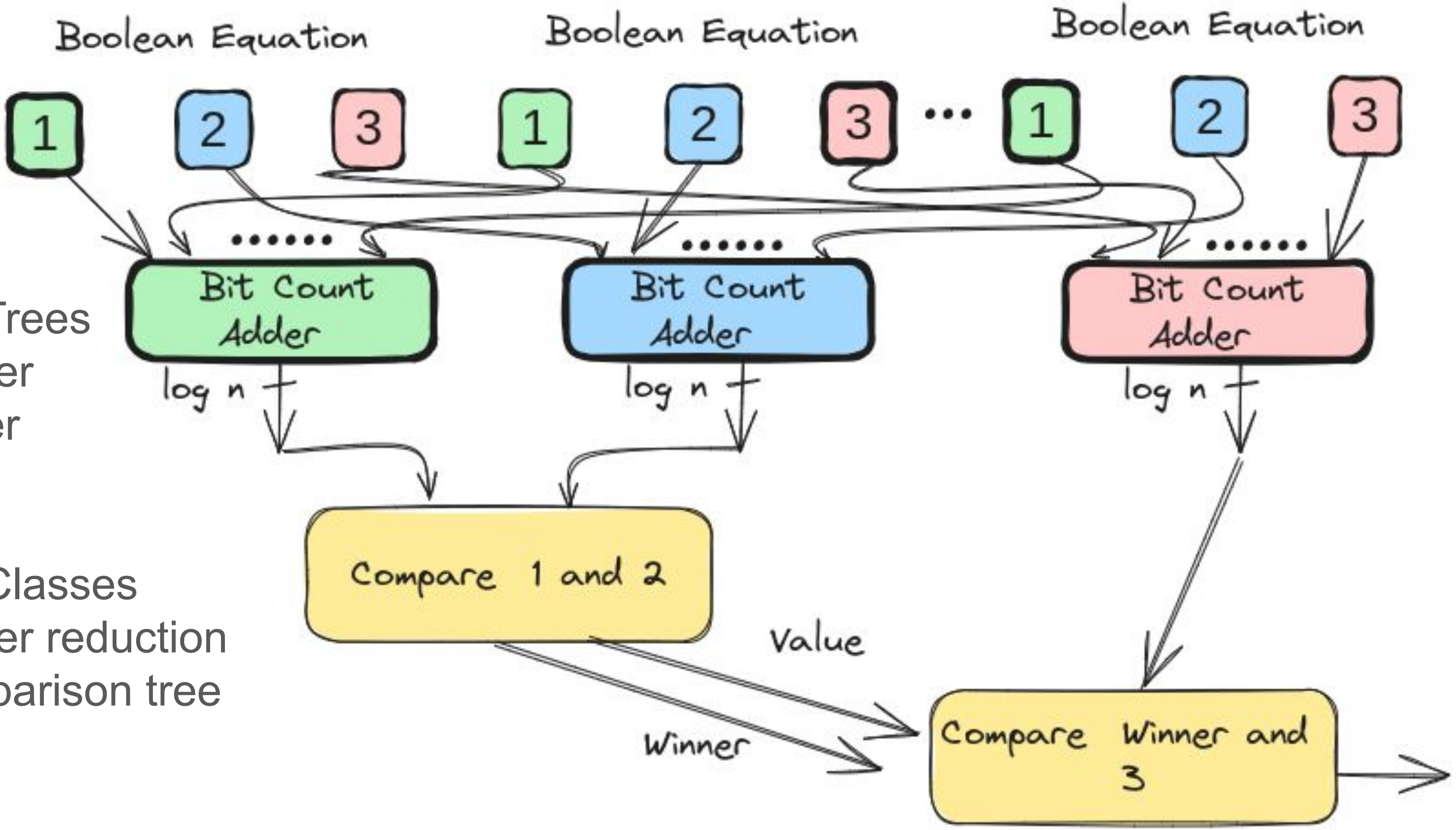
Random Forest



Random Forest



Random Forest



More Trees
= Larger
Adder

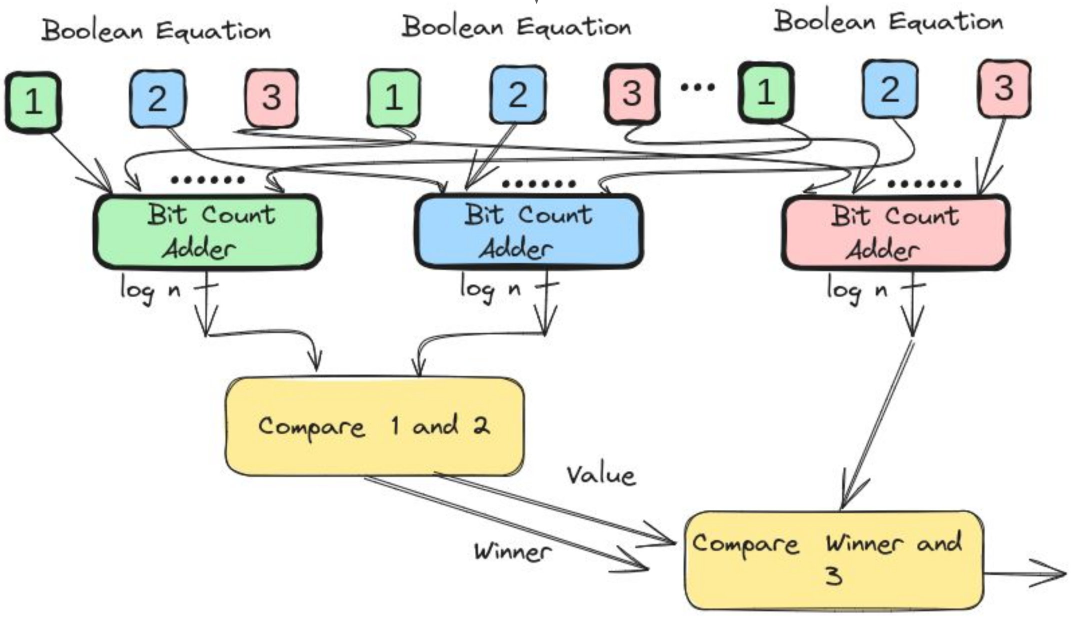
More Classes
= Larger reduction
comparison tree

SBESC

Our Proposal: Generate a single Boolean Function

Features

A= Length > 15.2 ?, B= Depth > 3.8 ?, …, F = width > 1.2 ?



Random
Decision
Single
Forest

Boolean Function

1 0 = **Class 2**

no adder, no maj

How to do that ? embedded the adder …..

- Tree 1 generates “paths” equation Tree1Class1 = A & B | C
- Tree 2 generates “paths” equation Tree2Class1 = D’ & E | A
- By using an “adder function operator”, we could merge Trees 1 and 2
- For X and Y, sum = X ^ Y and carry on = X & Y

- S1 = (A&B|C) ^ (D’&E|A) = F(A,B,C,D,E)

- Con1 = (A&B|C) & (D’&E|A) = F(A,B,C,D,E)

What to do next to “adding the class trees” ?

- Each Class generates a Sum Function Vector with Log N bits, for N trees
- Assuming a simple example of 3 trees => 2 bits
  - Class1Sum = $\boxed{C1S1}\boxed{C1S0}$
  - Class2Sum = $\boxed{C2S1}\boxed{C2S0}$
  - Class3Sum = $\boxed{C3S1}\boxed{C3S0}$
- Compare Function
  - C1S1 = 1 means Class 1 has “**1**0” ou “**1**1” votes (2 or 3)
  - Therefore   Maj1 = C1S1 or ….
  - Maj2 = C2S1 or ….
  - Maj3 = C3S1 or …
- One-Hot Code Maj (C1,C2,C3) function(a,b,c,d,e)

What to do event of a tie in the voting process ?

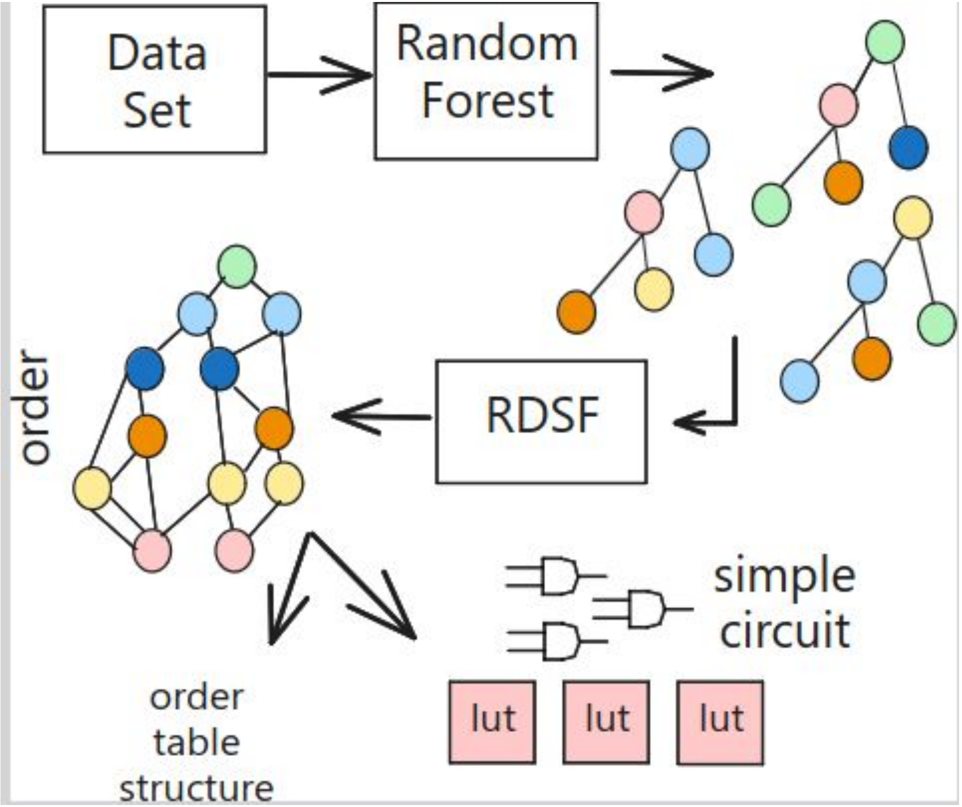
- Priority Encoder
- Assuming 4 class example One Hot Maj(c0,c1,c2,c3)
  - Encoder E1= C3 or C2
  - Encoder E0 = not E1 and C1 or C3
  - F= 0,0,1,1 E1E0 = 11 = C3
  - F= 1,0,1,0 E1E0 = 10 = C2
  - F= 1,1,0,0 E1E0 = 01 = C1

Embedded Boolean Function depends on the primary inputs

- Final = Encoder (Maj (Adding (primary input)))
- We use BDD to manipulate the Boolean Functions
  - compact
  - canonical representation
  - CUDD C++ efficient BDD package
  - Variable ORDER

Final Boolean Function depends on the primary inputs

BDD

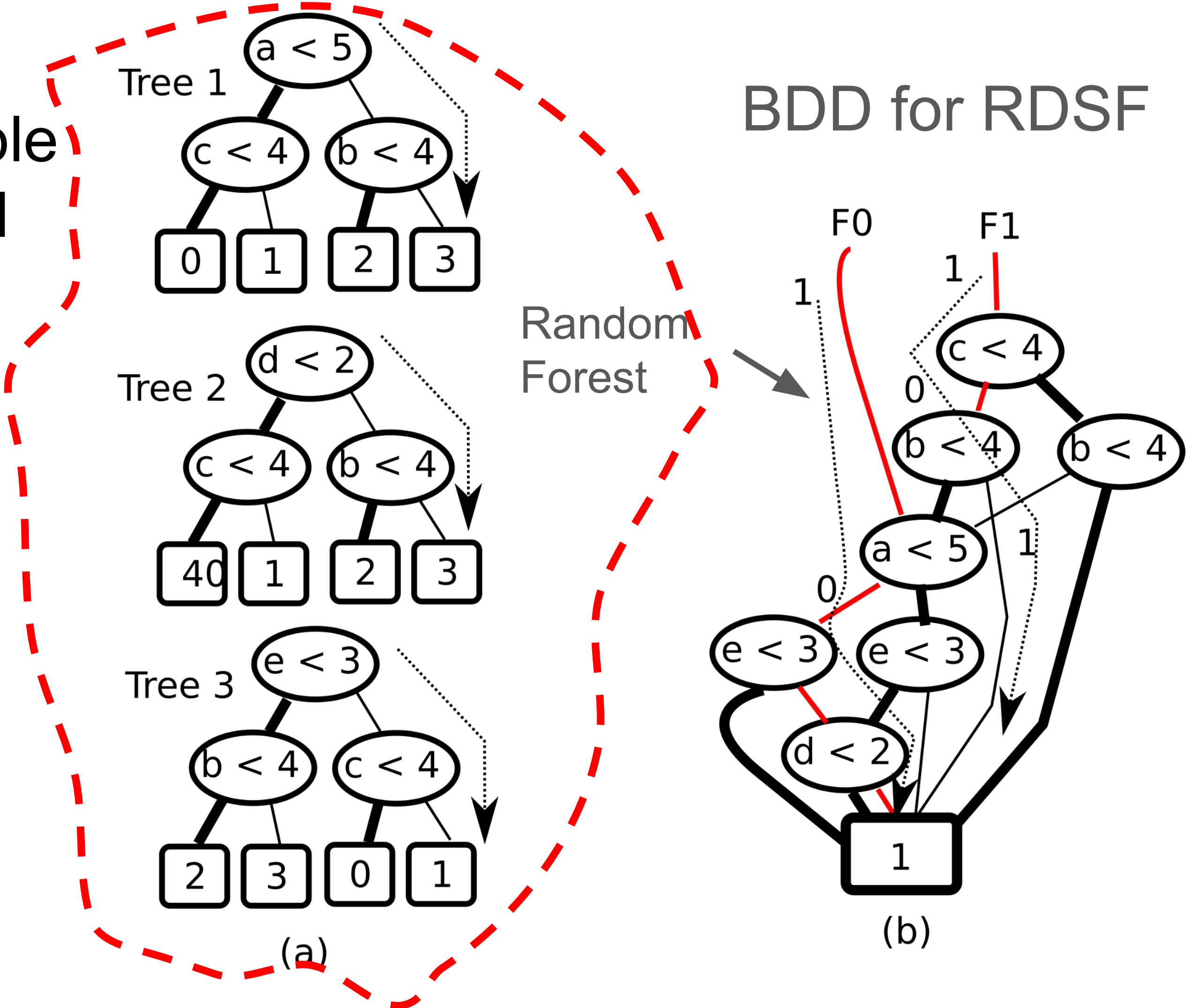


Random Forest

SBESC

A simple example with 3 trees and 3 classes

BDD for RDSF



(a)

(b)

SBESC

Features Order

Tree1

A, C, B

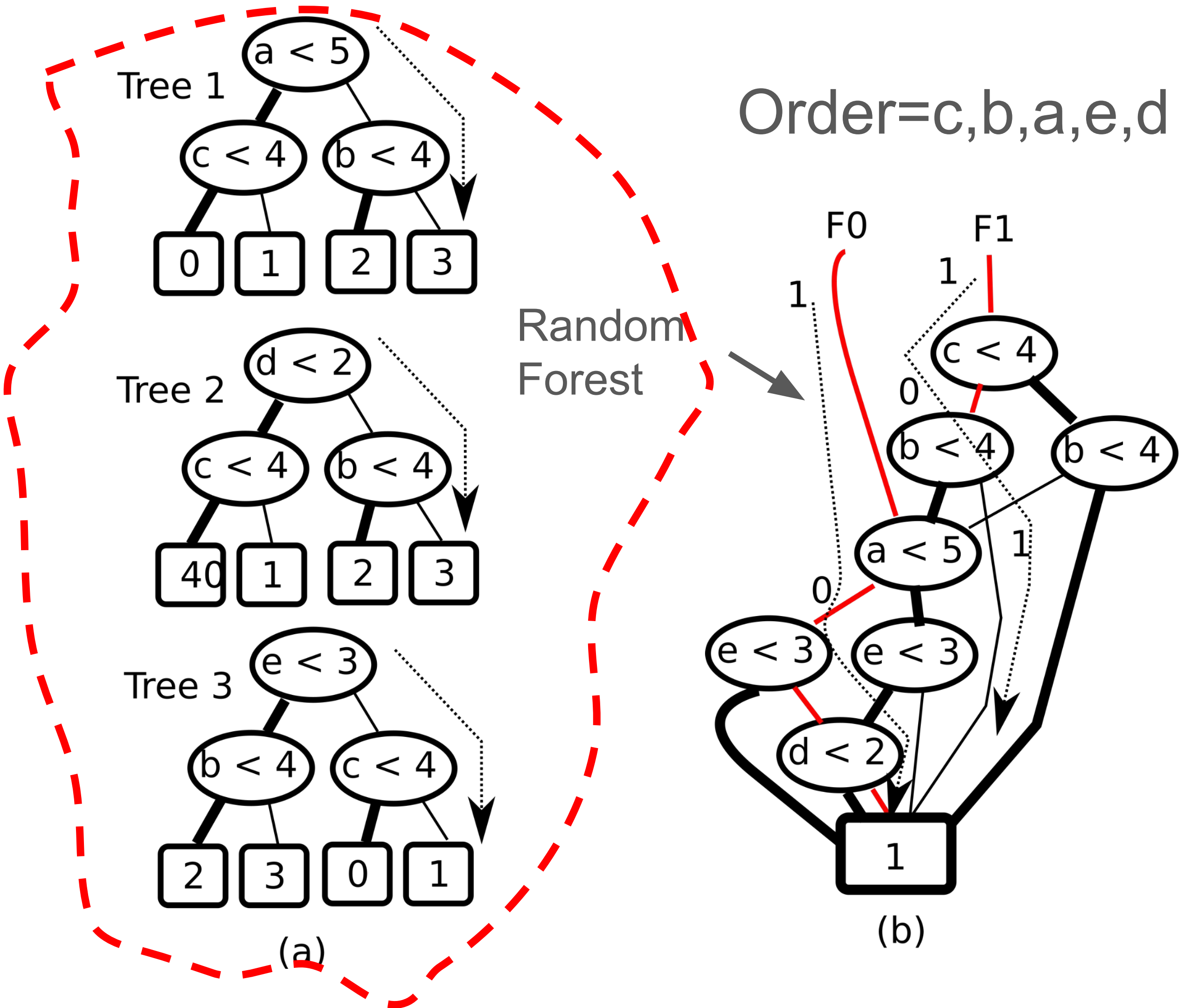
Tree 2

D,C, B

Tree 3

E, B, C

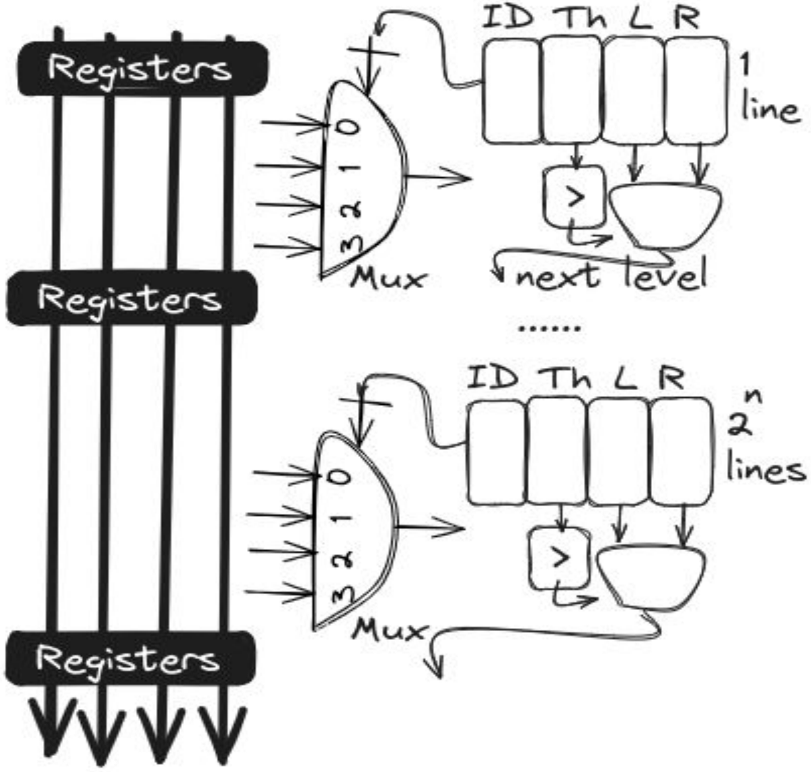
Order=c,b,a,e,d

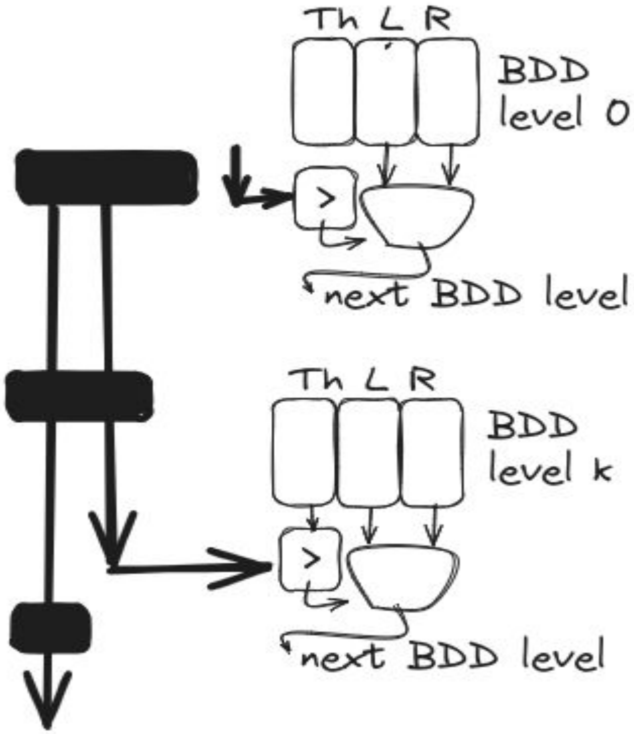


SBESC

Random Forest

Order=c,b,a,e,d





Experimental Results BDD versus Previous Approach

Execution Time (ms)					
		Generate		Inference	
Trees	Depth	CUDD	ADD-lib	CUDD	ADD-Lib
3	3	9	77	78	509
3	7	9	895	110	918
3	-	9	-	290	-
7	3	9	269	94	745

Dry beam Dataset 13K samples, 16 features, 4 classes

Experimental Results BDD generation Size

Tree	Depth	Var	Eq	Add	Vote	All
3	3	15	30	72	74	51
3	7	91	136	1364	1387	1006
3	-	271	377	29027	26059	20647
7	3	39	81	433	973	759
3	3	12	29	80	47	37

Dry beam Dataset 13K samples, 16 features, 4 classes

SBESC

Congress Voting, 16 attributes, 2 classes, 435 samples

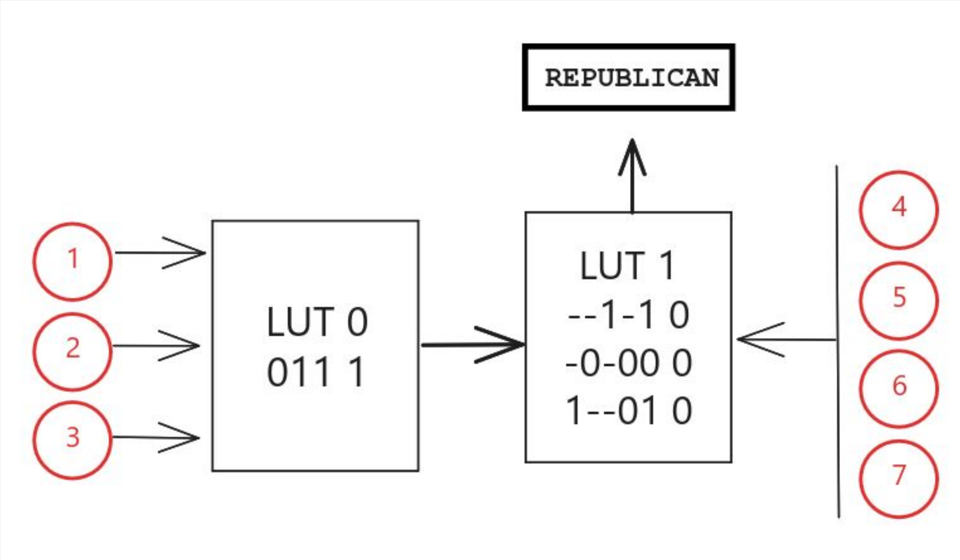
Tree	Depth	Var	Eq	Add	Vote
3	3	16	10	16	10
3	7	16	93	265	117
3	-	16	99	431	222
7	3	16	21	66	24
7	7	16	225	1639	470
7	-	16	289	2057	552

SBESC

Congress Voting, 16 attributes, 2 classes, 435 samples

Tree	Depth	Var	Eq	Add	Vote
3	3	16	10	16	10

1 - education spending
2 - immigration
3 - export administration act South Africa



4 - synfuels corporation cutback
5 - duty-free expor
6 - physician fee freeze
7 - adoption of the budget resolution

SBESC

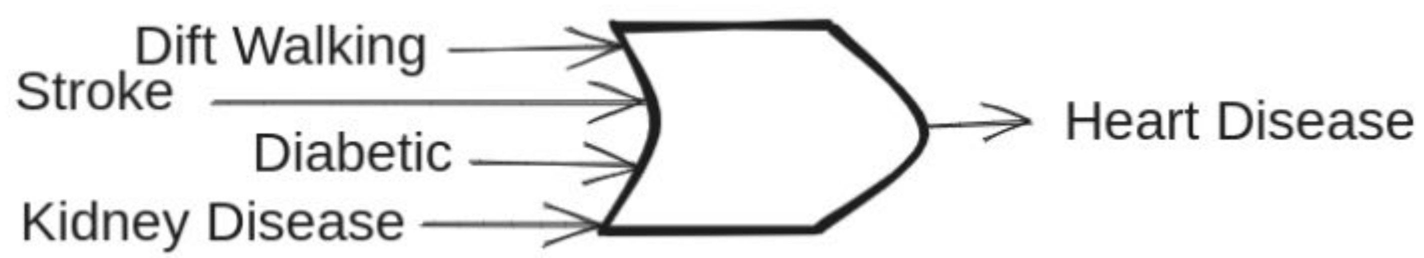
Heart disease, 9 features, 400k samples

Tree	Depth	Var	Eq	Add	Vote
3	3	9	5	7	5
3	7	9	7	13	10
3	-	9	189	147	74
7	3	9	8	13	6
7	7	9	218	163	46
7	-	9	383	213	79

SBESC

Heart disease, 9 features, 400k samples

Tree	Depth	Var	Eq	Add	Vote
3	3	9	5	7	5
3	7	9	7	13	10
3					
7					
7					
7					

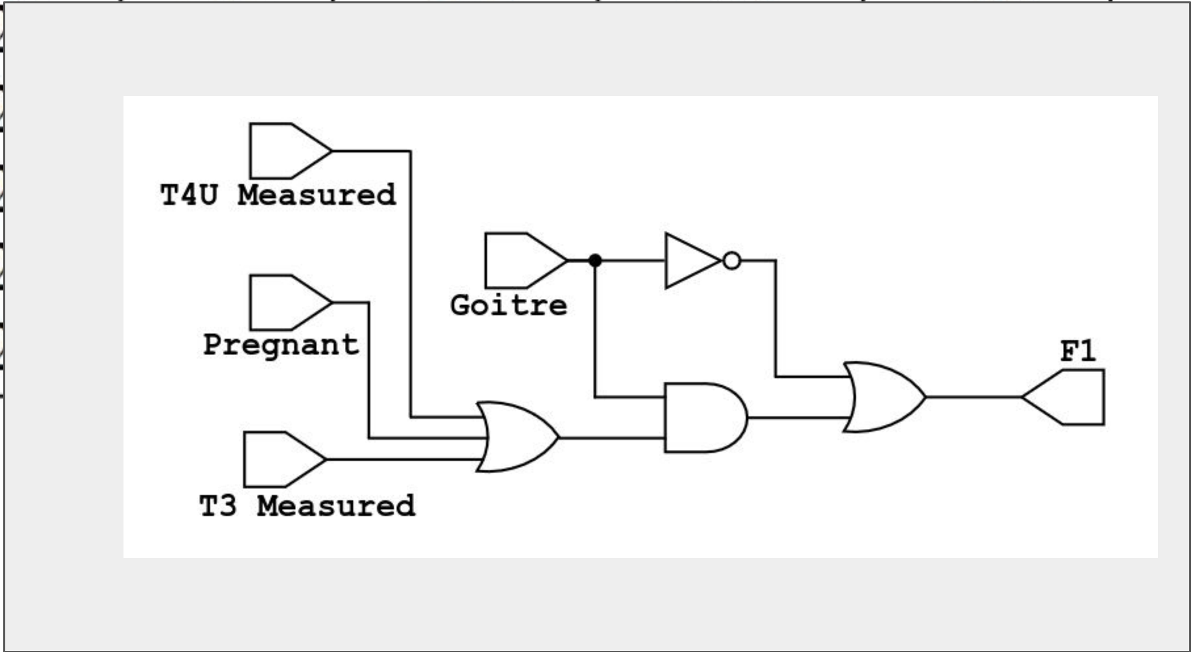


Thyroid, 20 features, 4 classes, 9k samples

Tree	Depth	Var	Eq	Add	Vote	All
3	3	20	5	7	5	5
3	7	20	105	282	132	109
3	-	20	364	875	333	287
7	7	20	18	53	19	15
7	7	20	252	2,612	868	718
7	-	20	968	7,055	1,383	1162

Thyroid, 20 features, 4 classes, 9k samples

Tree	Depth	Var	Eq	Add	Vote	All
3	3	20	5	7	5	5
3	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



Cluster Classification

- Maj vote (c0,c1,c2,c3,c4,c5,c6,c7)
- Maj vote (c0,c5,c6,c3,c4,c1,c2,c7)
- Maj vote (c0,c5,c6,c3,c4,c1,c2,c7)

Conclusions

- Random Decision Single Forest (RDSF)
- Improved scalability and reduced execution time compared to ADD approaches.

- Advantage of controlling input data order during inference
- Facilitating direct class clustering, enhancing its versatility.

- Numerical and categorical datasets.
- Mapping some datasets in simple functions for categorical ones

Questions ?

ricardo@ufv.br

Acknowledgments

Financial support from FAPEMIG APQ-01577-22, CNPq, and UFV. This work was also carried out with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brasil (CAPES) - Financing Code 001