Learning Optimal Decision Trees using Constraint Programming

CP2025 - Constraints Journal Prominent Paper Award

Hélène Verhaeghe[1], Siegfried Nijssen[1], Gilles Pesant[2], Claude-Guy Quimper[3], and Pierre Schaus[1]

13 August 2025

[1]ICTEAM, UCLouvain, Place Sainte Barbe 2, 1348 Louvain-la-Neuve, Belgium, {*firstname.lastname*}@*uclouvain.be*
[2]Polytechnique Montréal, Montréal, Canada, *gilles.pesant@polymtl.ca*
[3]Université Laval, Québec, Canada, *claude − guy.quimper@ift.ulaval.ca*

The Problem

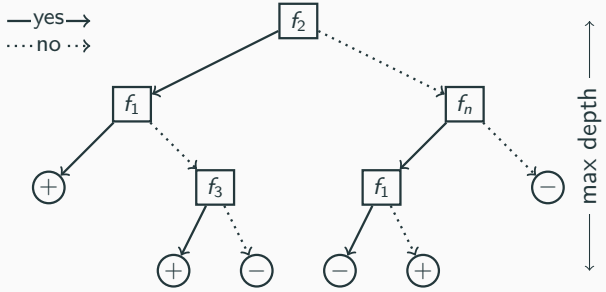
UCLouvain

Database					
f_1	f_2	f_3	$\dots$	f_n	c
1	0	1	$\dots$	1	$+$
0	1	0	$\dots$	1	$-$
1	1	0	$\dots$	0	$+$
0	0	0	$\dots$	0	$+$
1	0	0	$\dots$	0	$+$
0	1	1	$\dots$	1	$-$
1	1	1	$\dots$	0	$-$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
1	1	1	$\dots$	1	$+$

UCLouvain

Database

f_1	f_2	f_3	...	f_n	c
1	0	1	...	1	+
0	1	0	...	1	−
1	1	0	...	0	+
0	0	0	...	0	+
1	0	0	...	0	+
0	1	1	...	1	−
1	1	1	...	0	−
⋮	⋮	⋮	⋱	⋮	⋮
1	1	1	...	1	+



$$\min \sum (pred(i) - c(i))$$

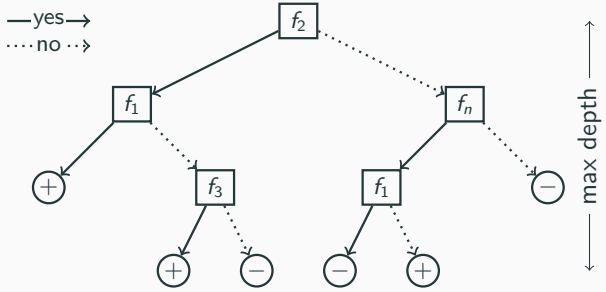
UCLouvain

Database

f_1	f_2	f_3	...	f_n	c
1	0	1	...	1	+
0	1	0	...	1	−
1	1	0	...	0	+
0	0	0	...	0	+
1	0	0	...	0	+
0	1	1	...	1	−
1	1	1	...	0	−
⋮	⋮	⋮	⋱	⋮	⋮
1	1	1	...	1	+

New sample

0	0	1	...	0	?



$$\min \sum (pred(i) - c(i))$$

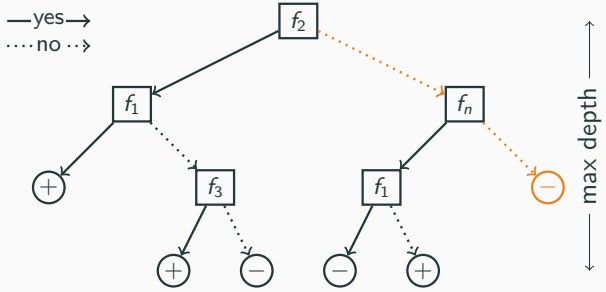
UCLouvain

Database

f_1	f_2	f_3	...	f_n	c
1	0	1	...	1	+
0	1	0	...	1	−
1	1	0	...	0	+
0	0	0	...	0	+
1	0	0	...	0	+
0	1	1	...	1	−
1	1	1	...	0	−
⋮	⋮	⋮	⋱	⋮	⋮
1	1	1	...	1	+

New sample

0	0	1	...	0	−



$$\min \sum (pred(i) - c(i))$$

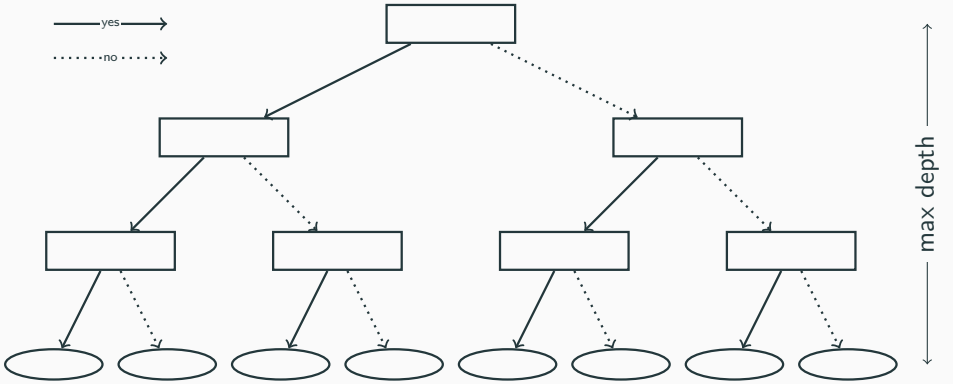
Greedy methods:

- ✔ easy construction
- ✗ hard to impose additional constraints
- ✗ potentially unnecessarily complex tree

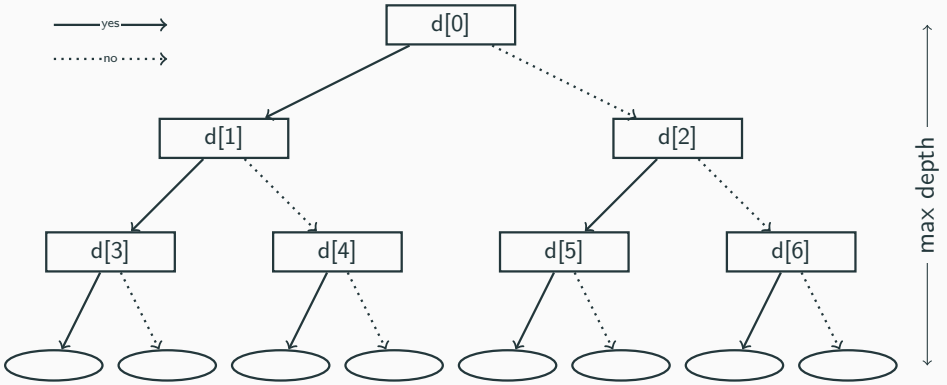
- Mining optimal decision trees from itemset lattices, Nijssen, S., Fromont, E., 2007
- Minimising decision tree size as combinatorial optimisation, Bessiere, C., Hebrard, E., O'Sullivan, B., 2009
- Optimal constraint-based decision tree induction from itemset lattices,Nijssen, S., Fromont, É., 2010
- **Optimal classification trees**, Bertsimas, D., Dunn, J., 2017
- Learning optimal decision trees with sat, Narodytska, N., Ignatiev, A., Pereira, F., Marques-Silva, J., RAS, I., 2018
- Learning optimal and fair decision trees for non-discriminative decision-making, Aghaei, S., Azizi, M.J., Vayanos, P., 2019
- Learning optimal classification trees using a binary linear program formulation, Verwer, S., Zhang, Y., 2019

CP Model

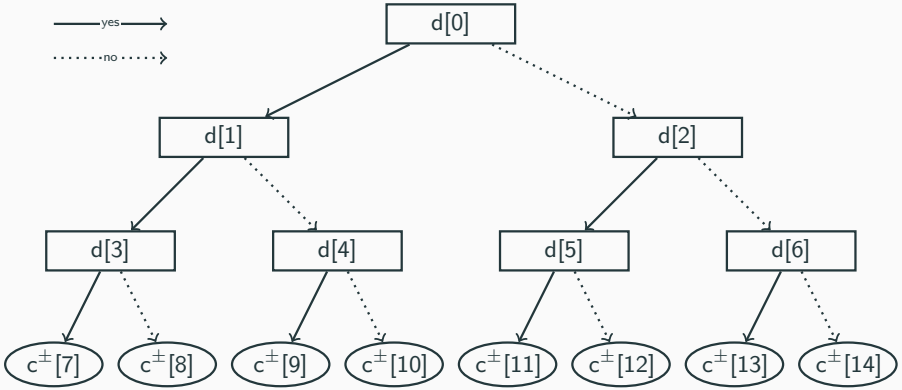
Model - Decision variables



Model - Decision variables

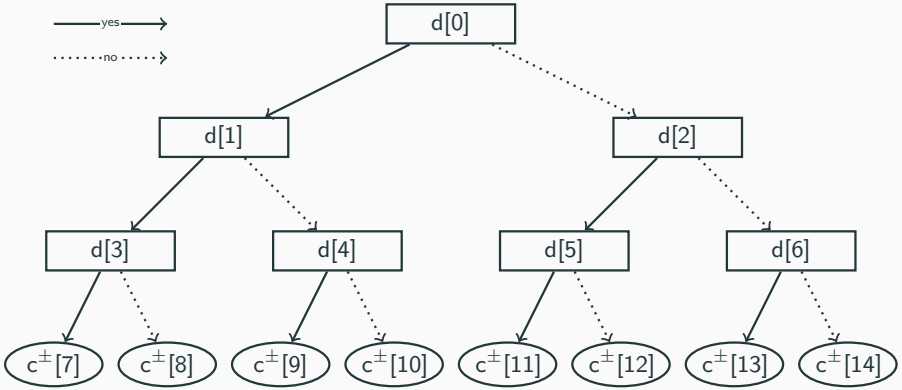


$dom(d[i]) = \{1, ..., n\}$



$dom(d[i]) = \{1, ..., n\}$

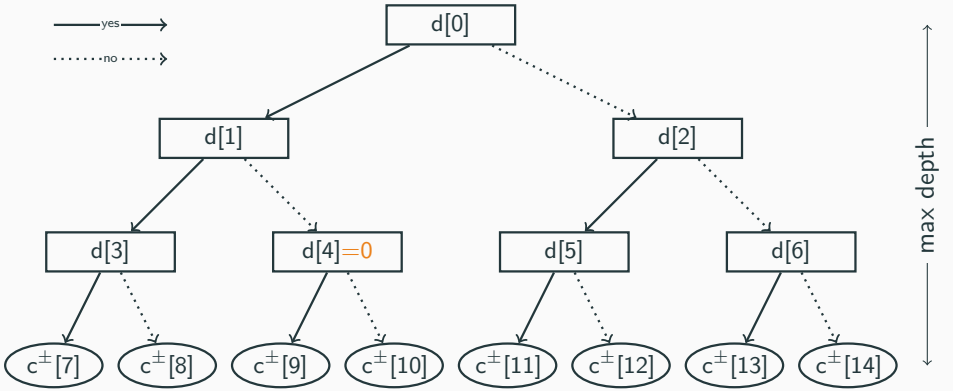
$dom(c[i]) = \{0, ..., N\}$



$dom(d[i]) = \{0, 1, ..., n\}$

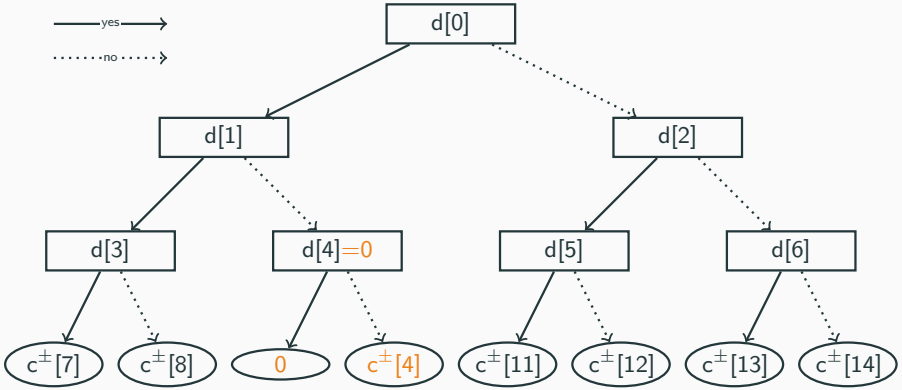
$dom(c[i]) = \{0, ..., N\}$

Model - Unbalanced Tree



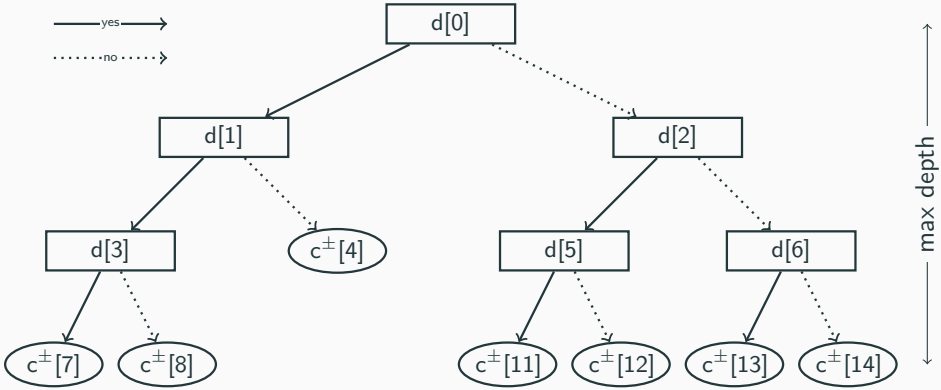
$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



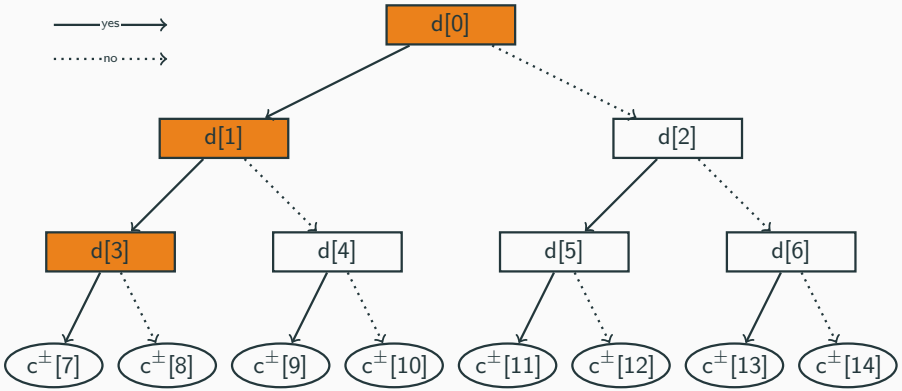
$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



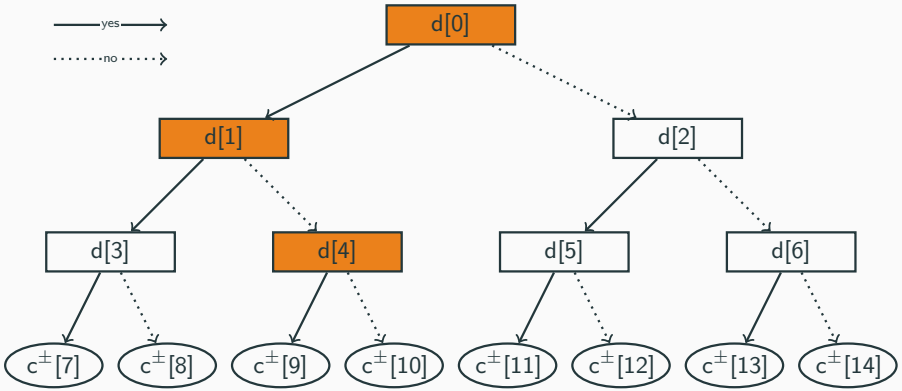
$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



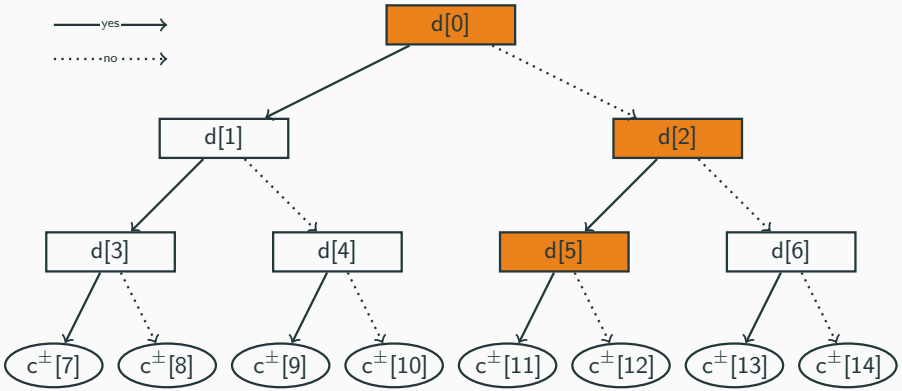
$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



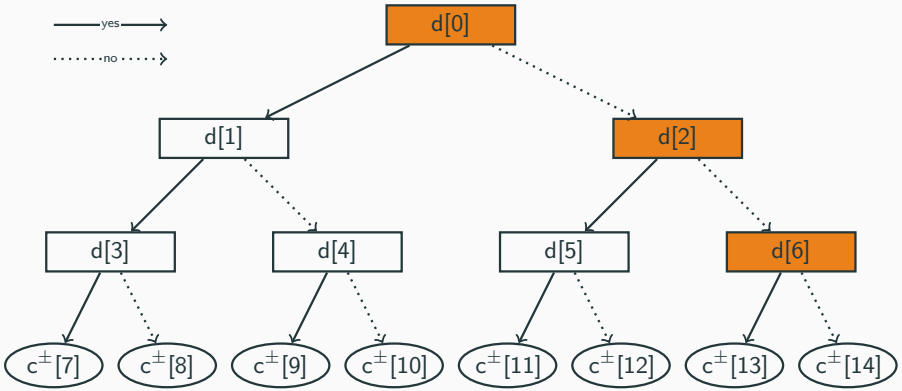
$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

Features (Dense)				Counter
x_1	x_2	x_3	x_4	

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

Features (Dense)				Counter
x_1	x_2	x_3	x_4	
0	1	0	1	

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

Features (Dense)				Counter
x_1	x_2	x_3	x_4	
0	1	0	1	3

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

Features (Dense)				Counter
x_1	x_2	x_3	x_4	
0	1	0	1	3

- Dense representation
- No feature rejection

UCLouvain

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

Features (Sparse)		Counter
y_1	y_2	
2	4	3

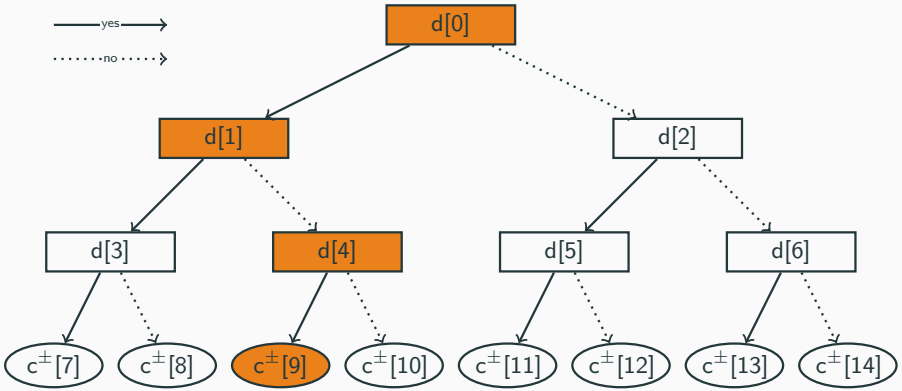
- ~~Dense representation~~
- No feature rejection

UCLouvain

f_1	f_2	f_3	f_4
1	0	1	1
0	1	0	1
1	1	0	0
0	0	0	0
1	0	0	0
0	1	1	1
1	1	1	0
1	1	1	1

✔Features (Sparse)		✗Features (Sparse)	Counter
y_1	y_2	z_1	
2	4	3	1

- ~~Dense representation~~
- ~~No feature rejection~~



$Coversize(\{d[0], d[4]\}, \{d[1]\}, c^{+}[9])$

$Coversize(\{d[0], d[4]\}, \{d[1]\}, c^{-}[9])$

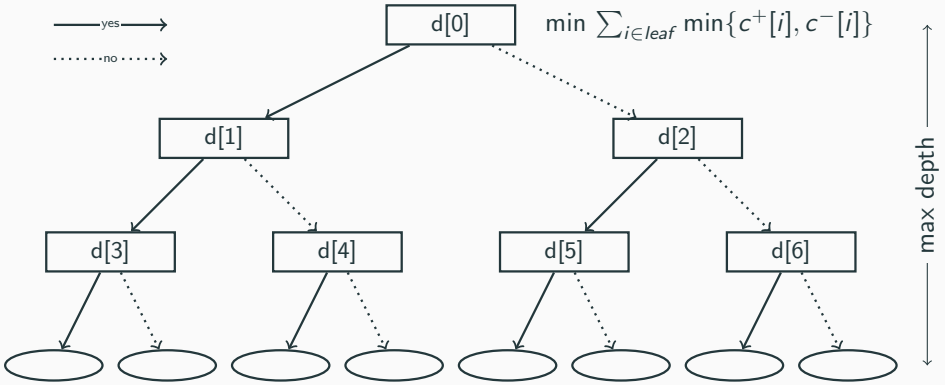
UCLouvain

- constraints imposing minimum at leaf

$$c^+[i] + c^-[i] \geq N_{min}$$

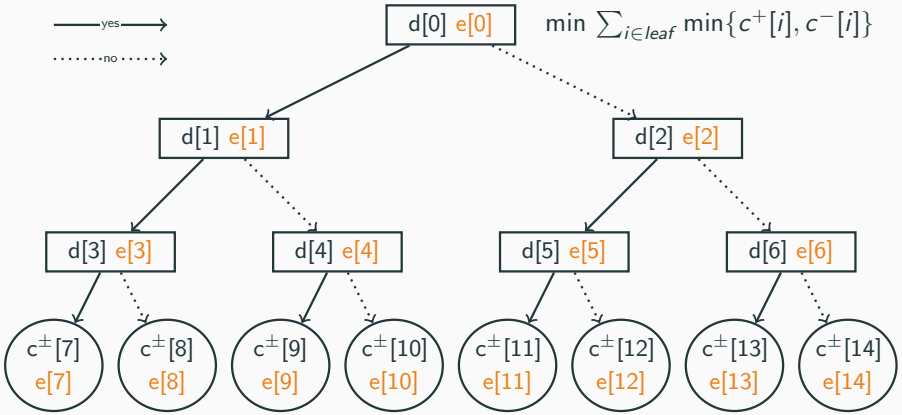
- constraints avoiding useless decisions

- redundant constraints improving speed

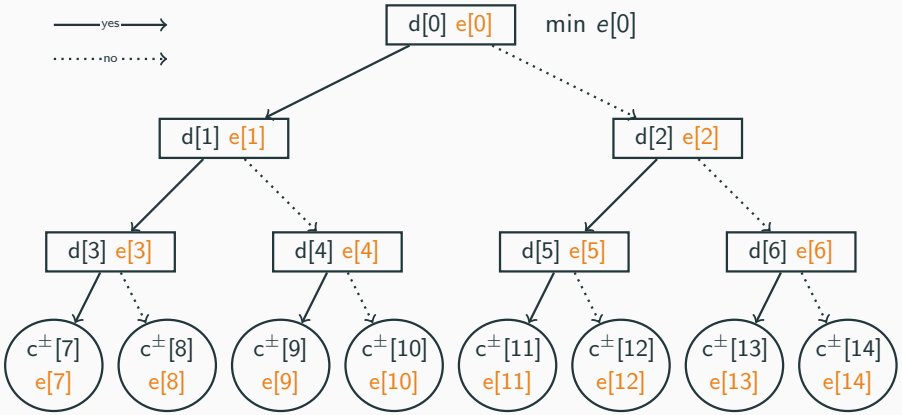


$dom(d[i]) = \{0, 1, ..., n\}$

$dom(c[i]) = \{0, ..., N\}$



$dom(d[i]) = \{0, 1, ..., n\}$ $dom(c[i]) = \{0, ..., N\}$ $dom(e[i]) = \{0, ..., N\}$



$dom(d[i]) = \{0, 1, ..., n\}$ $dom(c[i]) = \{0, ..., N\}$ $dom(e[i]) = \{0, ..., N\}$

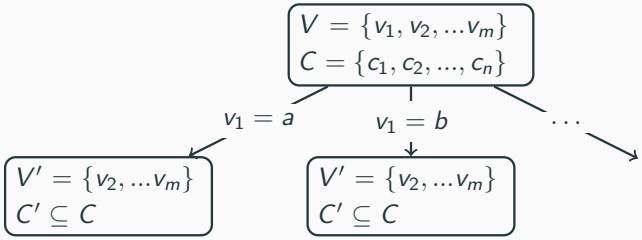
Search

UCLouvain

$V = \{v_1, v_2, ...v_m\}$
$C = \{c_1, c_2, ..., c_n\}$

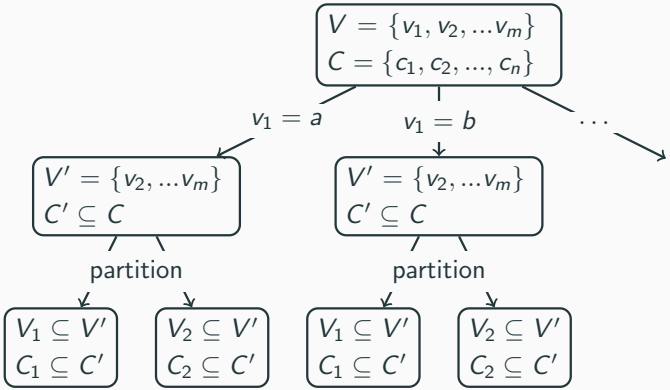
OR nodes

SOL = SOL_1 or SOL_2 or ...



OR nodes

SOL = SOL_1 or SOL_2 or ...

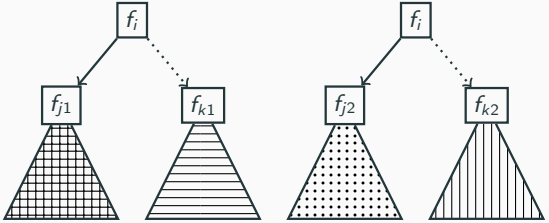


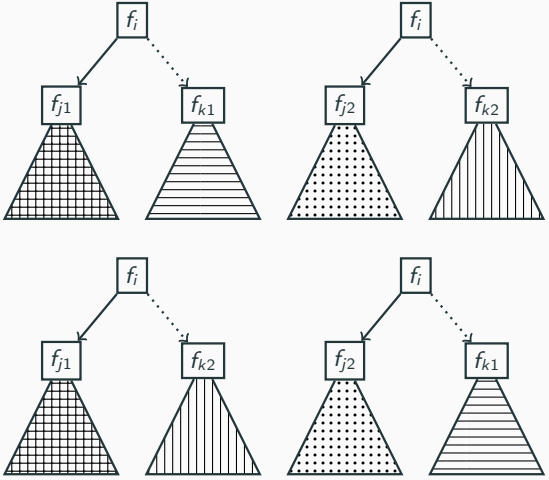
OR nodes

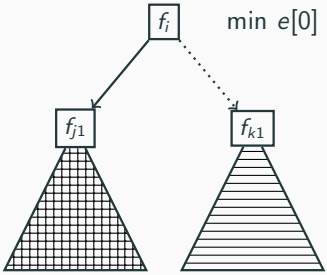
SOL = SOL_1 or SOL_2 or ...

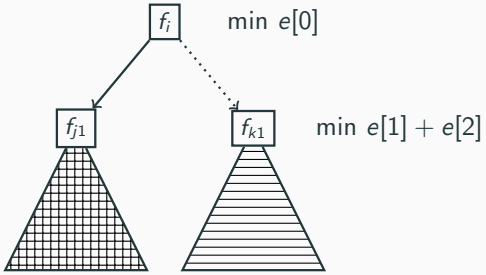
AND nodes

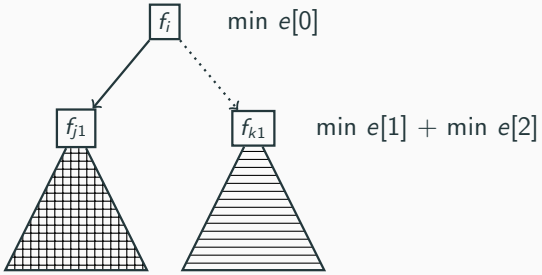
SOL = SOL_1 and SOL_2 and ...

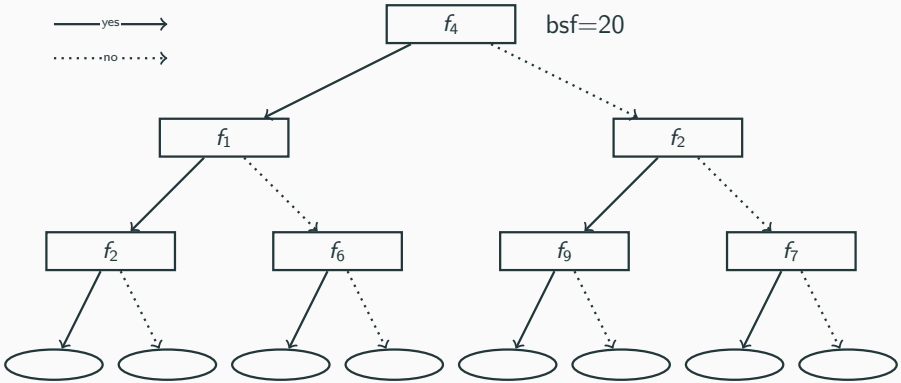

f_i
f_{j1}
f_{k1}
f_i
f_{j2}
f_{k2}

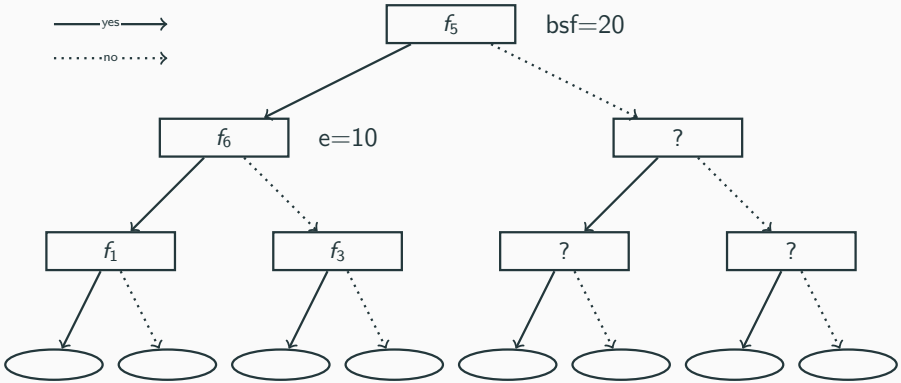

f_i
f_{j1}
f_{k1}
f_i
f_{j2}
f_{k2}
f_i
f_{j1}
f_{k2}
f_i
f_{j2}
f_{k1}

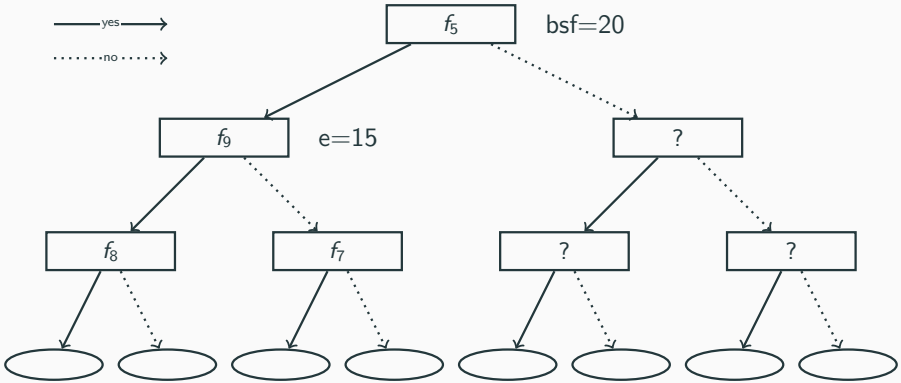

f_i
min $e[0]$
f_{j1}
f_{k1}

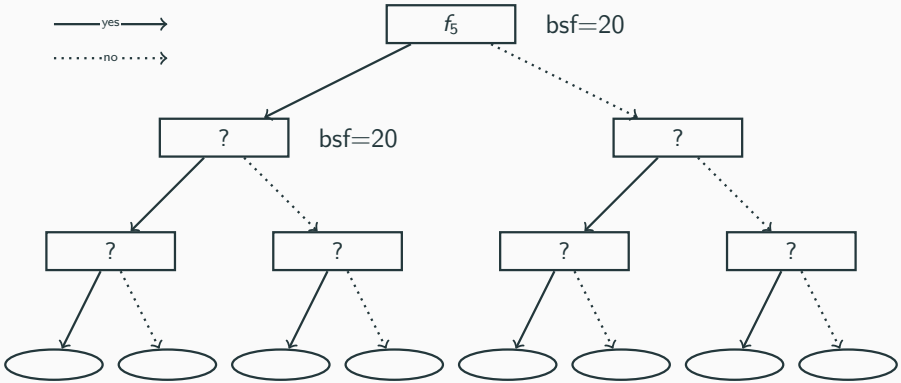

f_i
min $e[0]$
f_{j1}
f_{k1}
min $e[1] + e[2]$

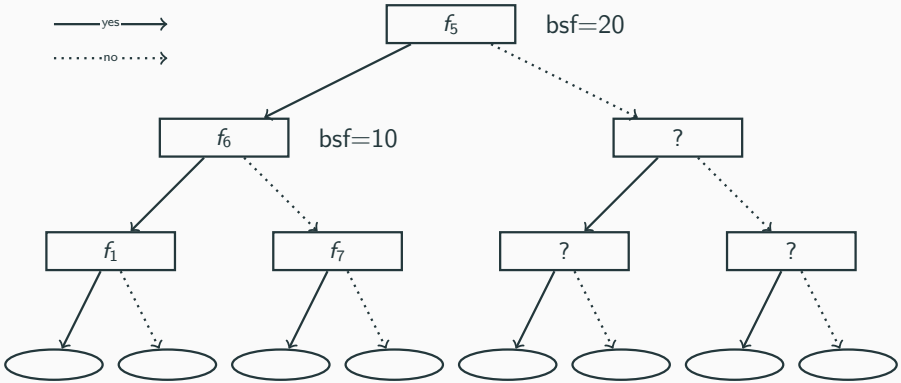

f_i
min $e[0]$
f_{j1}
f_{k1}
min $e[1]$ + min $e[2]$

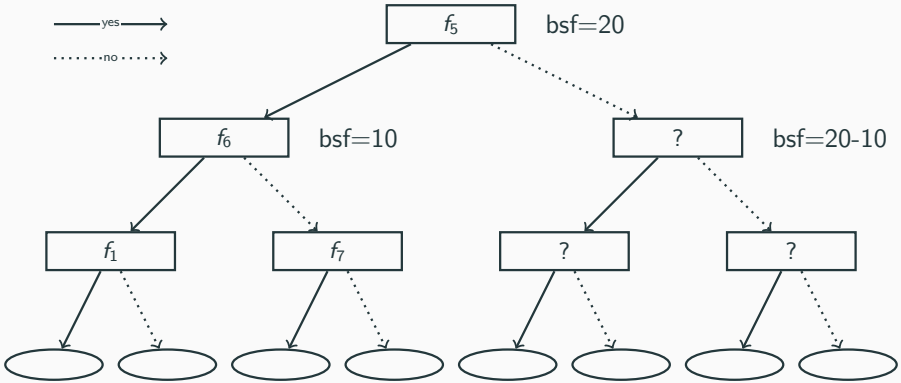

yes
no
f_4
bsf=20
f_1
f_2
f_2
f_6
f_9
f_7

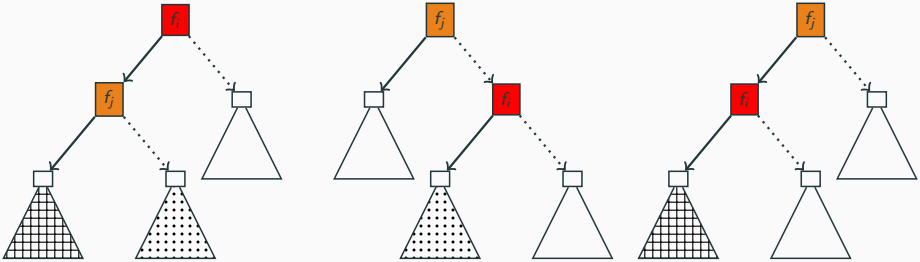

yes
no
f_5
bsf=20
f_6
e=10
?
f_1
f_3
?
?

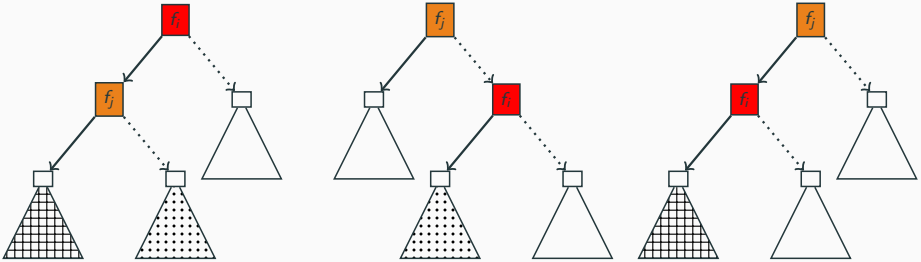

yes
no
f_5
bsf=20
f_9
e=15
?
f_8
f_7
?
?


yes
no
f_5
bsf=20
?
bsf=20
?
?
?
?
?


yes
no
f_5
bsf=20
f_6
bsf=10
?
f_1
f_7
?
?


yes
no
f_5
bsf=20
f_6
bsf=10
?
bsf=20-10
f_1
f_7
?
?


f_i
f_j
f_j
f_i
f_j
f_i



	yes	no	hash
	f_i f_j		$f_i, f_j -$
	f_i	f_j	$f_i - f_j$

Results

UCLouvain

	$N_{min} = 1$			$N_{min} = 5$			
	DL8	BinOCT	CP	DL8	CP	CP-c	CP-m
Proven optimality	49(61%)	13(16%)	**68**(85%)	54(67%)	**65**(81%)	63(79%)	59(74%)
Best solution found	49(61%)	21(26%)	**80**(100%)	54(67%)	**79**(99%)	77(96%)	72(90%)
Fastest	17(21%)	1(1%)	**63**(79%)	26(32%)	**52**(65%)	36(45%)	27(34%)
Time out	31(39%)	67(84%)	**12**(15%)	25(31%)	**15**(19%)	17(21%)	21(26%)

23 instances, depths from 2 to 5, 10 min TO

DL8: Dynamic programming approach using frequent itemsets mining

BinOCT: MIP-based approach running on CPLEX

To summarize

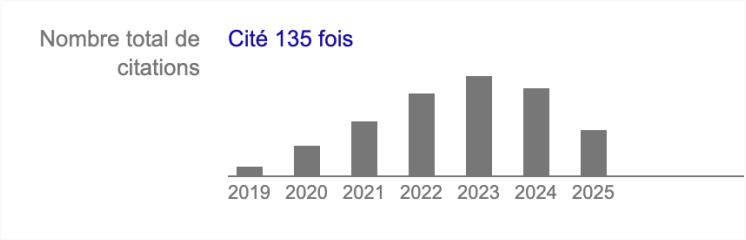
- efficient method
- cp based
- exploits the structure of the problem
- anytime best solution

To go further

- multi-class decision trees
- continuous features through binarization
- other sum-based cost functions
- ...

And since...

And since...


Nombre total de citations
Cité 135 fois
2019
2020
2021
2022
2023
2024
2025

The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Learning Optimal Decision Trees Using Caching Branch-and-Bound Search

Gaël Aglin, Siegfried Nijssen, Pierre Schaus
firstname.lastname@uclouvain.be
ICTEAM, UCLouvain
Louvain-la-Neuve, Belgium

Abstract

Several recent publications have studied the use of Mixed Integer Programming (MIP) for finding an optimal decision tree, that is, the best decision tree under formal requirements on accuracy, fairness or interpretability of the predictive model. These publications used MIP to deal with the hard computational challenge of finding such trees. In this paper, we introduce a new efficient algorithm, DL8.5, for finding optimal decision trees, based on the use of itemset mining

- the trees found are accurate while satisfying additional constraints such as on the *fairness* of the trees: in their predictions, the trees may favor one group of individuals over another.

With the increasing interest in explainable and fair models in machine learning, recent years have witnessed a renewed interest in alternative algorithms for learning decision trees that can provide such optimality guarantees.

Most attention has been given in recent years and in

MaxSat and integration in AdaBoost (IJCAI 2020)

UCLouvain

Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)

Learning Optimal Decision Trees with MaxSAT and its Integration in AdaBoost

Hao Hu[1*], **Mohamed Siala**[1], **Emmanuel Hebrard**[1], **Marie-José Huguet**[1]
LAAS-CNRS, Université de Toulouse, CNRS, INSA, Toulouse, France[1]
{hhu, siala, hebrard, huguet}@laas.fr

Abstract

Recently, several exact methods to compute decision trees have been introduced. On the one hand, these approaches can find optimal trees for vari-

methods might exist [Bennett, 1994]. Moreover, simpler (e.g. smaller) trees, are often more accurate on unknown data.

Recently, several exact methods have been introduced to find optimal decision trees for some combinations of criteria involving, their size (i.e., nodes), depth, and their empirical

Journal of Machine Learning Research 23 (2022) 1-47 Submitted 5/20; Revised 10/21; Published 2/22

MurTree: Optimal Decision Trees via Dynamic Programming and Search

Emir Demirović E.DEMIROVIC@TUDELFT.NL
Anna Lukina A.LUKINA@TUDELFT.NL
Delft University of Technology
Delft, The Netherlands

Emmanuel Hebrard HEBRARD@LAAS.FR
LAAS CNRS
Toulouse, France

Jeffrey Chan JEFFREY.CHAN@RMIT.EDU.AU
RMIT University
Melbourne, Australia

James Bailey BAILEYJ@UNIMELB.EDU.AU
Christopher Leckie CALECKIE@UNIMELB.EDU.AU
Kotagiri Ramamohanarao KOTAGIRI@UNIMELB.EDU.AU
University of Melbourne
Melbourne, Australia

Peter J. Stuckey PETER.STUCKEY@MONASH.EDU
Monash University and DATA61
Melbourne, Australia

DT-Slim with SAT and local encoding (JAIR 2024)

UCLouvain

Journal of Artificial Intelligence Research 80 (2024) 875-918 Submitted 01/2024; published 07/2024

SAT-based Decision Tree Learning for Large Data Sets

André Schidler ASCHIDLER@AC.TUWIEN.AC.AT
Stefan Szeider SZ@AC.TUWIEN.AC.AT
Algorithms & Complexity,
TU Wien,
Favoritenstrasse 9-11,
1040 Vienna, Austria

Abstract

Decision trees of low depth are beneficial for understanding and interpreting the data they represent. Unfortunately, finding a decision tree of lowest complexity (depth or size) that correctly represents given data is NP-hard. Hence known algorithms either (i) utilize heuristics that do not minimize the depth or (ii) are exact but scale only to small or medium-sized instances. We propose a new hybrid approach to decision tree learning, combining heuristic and exact methods in a novel way. More specifically, we employ SAT encodings repeatedly to local parts of a decision tree provided by a standard heuristic,

And many more...

UCLouvain

Decision trees: from efficient prediction to responsible AI
H Blockeel, L Devos, B Frénay, G Nanfack, ... - Frontiers in artificial ..., 2023 - frontiersin.org
This article provides a birds-eye view on the role of decision trees in machine learning and data science over roughly four decades. It sketches the evolution of decision tree research ...
☆ Enregistrer 99 Citer Cité 145 fois Autres articles Les 11 versions »

Machine learning in agriculture: A comprehensive updated review
L Benos, AC Tagarakis, G Dolias, R Berruto, D Kateris, ... - Sensors, 2021 - mdpi.com
The digital transformation of agriculture has evolved various aspects of management into artificial intelligent systems for the sake of making value from the ever-increasing data ...
☆ Enregistrer 99 Citer Cité 1002 fois Autres articles Les 15 versions »

Mathematical optimization in classification and regression trees
E Carrizosa, C Molero-Río, D Romero Morales - Top, 2021 - Springer
Classification and regression trees, as well as their variants, are off-the-shelf methods in Machine Learning. In this paper, we review recent contributions within the Continuous ...
☆ Enregistrer 99 Citer Cité 178 fois Autres articles Les 15 versions

On tackling explanation redundancy in decision trees
Y Izza, A Ignatiev, J Marques-Silva - Journal of Artificial Intelligence ..., 2022 - jair.org
Decision trees (DTs) epitomize the ideal of interpretability of machine learning (ML) models. The interpretability of decision trees motivates explainability approaches by so-called ...
☆ Enregistrer 99 Citer Cité 96 fois Autres articles Les 14 versions »

Logic-based explainability in machine learning
J Marques-Silva - ... Knowledge: 18th International Summer School 2022 ..., 2023 - Springer
The last decade witnessed an ever-increasing stream of successes in Machine Learning (ML). These successes offer clear evidence that ML is bound to become pervasive in a wide ...
☆ Enregistrer 99 Citer Cité 83 fois Autres articles Les 7 versions

Learning optimal decision trees using caching branch-and-bound search
G Aglin, S Nijssen, P Schaus - Proceedings of the AAAI conference on ..., 2020 - ojs.aaai.org
Several recent publications have studied the use of Mixed Integer Programming (MIP) for finding an optimal decision tree, that is, the best decision tree under formal requirements on ...
☆ Enregistrer 99 Citer Cité 190 fois Autres articles Les 7 versions »

On explaining decision trees
Y Izza, A Ignatiev, J Marques-Silva - arXiv preprint arXiv:2010.11034, 2020 - arxiv.org
Decision trees (DTs) epitomize what have become to be known as interpretable machine learning (ML) models. This is informally motivated by paths in DTs being often much smaller ...
☆ Enregistrer 99 Citer Cité 144 fois Autres articles Les 8 versions »

Murtree: Optimal decision trees via dynamic programming and search
E Demirović, A Lukina, E Hebrard, J Chan, ... - Journal of Machine ..., 2022 - jmlr.org
Decision tree learning is a widely used approach in machine learning, favoured in applications that require concise and interpretable models. Heuristic methods are ...
☆ Enregistrer 99 Citer Cité 132 fois Autres articles Les 14 versions »

SAT-based decision tree learning for large data sets
A Schidler, S Szeider - Journal of Artificial Intelligence Research, 2024 - jair.org
Decision trees of low depth are beneficial for understanding and interpreting the data they represent. Unfortunately, finding a decision tree of lowest complexity (depth or size) that ...
☆ Enregistrer 99 Citer Cité 79 fois Autres articles Les 13 versions »

Constraint enforcement on decision trees: A survey
G Nanfack, P Temple, B Frénay - ACM Computing Surveys (CSUR), 2022 - dl.acm.org
Decision trees have the particularity of being machine learning models that are visually easy to interpret and understand. Therefore, they are primarily suited for sensitive domains like ...

And many more...

UCLouvain

Decision trees: from
H Blockeel, L Devos, B Fr
This article provides a bird
data science over roughly
Enregistrer Citer

Strong optimal classification trees
S Aghaei, A Gómez, P Vayanos - arXiv preprint arXiv:2103.15965, 2021 - arxiv.org
Decision trees are among the most popular machine learning models and are used routinely in applications ranging from revenue management and medicine to bioinformatics. In this …
Enregistrer Citer Cité 74 fois Autres articles Les 5 versions

Machine learning in
L Benos, AC Tagarakis, G
The digital transformation
artificial intelligent systems
Enregistrer Citer

Necessary and sufficient conditions for optimal decision trees using dynamic programming
J van der Linden, M de Weerdt… - Advances in Neural …, 2023 - proceedings.neurips.cc
Global optimization of decision trees has shown to be promising in terms of accuracy, size, and consequently human comprehensibility. However, many of the methods used rely on …
Enregistrer Citer Cité 24 fois Autres articles Les 9 versions

Mathematical optimi
E Carrizosa, C Molero-Ríc
Classification and regress
Machine Learning. In this
Enregistrer Citer

Optimal survival trees
D Bertsimas, J Dunn, E Gibson, A Orfanoudaki - Machine learning, 2022 - Springer
Tree-based models are increasingly popular due to their ability to identify complex relationships that are beyond the scope of parametric models. Survival tree methods adapt …
Enregistrer Citer Cité 64 fois Autres articles Les 9 versions

On tackling explana
Y Izza, A Ignatiev, J Marqu
Decision trees (DTs) epito
The interpretability of deci
Enregistrer Citer

[HTML] Solid waste management techniques powered by in-silico approaches with a special focus on municipal solid waste management: Research trends and …
S Vyas, K Dhakar, S Varjani, RR Singhania… - Science of the Total …, 2023 - Elsevier
Many technical, climatic, environmental, biological, financial, educational, and regulatory factors are typically involved in solid waste management (SWM). Artificial Intelligence (AI) …
Enregistrer Citer Cité 21 fois Autres articles Les 7 versions

Logic-based explain
J Marques-Silva - … Know
The last decade witnessed
(ML). These successes of
Enregistrer Citer

Optimal survival trees: a dynamic programming approach
T Huisman, JGM van der Linden… - Proceedings of the AAAI …, 2024 - ojs.aaai.org
Survival analysis studies and predicts the time of death, or other singular unrepeated events, based on historical data, while the true time of death for some instances is unknown …
Enregistrer Citer Cité 15 fois Autres articles Les 5 versions

Learning optimal de
G Aglin, S Nijssen, P Scha
Several recent publication
finding an optimal decision
Enregistrer Citer

Optimal policy trees
M Amram, J Dunn, YD Zhuo - Machine Learning, 2022 - Springer
We propose an approach for learning optimal tree-based prescription policies directly from data, combining methods for counterfactual estimation from the causal inference literature …
Enregistrer Citer Cité 57 fois Autres articles Les 7 versions

On explaining decis
Y Izza, A Ignatiev, J Marqu
Decision trees (DTs) epito
learning (ML) models. This
Enregistrer Citer

[PDF] Learning optimal decision trees with maxsat and its integration in adaboost
H Hu, M Siala, E Hebrard, MJ Huguet - IJCAI-PRICAI 2020, 29th …, 2020 - ijcai.org
Recently, several exact methods to compute decision trees have been introduced. On the one hand, these approaches can find optimal trees for various objective functions including …
Enregistrer Citer Cité 72 fois Autres articles Les 7 versions

Murtree: Optimal de
E Demirović, A Lukina, E H
Decision tree learning is a
applications that require c
Enregistrer Citer

[HTML] Model tree methods for explaining deep reinforcement learning agents in real-time robotic applications
VB Gjærum, I Strümke, J Løver, T Miller, AM Lekkas - Neurocomputing, 2023 - Elsevier
Deep reinforcement learning has shown useful in the field of robotics but the black-box nature of deep neural networks impedes the applicability of deep reinforcement learning …
Enregistrer Citer Cité 30 fois Autres articles Les 3 versions

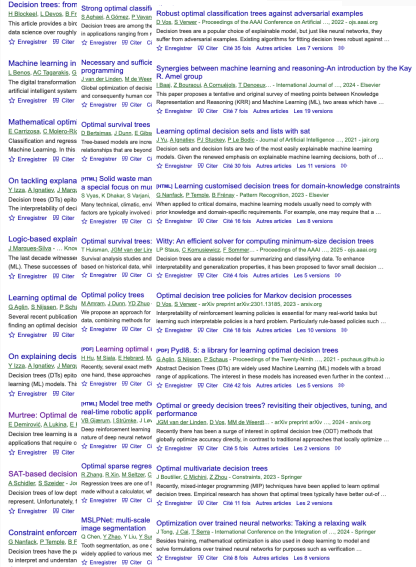
SAT-based decision
A Schidler, S Szeider - Jou
Decision trees of low dept
represent. Unfortunately, f
Enregistrer Citer

Optimal sparse regression trees
R Zhang, R Xin, M Seltzer, C Rudin - … of the AAAI Conference on Artificial …, 2023 - ojs.aaai.org
Regression trees are one of the oldest forms of AI models, and their predictions can be made without a calculator, which makes them broadly useful, particularly for high-stakes …
Enregistrer Citer Cité 21 fois Autres articles Les 11 versions

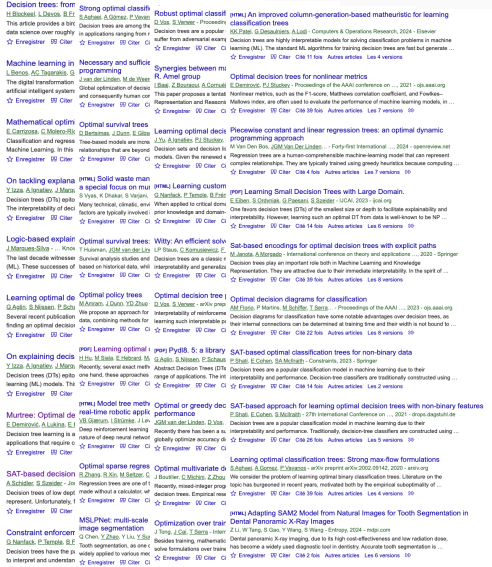
Constraint enforcem
G Nanfack, P Temple, B F
Decision trees have the pa
to interpret and understan

MSLPNet: multi-scale location perception network for dental panoramic X-ray image segmentation
Q Chen, Y Zhao, Y Liu, Y Sun, C Yang, P Li… - Neural Computing and …, 2021 - Springer
Tooth segmentation, as one of the key techniques of medical image segmentation, can be widely applied to various medical applications, eg, orthodontic treatment, corpse …
Enregistrer Citer Cité 43 fois Autres articles Les 5 versions

And many more…

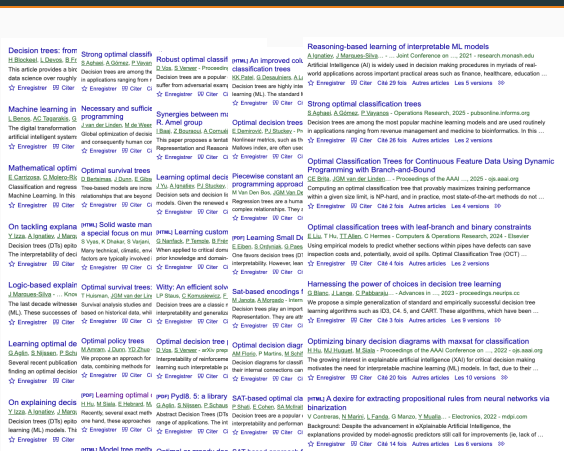


And many more…



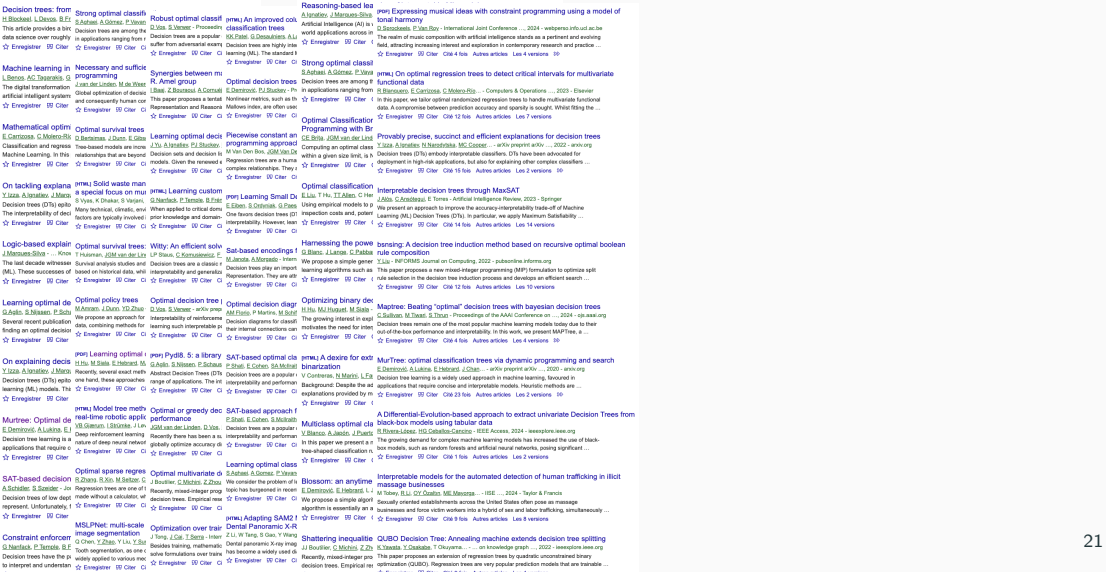
And many more…

UCLouvain



And many more…

UCLouvain



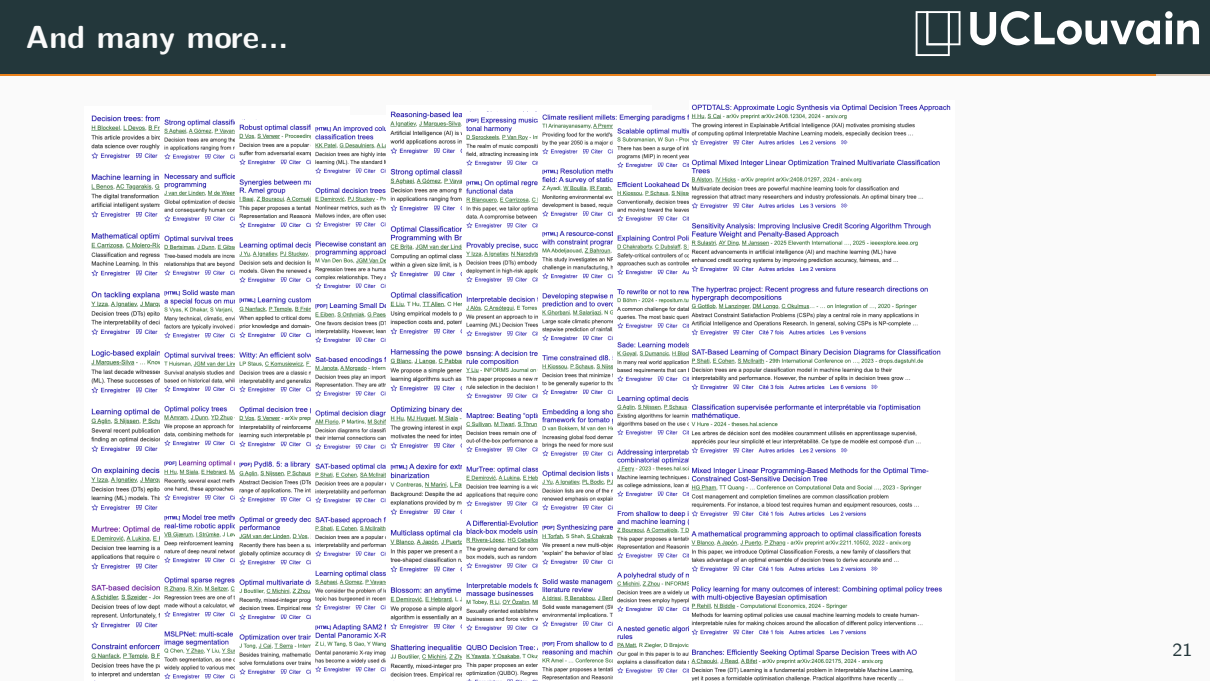
And many more…

UCLouvain

And many more...

UCLouvain

And many more…



And many more...

And many more...

And many more...

UCLouvain

And many more…

UCLouvain

Thank you for listening!

Any questions?

https://hverhaeghe.bitbucket.io/