Optimal decision trees using optimization techniques

MASTER THESIS

Master in Data Science

Josep Alòs Pascual
jalospa@uoc.edu

Universitat Oberta de Catalunya
Course instructor (UOC): Raúl Parada Medina
Area Responsible Professor (UOC): Jordi Casas Roma
External supervisor (UdL): Carlos Ansotegui

July 8, 2019

Outline

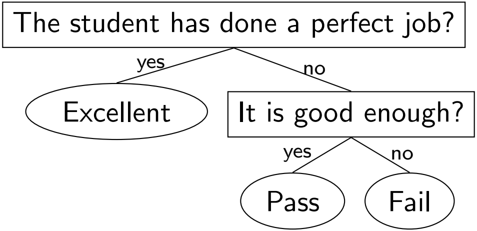
1. Introduction
   - Optimal decision trees

Outline

Outline

Project's introduction

Decision trees. An introduction



Optimal decision trees. The "WHAT"

Optimal decision tree iif:

1. 100% accuracy on train
2. With a smaller size $\implies$ *accuracy* $\neq$ 100%

Proof

We need proof that this specific tree is optimal

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1} \qquad j \in \mathrm{LR}(i)$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1} \qquad \neg v_i \to \left(\sum_{j \in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \rightarrow \left(\sum_{\in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

Optimal decision trees. The “HOW”

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \rightarrow \left(\sum_{j \in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum_{i=\lfloor \frac{j}{2} \rfloor}^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \to \left(\sum_{\in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} \left((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij}) \right) \right); \; d^0_{r,1} = 0.$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \to \left(\sum_{\mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} \left((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij}) \right) \right); \; d^0_{r,1} = 0.$$

$$d^1_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} \left((p_{ji} \wedge d^1_{ri}) \vee (a_{ri} \wedge l_{ij}) \right) \right); \; d^1_{r,1} = 0.$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \to \left(\sum_{\mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} \left((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij}) \right) \right);\ d^0_{r,1} = 0.$$

$$\left(\bigvee^{j-1} \ldots _{ri}) \vee (a_{ri} \wedge l_{ij}) \right) \right);\ d^1_{r,1} = 0.$$

$$\bigwedge_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji} \to \neg a_{rj})$$

$$u_{rj} \leftrightarrow \left(a_{rj} \vee \bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji}) \right)$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \rightarrow \left(\sum_{j \in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} ((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij})) \right); \; d^0_{r,1} = 0.$$

$$\left(\bigvee^{j-1} \ldots {}^{1}_{ri}) \vee (a_{ri} \wedge l_{ij})) \right); \; d^1_{r,1} = 0.$$

$$\bigwedge_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji} \rightarrow \neg a_{rj})$$

$$u_{rj} \leftrightarrow \left(a_{rj} \vee \bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji}) \right)$$

$$\neg v_j \rightarrow \left(\sum_{r=1}^{K} a_{rj} = 1 \right) \qquad \text{with } j = 1, \ldots, N$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \rightarrow \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \rightarrow \left(\sum_{j \in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} ((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij})) \right); \; d^0_{r,1} = 0.$$

$$\left(\bigvee^{j-1} \ldots {}_{ri}) \vee (a_{ri} \wedge l_{ij})) \right); \; d^1_{r,1} = 0.$$

$$\bigwedge_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji} \rightarrow \neg a_{rj})$$

$$u_{rj} \leftrightarrow \left(a_{rj} \vee \bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji}) \right)$$

$$\neg v_j \rightarrow \left(\sum^{K} a_{rj} = 1 \right) \qquad \text{with } j = 1, \ldots, N$$

$$v_j \rightarrow \left(\sum_{r=1}^{K} a_{rj} = 0 \right) \qquad \text{with } j = 1, \ldots, N$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \to \left(\sum_{j \in \mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} ((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij})) \right); \; d^0_{r,1} = 0.$$

$$\ldots_{ri}) \vee (a_{ri} \wedge l_{ij})) \Big); \; d^1_{r,1} = 0.$$

$$\bigwedge_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji} \to \neg a_{rj})$$

$$u_{rj} \leftrightarrow \left(a_{rj} \vee \bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji}) \right)$$

$$\neg v_j \to \left(\sum^{K} a_{rj} = 1 \right) \qquad \text{with } j = 1, \ldots, N$$

$$v_j \to \left(\sum_{r=1}^{K} a_{rj} = 0 \right)$$

$$v_j \wedge \neg c_j \to \bigvee_{r=1}^{K} d_{r,j}^{\sigma(r,q)}$$

Optimal decision trees. The "HOW"

$$(\neg v_1)$$

$$v_i \to \neg l_{ij} \qquad j \in \mathrm{LR}(i)$$

$$l_{ij} \leftrightarrow r_{ij+1}$$

$$\neg v_i \to \left(\sum_{:\mathrm{LR}(i)} l_{ij} = 1 \right)$$

$$p_{ji} \leftrightarrow l_{ij}, \quad j \in \mathrm{LR}(i)$$

$$p_{ji} \leftrightarrow r_{ij}, \quad j \in \mathrm{RR}(i)$$

$$\left(\sum^{\min(j-1,N)} p_{ji} = 1 \right) \qquad \text{with } j = 2, \ldots, N$$

$$d^0_{rj} \leftrightarrow \left(\bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} \left((p_{ji} \wedge d^0_{ri}) \vee (a_{ri} \wedge r_{ij}) \right) \right); \; d^0_{r,1} = 0.$$

$$\ldots_{ri}) \vee (a_{ri} \wedge l_{ij})) \Big); \; d^1_{r,1} = 0.$$

$$\bigwedge_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji} \to \neg a_{rj})$$

$$u_{rj} \leftrightarrow \left(a_{rj} \vee \bigvee_{i=\lfloor \frac{j}{2} \rfloor}^{j-1} (u_{ri} \wedge p_{ji}) \right)$$

$$\neg v_j \to \left(\sum^{K} a_{rj} = 1 \right) \qquad \text{with } j = 1, \ldots, N$$

$$v_j \to \left(\sum_{r=1}^{K} a_{rj} = 0 \right)$$

$$v_j \wedge \neg c_j \to \bigvee_{r=1}^{K} d^{\sigma(r,q)}_{r,j}$$

$$v_j \wedge c_j \to \bigvee_{r=1}^{K} d^{\sigma(r,q)}_{r,j}$$

Optimal decision trees. The “Problem”?

Problem

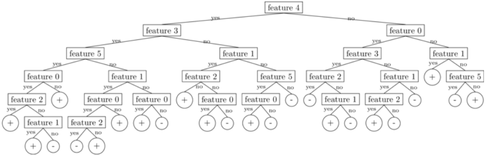
Obtaining optimal decision trees is hard

Is it really a problem, though?

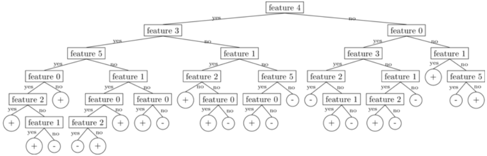
Why would an optimal decision tree interest me?

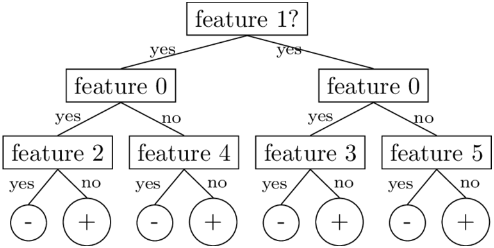
Sklearn is good enough, right?

Optimal decision trees. The “WHY”

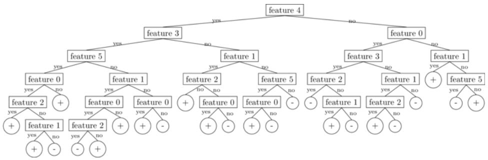


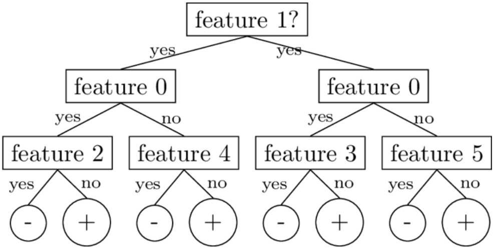
Optimal decision trees. The "WHY"

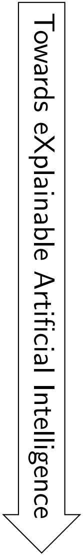




Optimal decision trees. The “WHY”







The problem

We have tools for greedy algorithms

The problem

We have tools for greedy algorithms

We don't have tools for optimal trees

Our solution

Software package that makes optimal algorithms user friendly

1. Prepare the data

Our solution

Software package that makes optimal algorithms user friendly

1. Prepare the data
2. Find optimal trees

Our solution

Software package that makes optimal algorithms user friendly

1. Prepare the data
2. Find optimal trees
3. Abstract different CP solvers and algorithms into a common API

Our solution

Software package that makes optimal algorithms user friendly

1. Prepare the data
2. Find optimal trees
3. Abstract different CP solvers and algorithms into a common API
4. Return the solutions as a ready-to-use class

Our solution

Convert this...

Algorithm 1: Searching for the optimal decision tree

input: A dataset *data* as a matrix of $n\_feats + 1 \times n\_rows$

search_space ← SOMETHING ; `//specific of each encoding`

while `more_params_available`(*search_space*) **do**

  params ← `next_params`(*search_space*);

  model ← `encode`(*data*, *params*);

  solution ← `solve`(*model*);

  if *solution* is *optimal* **then**

    return *solution*

  end

end

return *"No solution"*

Our solution

...into this:

```
model = SomeDTModel(...)
data = np.array(...)

tree = model.find_tree(data)
```

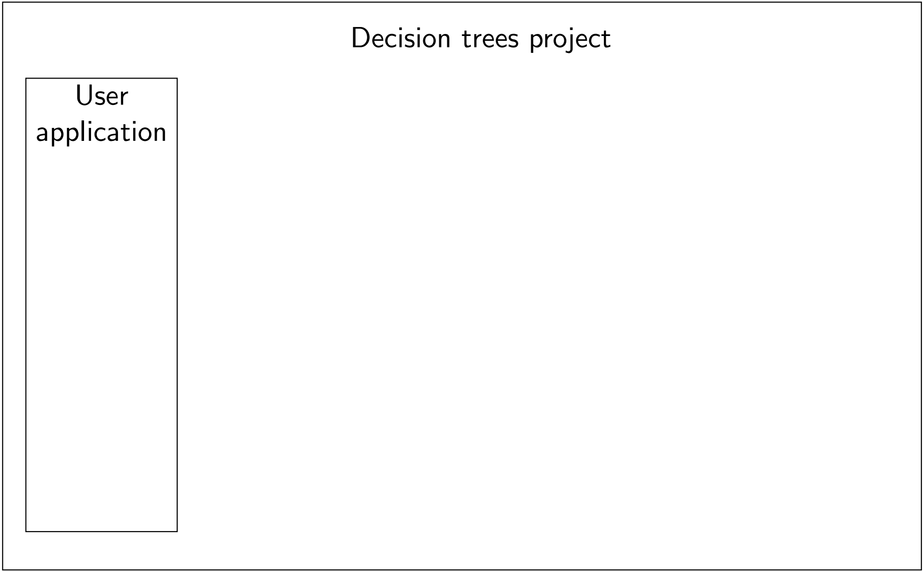
`tree` **is proven to be optimal!**

Our solution - Implemented models

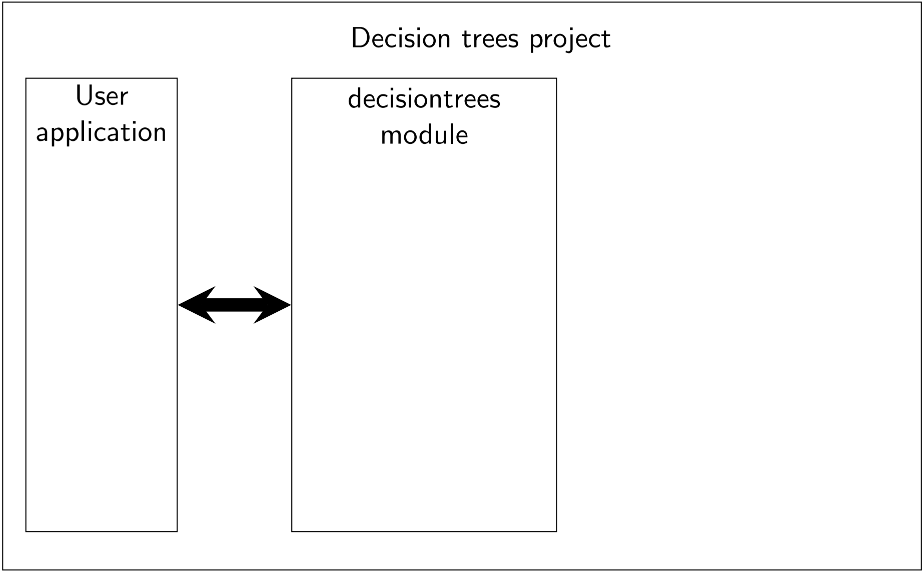
Implemented models:

- Optimize depth: Avellaneda, 2020
- Optimize size (using all observations)
  - 2x Narodytska et al, 2018
  - 1x Avellaneda, 2020
- Optimize size (using a subset of observations): Avellaneda, 2020
- Bonus, non-optimal Scikit-Learn DecisionTreeClassifier

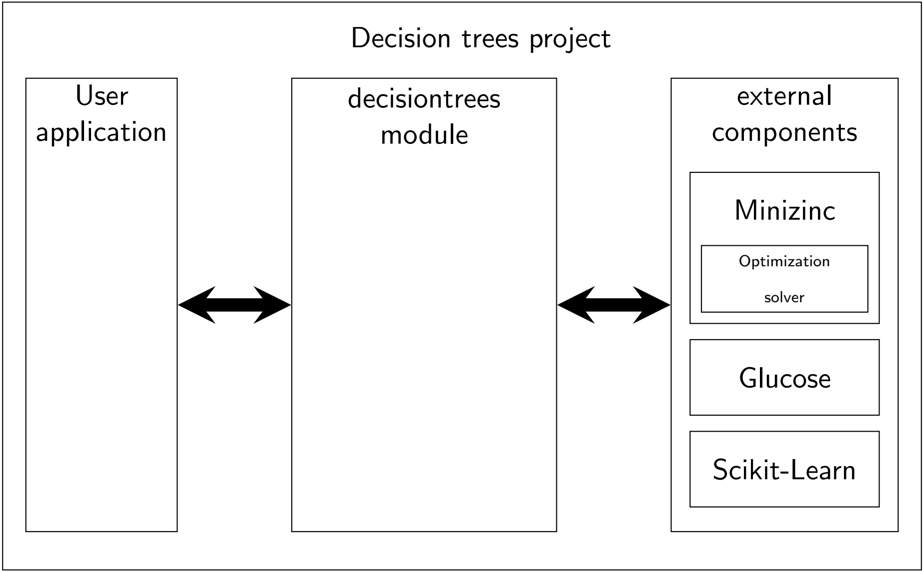
Our solution - Overview



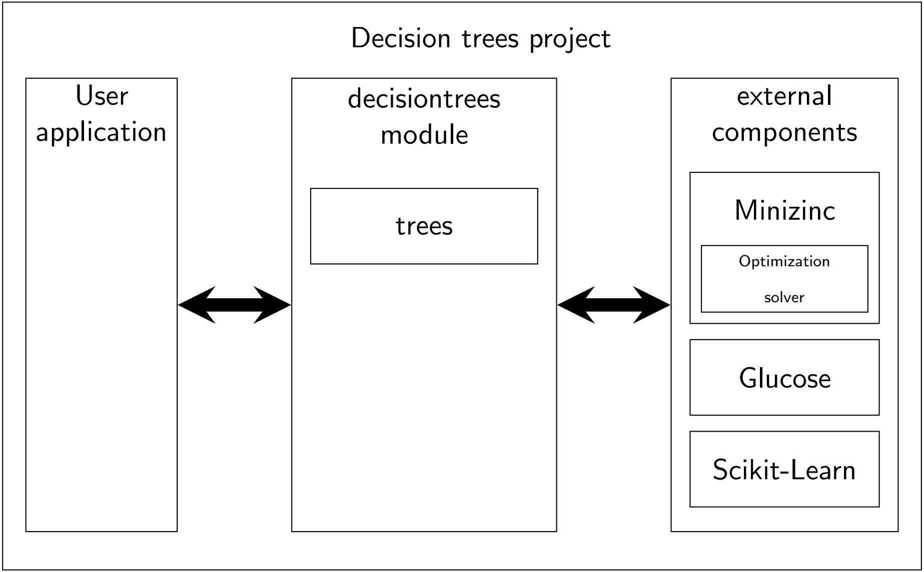
Our solution - Overview



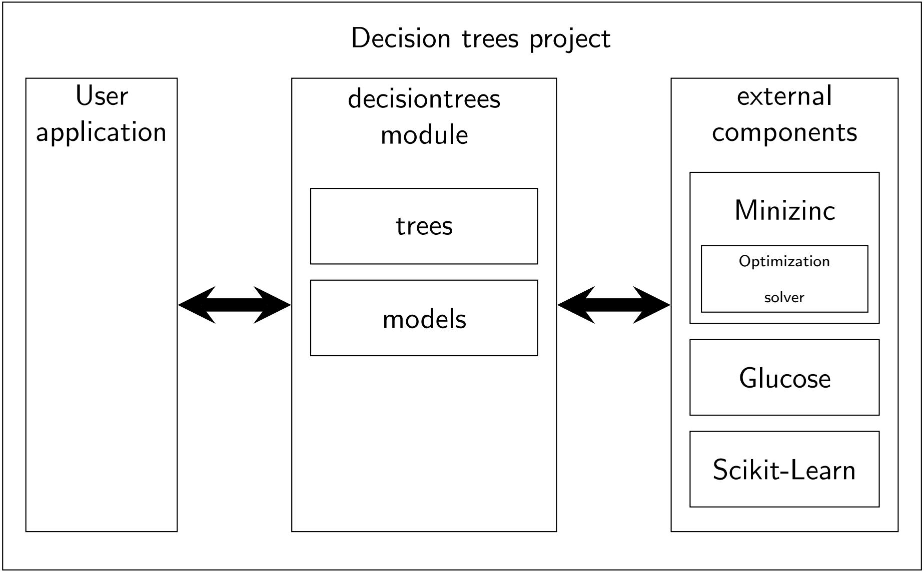
Our solution - Overview



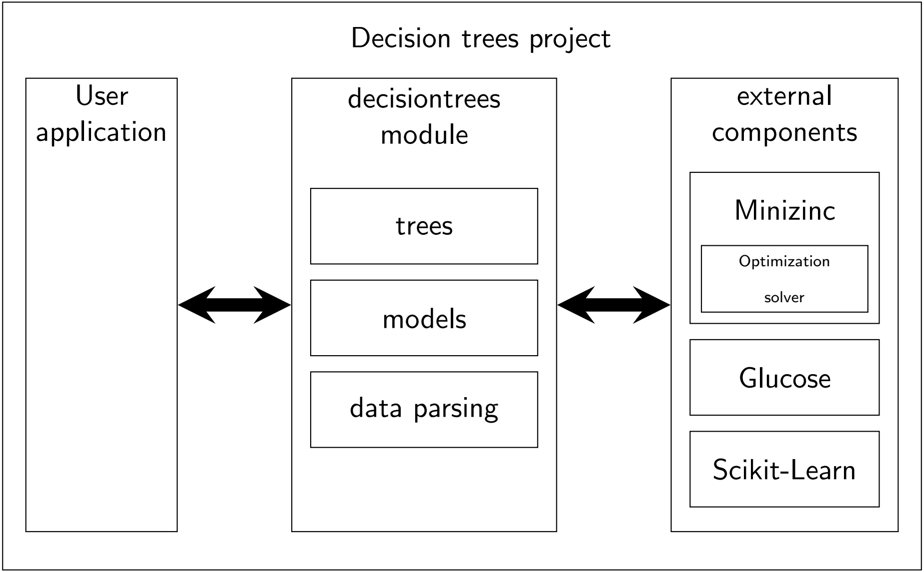
Our solution - Overview



Our solution - Overview



Our solution - Overview



Conclusions - Experimental results

- Mathematical optimization approaches: $\uparrow$ smaller trees; $\downarrow$ slow

Conclusions - Experimental results

- Mathematical optimization approaches: $\uparrow$ smaller trees; $\downarrow$ slow
- Greedy approaches: $\uparrow$ fast; $\downarrow$ don't find (nor proof) optimal trees

Conclusions - Experimental results

- Mathematical optimization approaches: $\uparrow$ smaller trees; $\downarrow$ slow
- Greedy approaches: $\uparrow$ fast; $\downarrow$ don't find (nor proof) optimal trees
- Slowness $\leftarrow$ lack of interfaces between **Minizinc** and SAT solvers

Conclusions - Experimental results

- Mathematical optimization approaches: $\uparrow$ smaller trees; $\downarrow$ slow
- Greedy approaches: $\uparrow$ fast; $\downarrow$ don't find (nor proof) optimal trees
- Slowness $\leftarrow$ lack of interfaces between **Minizinc** and SAT solvers **POTENTIAL FUTURE WORK!**

Future work

- Interface Minizinc with SAT solvers

Future work

- Interface Minizinc with SAT solvers
- Add more approaches!!!

Future work

- Interface Minizinc with SAT solvers
- Add more approaches!!!
- (Neverending future work) Improve source code!

Conclusions

- Towards XAI: Optimal decision trees are smaller
- Created a successful tool to find optimal decision trees
- Breaking entry barriers for non-technical users

Thanks for watching