Open Universiteit

Embedding ML models into Linear Optimization

Case: Project Scheduling of a multi-skilled workforce

dr. Murat Firat, Computer Science - OU

Joint work with
Daan van Wely, MSc - JADS
Bob Gijstra, Building Blocks

12-01-2023

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

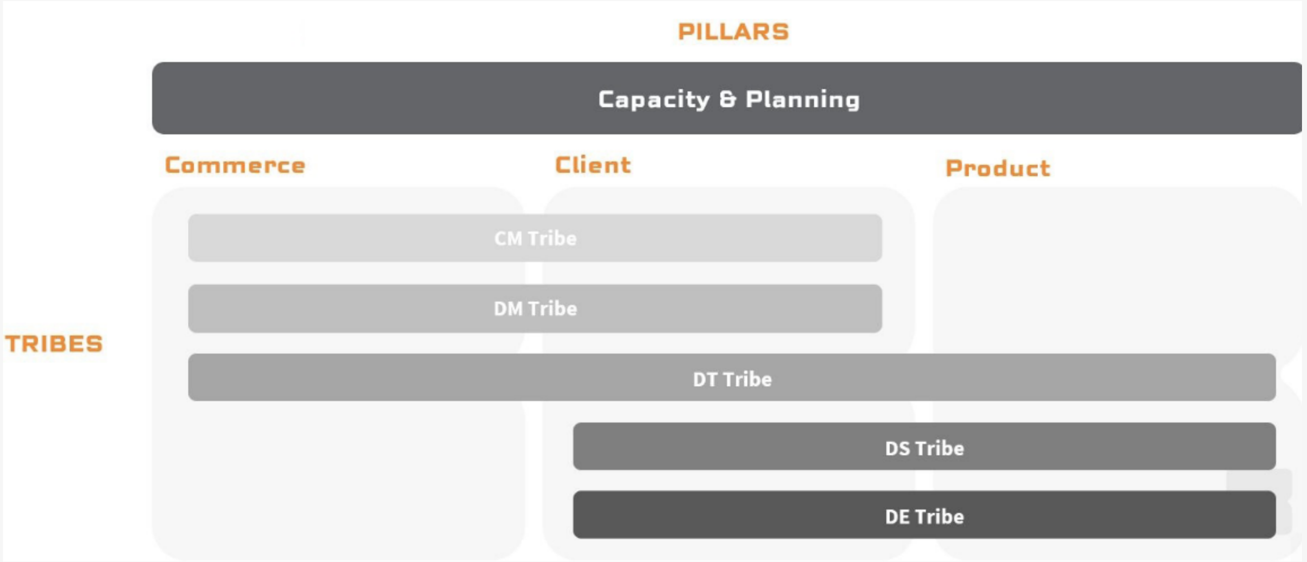
Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Company background

Business and organization

- ICT and DS Projects (mainly) with online retailer companies
- Employee position is characterized by an affiliation with pillar and tribes.



Company background

Client solution lifecycle

- Prerequisite preparations
- Impact of the project: KPIs
- Required skills and workload
- Uniform effort over time

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Problem description

Basic properties

- Projects:

  Release dates, deadlines
  Required skills, estimated workload per employee.
  No simultaneous skill use

Problem description

Basic properties

- Projects:

    Release dates, deadlines
    Required skills, workload per employee.
    No simultaneous skill use

- Employees:

    Possessed skills, availability, contract-defined capacity.
    Pairwise matches, project (topic) preferences.

Problem description

Basic properties

- Projects:

  Release dates, deadlines
  Required skills, workload per employee.
  No simultaneous skill use

- Employees:

  Possessed skills, availability, contract-based capacity.
  Pairwise matches, project (topic) preferences.

- Project teams:

  Working in groups
  Average pairwise employee match
  Average project preferences of employees

Problem description

Project efficiency

- Project teams (Planning Phase):

  Working in groups

  Average pairwise employee match

  Average project preferences of employees

- Project Execution (Available after completion):

  Cooperation of team members, progress in milestones

  Completion time, output quality

  Communication to customer

  Customer satisfaction

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Literature review

Project scheduling

1- Resource Constrained Project Scheduling Problem: schedule activities subject to precedence and resource constraints.

2- Scheduling and staffing multiple projects with a multi-skilled workforce: Worker efficiencies are considered. MILP formulation is proposed as solution approach.

3- *A MILP model for an integrated project scheduling and multi-skilled workforce allocation with flexible working hours:* Worker efficiencies, task workloads, use of single skill.

[1] Brucker et al, "*Resource-constrained project scheduling: Notation, classification, models, and methods*", 1999, EJOR, 112(1).
[2] Heimerl and Kolisch, "*Scheduling and staffing multiple projects with a multi-skilled workforce*", 2010, OR Spectrum, 32.
[3] Karam et al., "A MILP model for an integrated project scheduling and multi-skilled workforce allocation with flexible working hours", 2017, IFAC-PapersOnLine, 50.

Literature review

Project efficiency

1- Measuring the efficiency of project control using fictitious and empirical project data: The efficiency of controlling a project is measured and evaluated using a Monte-Carlo simulation.

2- Support Vector Machine Regression for project control forecasting: Predicting time and cost of a project execution.

[1] Vanhoucke, M., "*Measuring the efficiency of project control using fictitious and empirical project data*", 2012, IJPM, 30(2).
[2] Wauters and Vanhuocke, "*Support Vector Machine Regression for project control forecasting*", 2014, Automation in Construction, 47.

Literature review

Project efficiency

The Relationship between Project Success and Project Efficiency:

Table 1: The five dimensions of project success after Shenhar and Dvir (2007)

Success dimension	*Measures*	*Time*
1. Project efficiency	Meeting schedule goal Meeting budget goal	End of project
2. Team satisfaction	Team morale Skill development Team member growth Team member retention	End of project
3. Impact on the customer	Meeting functional performance Meeting technical specifications Fulfilling customer needs Solving a customer's problem The customer is using the product Customer satisfaction	Months following project
4. Business success	Commercial success Creating a large market share	Years following project
5. Preparing for the future	Creating a new market Creating a new product line Developing a new technology	Years following project

Sarrador and Turner, "*The Relationship between Project Success and Project Efficiency*", 2014, Procedia, 119.

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Methodology

Formulating as Mixed Integer Linear Programming (MILP) model

Decision Variables

x_{ept}	Schedules employee e to work on project p at time t	Employee-Project working time scheduling
y_{eps}	Allocates employee e to work on project p as skill s	Employee-Project skill assignment
τ_p	Tardiness of project p measured in workdays	
κ_{ew}	Number of idle workdays of employee e in week w	

Methodology

Formulating as Mixed Integer Linear Programming (MILP) model

Objective function:

$$\text{Min} \sum_{p \in P} \alpha w_p \tau_p + \frac{1-\alpha}{|E||W|} \sum_{e \in E} \frac{1}{c_e} \sum_{w \in W} \kappa_{ew} \tag{3.1}$$

Minimize project tardiness and idle times of employees

s.t.

Workforce allocation constraints

$$\sum_{e \in E} Sk_{es} y_{eps} \geq Rq_{ps} \qquad \forall p \in P; \forall s \in S \tag{3.2}$$

Skill requirements of the projects

$$y_{eps} \leq Sk_{es} \qquad \forall e \in E; \forall p \in P; \forall s \in S \tag{3.3}$$

Assigning only skilled workers to projects

$$\sum_{s \in S} y_{eps} \leq 1 \qquad \forall e \in E; \forall p \in P \tag{3.4}$$

Only single skill use of employees

$$y_{eps} \geq 1 \qquad \forall p \in P, \forall s \in S; \forall e \in E^{p,s} \tag{3.5}$$

Pre-selected employees for projects

$$\sum_{s \in S} y_{ep^*s} \leq \sum_{s \in S} y_{eps} \qquad \forall e \in E; \forall p \in P \tag{3.6}$$

Employees of preliminary projects also work for final projects

Methodology

Formulating as Mixed Integer Linear Programming (MILP) model

Workload scheduling constraints

$$\sum_{t=r_p}^{T} x_{ept} \geq \rho_p \sum_{s \in S} y_{eps} \qquad \forall e \in E; \forall p \in P \qquad (3.7)$$

Lower bound of workload for assigned employees

$$\sum_{t=r_p}^{T} x_{ept} \leq \rho'_p \sum_{s \in S} y_{eps} \qquad \forall e \in E; \forall p \in P \qquad (3.8)$$

Upper bound of workload for assigned employees

$$\sum_{t=0}^{r_p - 1} x_{ept} \leq 0 \qquad \forall e \in E; \forall p \in P \qquad (3.9)$$

Projects start after (incl.) release date

$$\frac{t}{|T|} x_{ept} - \frac{d_p}{|T|} \leq \tau_p \qquad \forall e \in E, \forall p \in P, \forall t : d_p \leq t \leq T \qquad (3.10)$$

Tardiness of projects

Methodology

Formulating as Mixed Integer Linear Programming (MILP) model

Scheduling conditions constraints

$$\sum_{p \in P} x_{ept} \leq A_{et} \qquad \forall e \in E; \forall t \in T \qquad (3.11)$$

Availability of employees

$$\sum_{e \in E} x_{ept} \leq Rq_p \sigma_{pt} \qquad \forall p \in P; \forall t \in T \qquad (3.12)$$

Project works on allowed days of the week

$$\sum_{p \in P} \sum_{t \in T^w} x_{ept} + \kappa_{ew} = c_e \qquad \forall e \in E; \forall w \in W \qquad (3.13)$$

Capacity of employees

Bounds of decision variables

$$x_{e,p,t} \in \{0,1\} \qquad \forall e \in E; \forall p \in P, \forall\, t \in T \qquad (3.14)$$

$$y_{e,p,s} \in \{0,1\} \qquad \forall e \in E; \forall p \in P, \forall\, s \in S \qquad (3.15)$$

$$\tau_p \geq 0 \qquad \forall p \in P \qquad (3.16)$$

$$0 \leq \kappa_{e,w} \leq c_e \qquad \forall e \in E; \forall w \in W \qquad (3.17)$$

Methodology

Project efficiency

- Planning Phase:

  Working in groups

  Average pairwise employee match

  Average project preferences of employees

- Project Execution:

  Cooperation of team members, progress in milestones

  Completion time, output quality

  Communication to customer

  Customer satisfaction

Methodology

Project efficiency

- Planning Phase:

  Working in groups
  Average pairwise employee match
  Average project preferences of employees

- Project Execution:

  Cooperation of team members,
  Progress in milestones
  Completion time, output quality
  Communication to customer
  Customer satisfaction

Project efficiency!

Methodology

Project efficiency

- Planning Phase:

  Working in groups
  Average pairwise employee match
  Average project preferences of employees

- Project Execution:

  Cooperation of team members,
  Progress in milestones
  Completion time, output quality
  Communication to customer
  Customer satisfaction

Project efficiency!

Question: How to form project teams to maximize predicted project efficiency?

Methodology

Project efficiency

Question: How to form project teams to maximize predicted project efficiency?

Supervised learning:

Data Features: Features of the planning phase and project execution

Data Label: Project efficiency, a qualitative measure.

Predictive model: Construct a decision tree* with only planning phase features

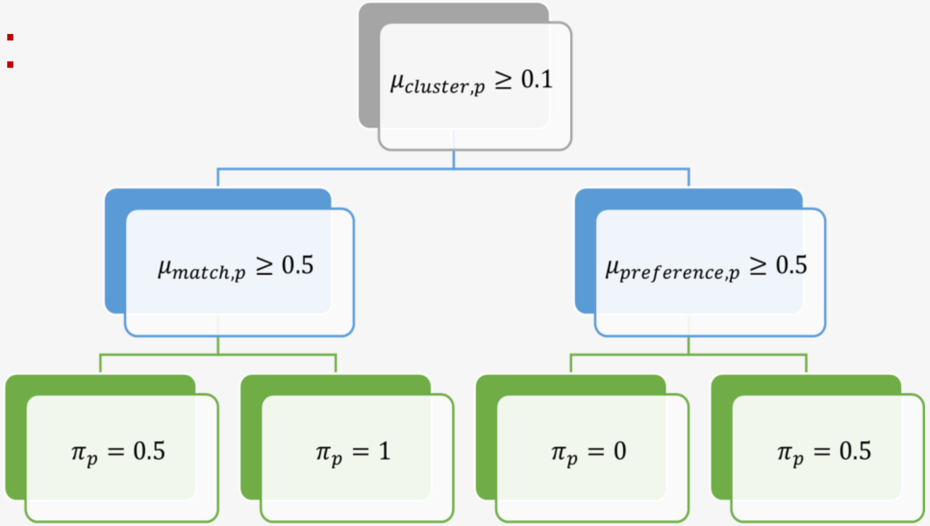
* Binary type with univariate splits at internal nodes.

Methodology

Project efficiency

Question: How to form project teams to maximize predicted project efficiency?

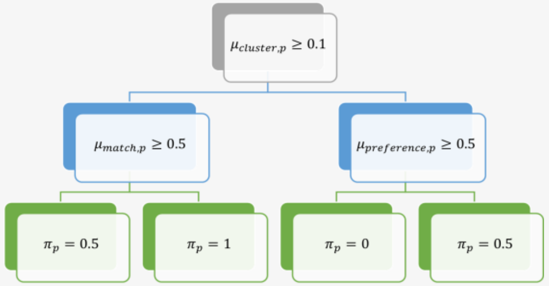
Supervised learning:



A hypothetical decision tree

Methodology

MILP model with embedded decision tree



Extended objective function:

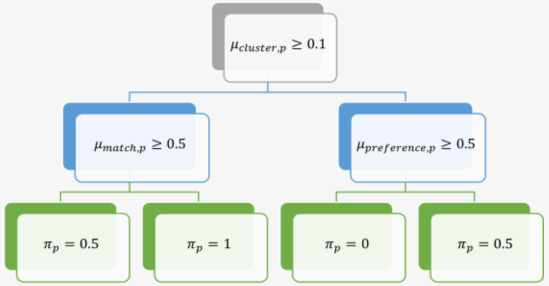
$$\text{Min} \sum_{p \in P} (\alpha w_p \tau_p - \beta \xi_p \pi_p) + \frac{1-\alpha}{|E||W|} \sum_{e \in E} \frac{1}{c_e} \sum_{w \in W} \kappa_{ew} \tag{4.1}$$

s.t.

Minimize tardiness, idle times and maximize (predicted) project efficiency

Methodology

MILP model with embedded decision tree



Clustering feature constraints

$$\sum_{e \in E} x_{ept} = \sum_{i=0}^{Rq_p} i\gamma_{pt}^{i} \qquad \forall p \in P; \forall t \in T \qquad (4.2)$$

$$\sum_{i=0}^{Rq_p} \gamma_{pt}^{i} \leq 1 \qquad \forall p \in P; \forall t \in T \qquad (4.3)$$

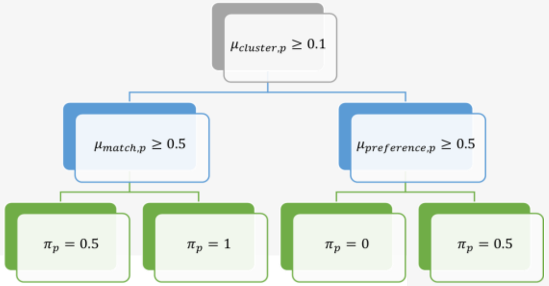
Detecting times of working groups

$$\mu_{cluster,p} = \frac{1}{Rq_p \rho_p} \left(\sum_{t \in T} \gamma_{pt}^{1} \right) \qquad \forall p \in P \qquad (4.4)$$

Days when single employees work on the projects

Methodology

MILP model with embedded decision tree



Match and preference feature constraints

$$\sum_{s \in S} (y_{eps} + y_{e'ps}) \leq 1 + \lambda_{ee'p} \qquad \forall\, e, e' \subseteq E \times E; \forall\, p \in P \qquad (4.5)$$

$$2\lambda_{ee'p} \leq \sum_{s \in S} (y_{eps} + y_{e'ps}) \qquad \forall\, e, e' \subseteq E \times E; \forall\, p \in P \qquad (4.6)$$

$$\mu_{match,p} = \frac{2}{Rq_p(Rq_p - 1)} \sum_{(e,e') \in E} \zeta_{ee'} \lambda_{ee'p} \qquad \forall\, p \in P \setminus P' \qquad (4.7)$$

Detecting pairwise matches of employees of a project team

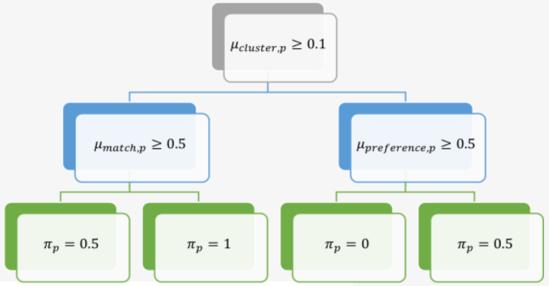
$$\mu_{match,p} = 0 \qquad \forall\, p \in P' \qquad (4.8)$$

$$\mu_{preference,p} = \frac{1}{Rq_p} \sum_{e \in E} \sum_{s \in S} \omega_{ep} y_{eps} \qquad \forall\, p \in P \qquad (4.9)$$

Detecting employee preferences to assigned project topics

Methodology

MILP model with embedded decision tree



Decision tree execution constraints

$$\theta_{p,n_{root}} = 1 \qquad \forall p \in P, \qquad (4.10)$$

Start at the root node

$$\theta_{p,n_l} + \theta_{p,n_r} = \theta_{p,n} \qquad \forall n \in N_{int}; \forall p \in P \qquad (4.11)$$

At current internal node, branch one of the child nodes

$$\frac{1}{\mu_{f,p}^{max}}(\mu_{f,p} - \psi_{f,n}) - (1 - \theta_{p,n}) \leq \theta_{p,n_r} \qquad \forall n \in N_{int}; \forall p \in P \qquad (4.12)$$

If feature value greater than or equal to threshold, branch right

$$\frac{1}{\mu_{f,p}^{max}}(\psi_{f,n} - \mu_{f,p}) - (1 - \theta_{p,n}) \leq \theta_{p,n_l} \qquad \forall n \in N_{int}; \forall p \in P \qquad (4.13)$$

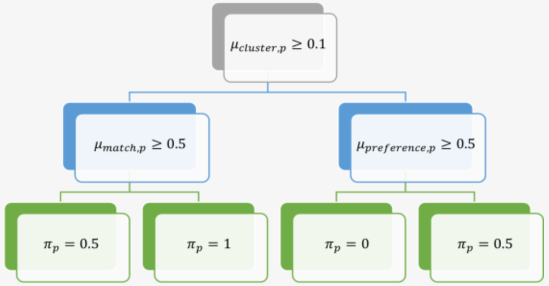
If feature value smaller than the threshold, branch left

$$\pi_p = \sum_{l \in N_{leaf}} \epsilon_l \theta_{p,l} \qquad \forall p \in P \qquad (4.14)$$

Find predicted project efficiency value

Methodology

MILP model with embedded decision tree



Bounds of decision variables

$\theta_{p,n} \in \{0,1\}$	$\forall p \in P, \forall\, n \in N_{int} \cup N_{leaf}$	(4.15)
$\mu_{f,p} \geq 0$	$\forall p \in P, \forall\, f \in F$	(4.16)
$\lambda_{e,e',p} \in \{0,1\}$	$\forall\,(e,e') \in E \times E; \forall p \in P$	(4.17)
$\pi_p \geq 0$	$\forall p \in P$	(4.18)

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Experimentation

- Instances:

Total 86 projects with max 10 skills types required

Number of employees 26.

Schedule horizon varies from 10 workdays to 40 workdays

Model variations

Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

Experimentation

- Computation times:

Model variations

Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

Experimentation

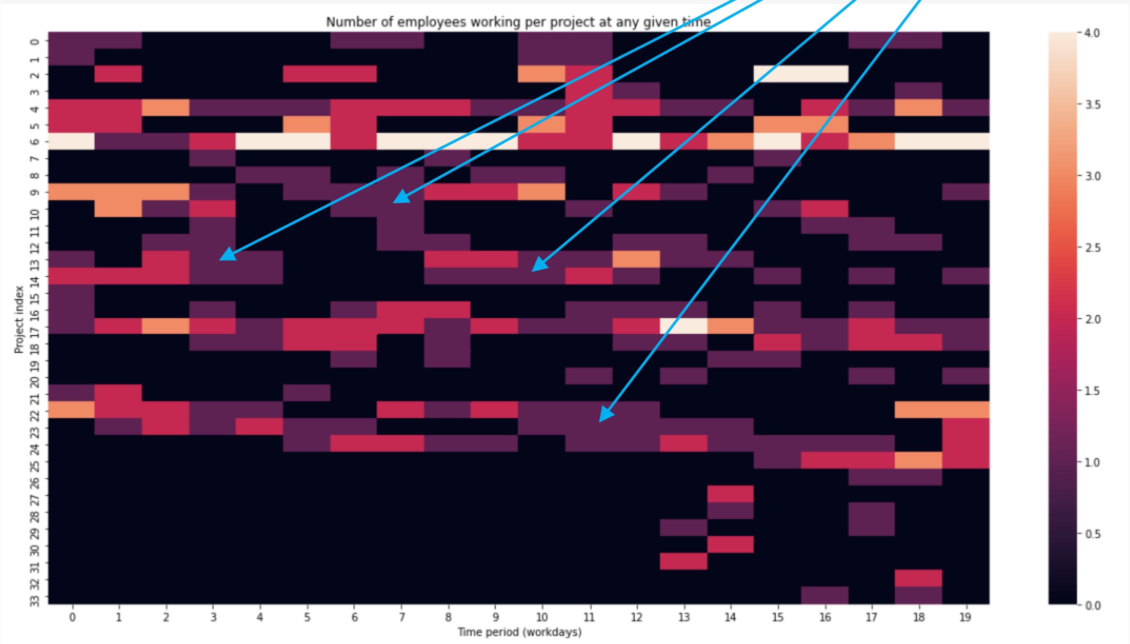
- Computation times:

Experimentation

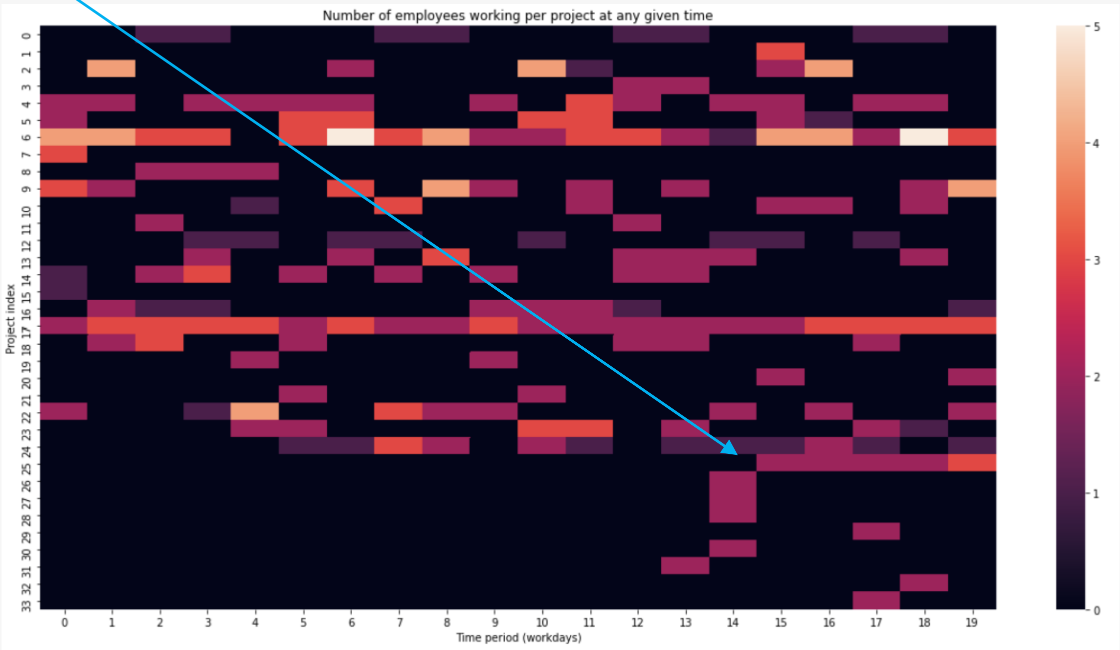
Model variations			
Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

Working in groups

Times when an employee works single



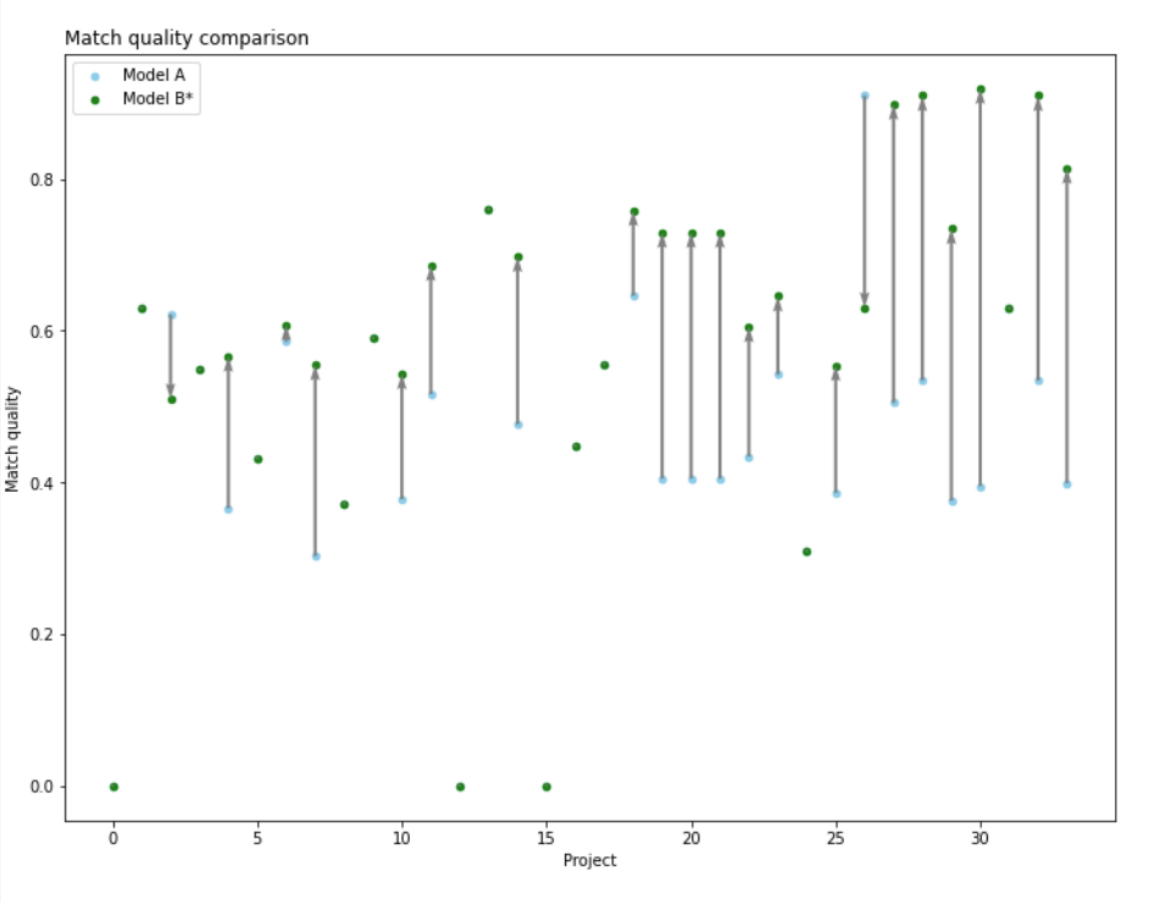
Model A



Model B*

Experimentation

Pairwise matches



Model variations			
Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

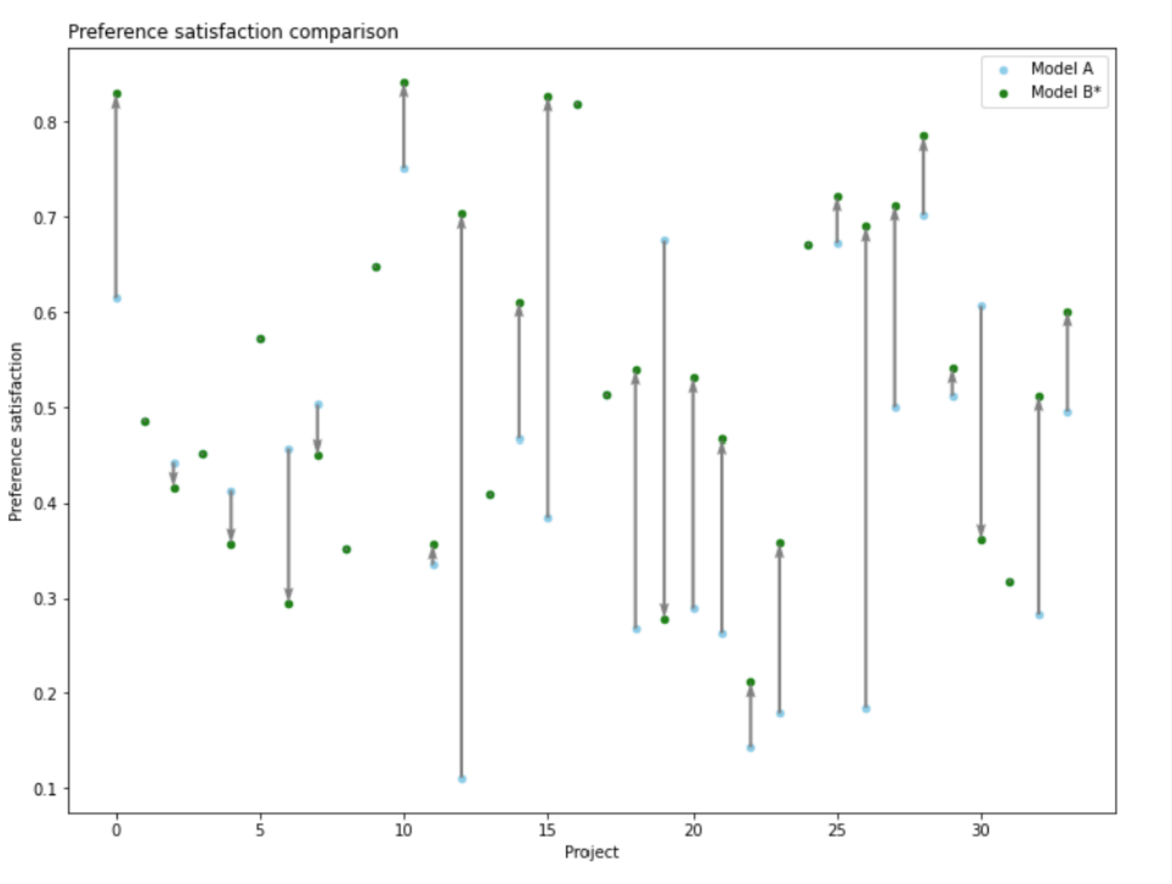
↑ 19

↓ 2

= 13

Experimentation

Employee preferences



Model variations

Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

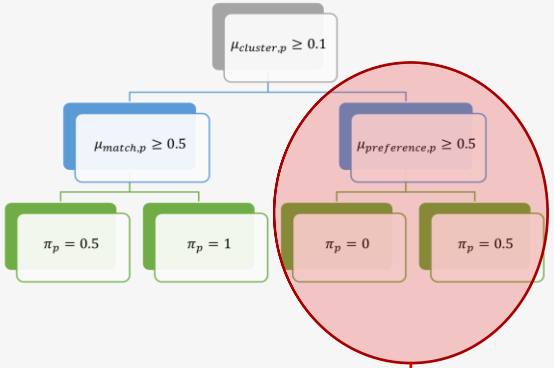
↑ 17

↓ 7

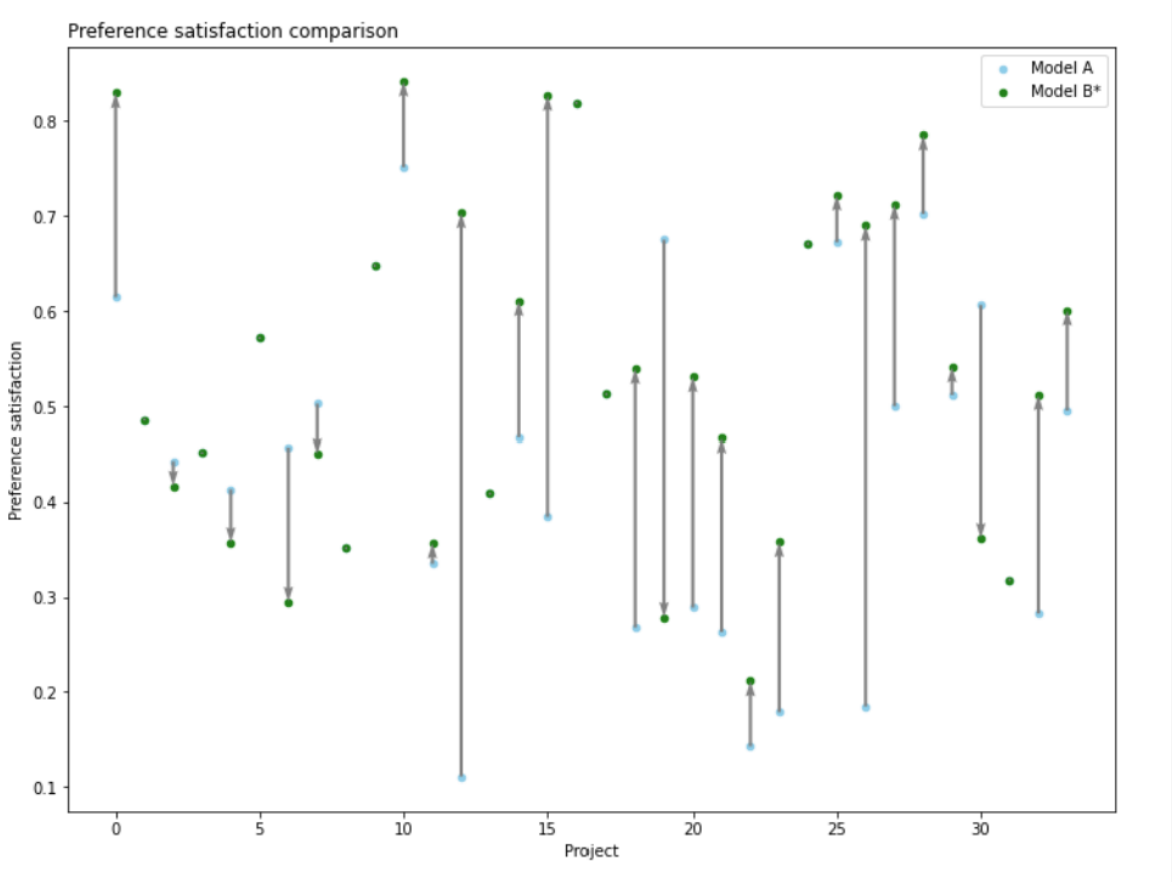
= 10

Experimentation

Employee preferences



Least preferred branching!



Model variations

Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

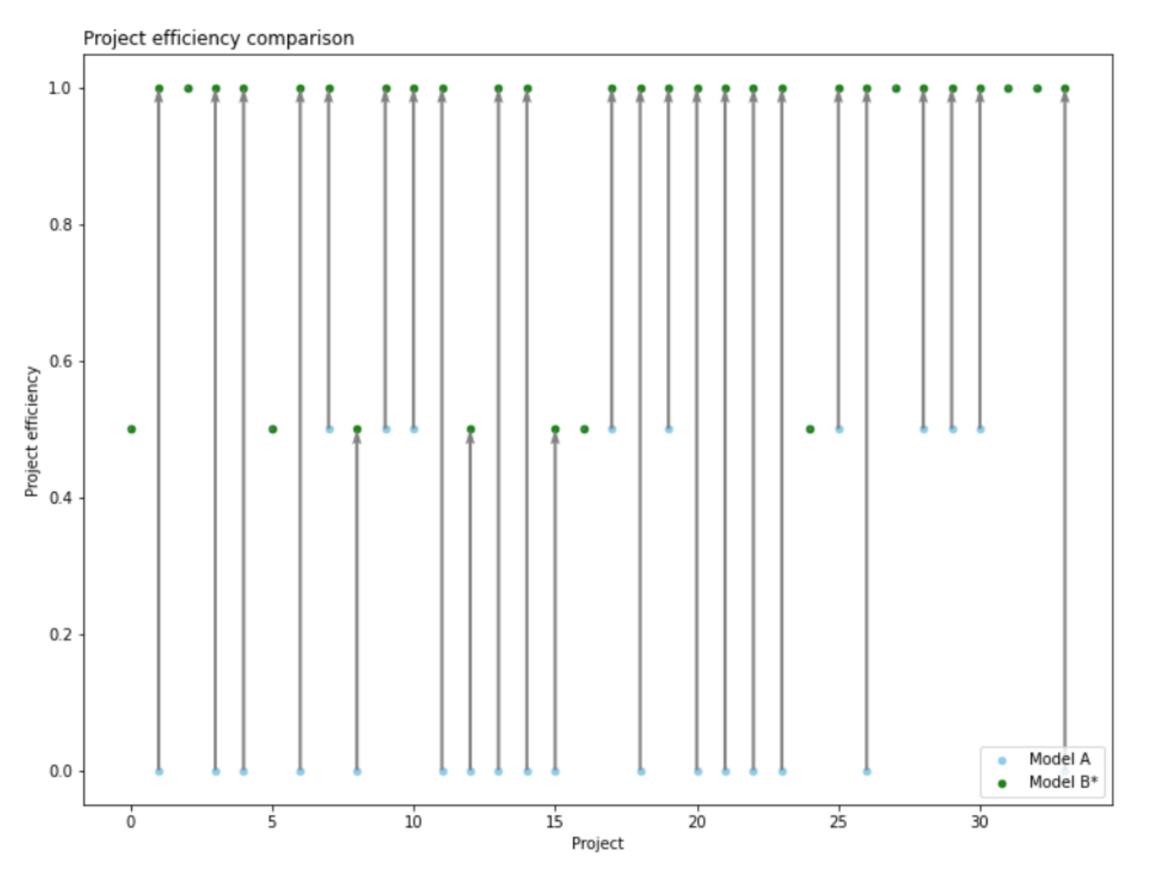
↑ 17

↓ 7

= 10

Experimentation

Predicted project efficiency values



Model variations			
Model variation	Min tardiness	Min idle time	Max efficiency
A	X	X	-
B*	X	X	X
C*	-	X	X
D	X	-	-
E	-	X	-
F*	-	-	X

 27

 0

= 7

Outline

- Company background
- Problem description
- Literature review
- Methodology
- Experimentation
- Conclusions

Conclusions

Further research

- **Other embeddings:**

  Regression Trees, Fuzzy Inferences Systems

- **Application:** Predictive maintenance in manufacturing

  Data: Production execution data with data label 'health index'

  Predict: health index HI_t using features workload amount/type, HI_{t-1}, .., HI_{t-k} values

  Predictive model: Regression tree

  Decision: Plan maintenance activities according to the course of the production plan.

THANKS FOR YOUR ATTENTION

Open Universiteit

WWW.OU.NL