FAIRNESS-AWARE MACHINE LEARNING IN EDUCATIONAL DATA MINING

PhD candidate: Tai Le Quy
Examiner 1: Prof. Dr. Eirini Ntoutsi
Examiner 2: Prof. Dr. Gunnar Friege
Chair: Prof. Dr. Johannes Krugel

Hannover, 16.10.2023

LernMINT

Did you know?

26% of respondents experienced discrimination during their studies at least once

46% observed discrimination against others

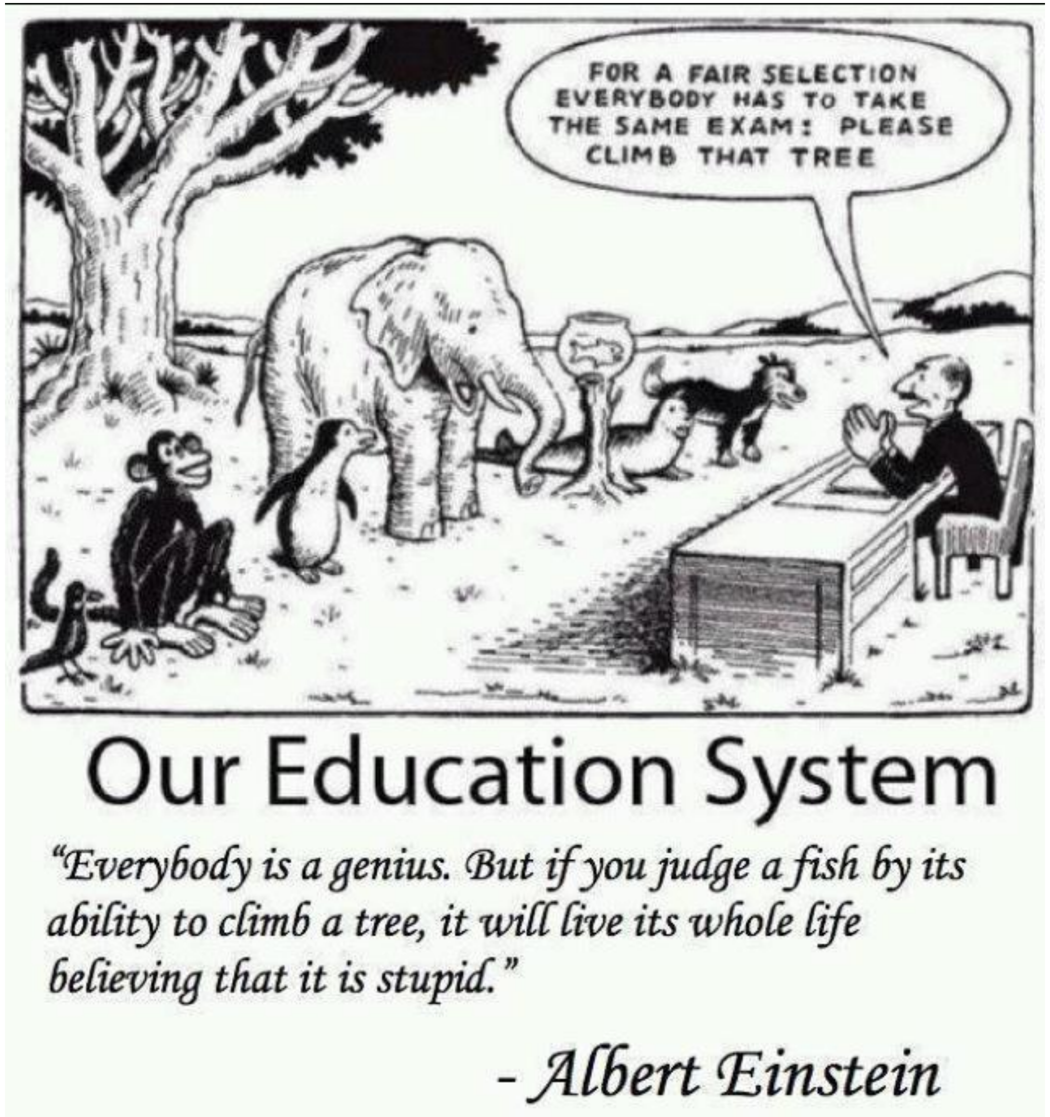
Online survey in 2021, included about 180,000 students from around 250 German universities by DZHW (German Centre for Higher Education Research and Science Studies) DZHW.

Source: https://www.dzhw.eu/pdf/pub_brief/dzhw_brief_08_2022.pdf

Source: http://gem-report-2017.unesco.org/en/chapter/gender_accountability_through_school/

Fairness in education

- Fairness is a fundamental concept of education
  - All students must have an equal opportunity to study
  - Be treated fairly regardless of their socioeconomic, assets, gender, or race
- Having a fair education system is crucial to achieving justice in a society (Meyer, 2014)



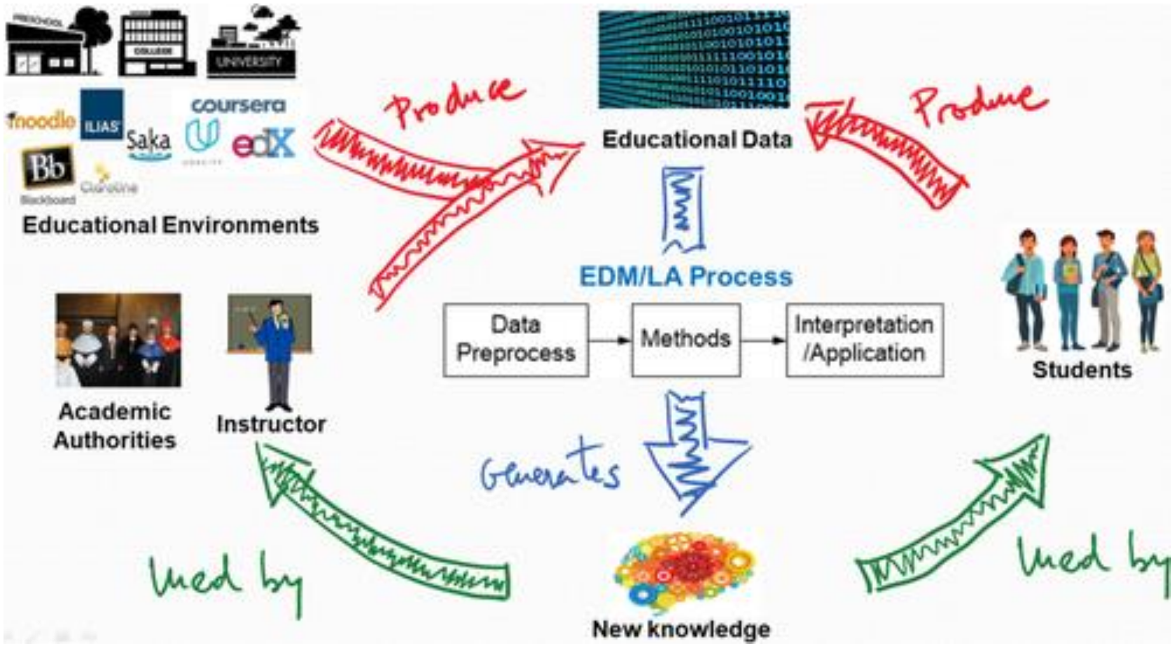
Source: https://blogs.ubc.ca/evaluation/2013/04/23/evaluation-and-fairness/

LernMINT

Educational Data Mining (EDM)

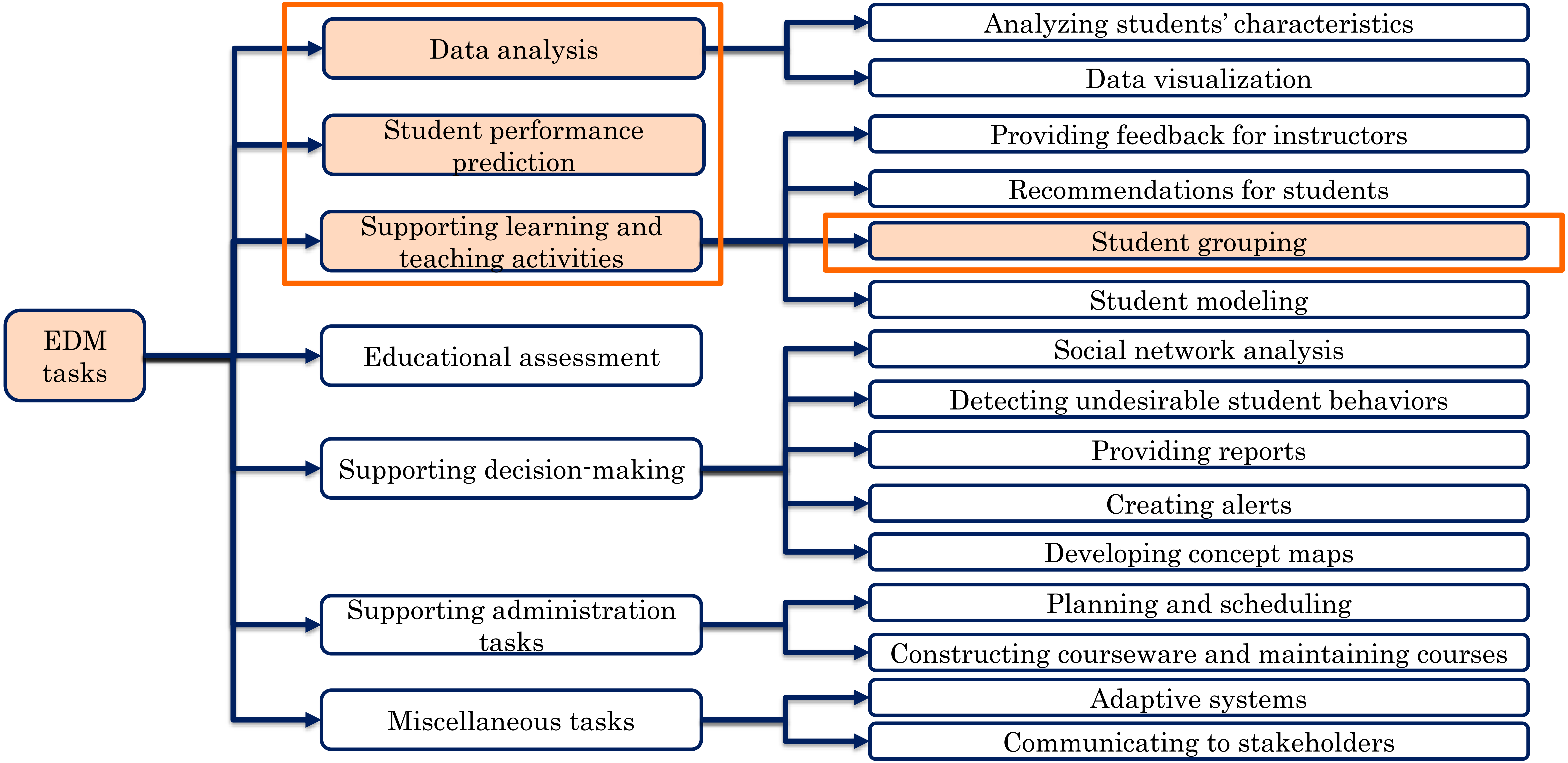
- Educational Data Mining
  "is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to understand better students, and the settings which they learn in" (EDM society, 2011).

- Important aspects
  - Objectives
  - Data
  - Techniques



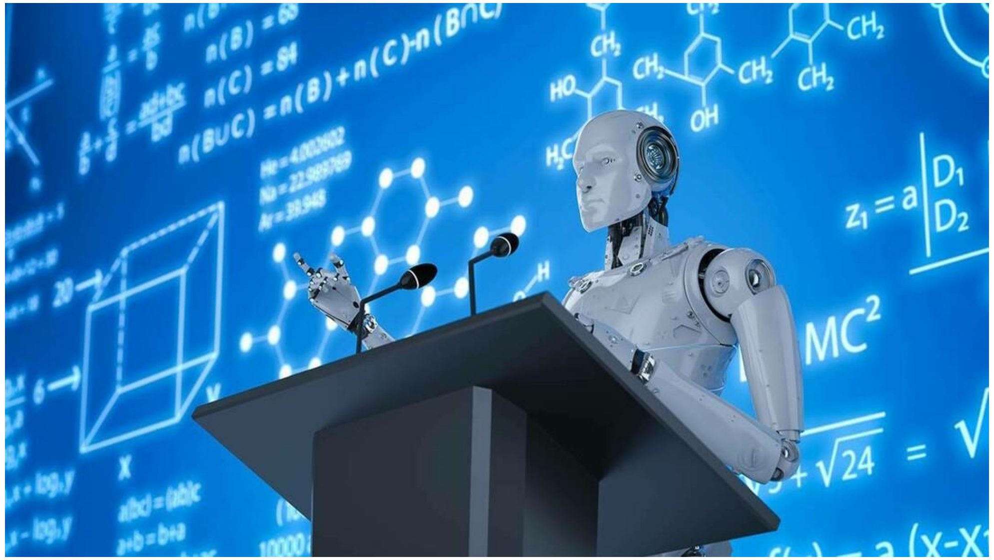
Source: Romero and Ventura (2020)

EDM tasks


EDM tasks
Data analysis
Student performance prediction
Supporting learning and teaching activities
Educational assessment
Supporting decision-making
Supporting administration tasks
Miscellaneous tasks
Analyzing students' characteristics
Data visualization
Providing feedback for instructors
Recommendations for students
Student grouping
Student modeling
Social network analysis
Detecting undesirable student behaviors
Providing reports
Creating alerts
Developing concept maps
Planning and scheduling
Constructing courseware and maintaining courses
Adaptive systems
Communicating to stakeholders

Machine Learning in EDM

- Machine Learning (ML) have been applied in a wide variety of EDM tasks
- The results of ML models are the basis for building applications in EDM (student data analysis, learning support, decision support systems, etc.)



Source: https://www.meer.com/en/72625-artificial-intelligence-in-education

The UK's A-level grading

The UK used a formula to predict students' scores for canceled exams. Guess who did well.

The formula predicted rich kids would do better than poor kids who'd earned the same grades in class.

By Kelsey Piper | Aug 22, 2020, 7:30am EDT

UK ditches exam results generated by biased algorithm after student protests

Protesters in London objected to the government's handling of exam results after exams were canceled due to the coronavirus outbreak. | Aaron Chown/PA Images via Getty Images

The UK has said that students in England and Wales will no longer receive exam results based on a controversial algorithm after accusations that the system was biased against students from poorer backgrounds, *Reuters* and *BBC News* report. The announcement followed a weekend of demonstrations at which protesters chanted "fuck the algorithm" outside the country's Department for Education.

Instead, students will receive grades based on their teachers' estimates after formal exams were canceled due to the pandemic. The announcement follows a similar U-turn in Scotland, which had previously seen 125,000 results downgraded.

Challenges

- ML-based decisions can be made based on protected attributes (gender, race, etc.) leading to discrimination
- Many fairness measures
  - Choosing proper measures can be cumbersome
  - There is no metric that fits all circumstances !!!
- Clustering models
  - Focus solely on the similarity objective
  - Do not consider the fairness of the resulting clusters w.r.t. protected attributes
  - Fail to account for the cardinality constraint of clusters
  - Do not consider the preferences of students

LernMINT

Research questions

RQ$_1$: How are protected attributes related to the class attribute in educational datasets? Does this relationship imply a dataset bias towards specific protected attributes?

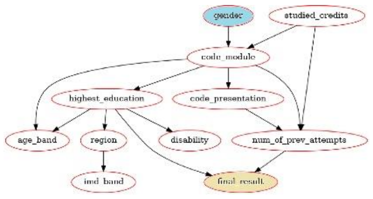
RQ$_2$: To what extent does the performance of (fairness-aware) classification models differ when applied to student performance prediction problems, considering various group fairness measures?

RQ$_3$: Which strategies can be utilized to achieve fairness in clustering models concerning both the protected attribute and cardinality constraints while dealing with student grouping problems?

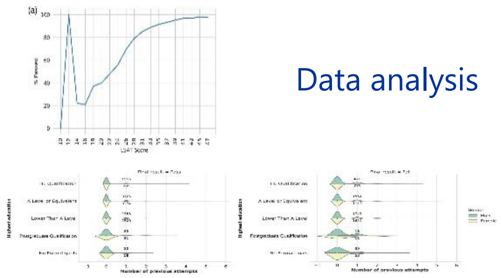
RQ$_4$: What approaches can be employed to achieve fairness in the students-topics grouping problem while considering multi-fairness constraints, cardinality, and taking into account students' preferences?

Main contributions

Bias-aware exploratory data analysis

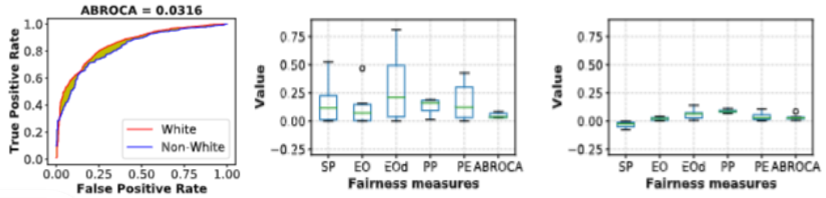


Bayesian network



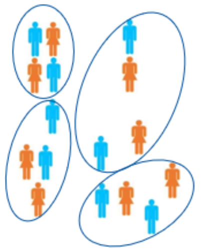
Data analysis

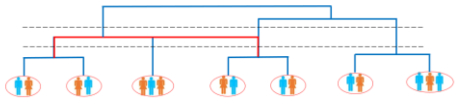
Evaluation of group fairness measures



Fairness-aware ML in EDM

Fair-capacitated clustering



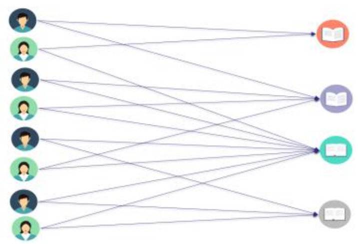


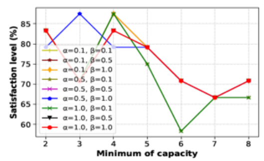
$$\text{maximize} \sum_{j=0}^{l} v_j y_j$$

$$\text{subject to} \sum_{j=1}^{l} w_j y_j \le q \quad \text{and } y_j \in \{0,1\}$$

Fair-capacitated clustering Knapsack

Multi-fair capacitated students-topics grouping problem





$$\text{maximize} \sum_{i=1}^{n} welfare_{ij_r} * y_{ij_r} \quad \text{s.t.} \begin{cases} \sum_{i=1}^{n} capacity_i * y_{ij_r} \le C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i * y_{ij_r} \le C^l \\ balance(G_r) \text{ is maximized} \end{cases}$$

Students-Topics Multi-fair capacitated Knapsack

Bias-aware exploratory data analysis

LernMINT

The goal

- Characterize datasets
- For each dataset
  - Use the Bayesian network (BN) to identify the relationships among attributes
  - Provide a graphical analysis of the attributes
  - Analyze features having a direct or indirect relationship with the protected attributes
- Quantitative evaluation of measures (predictive, fairness performance)

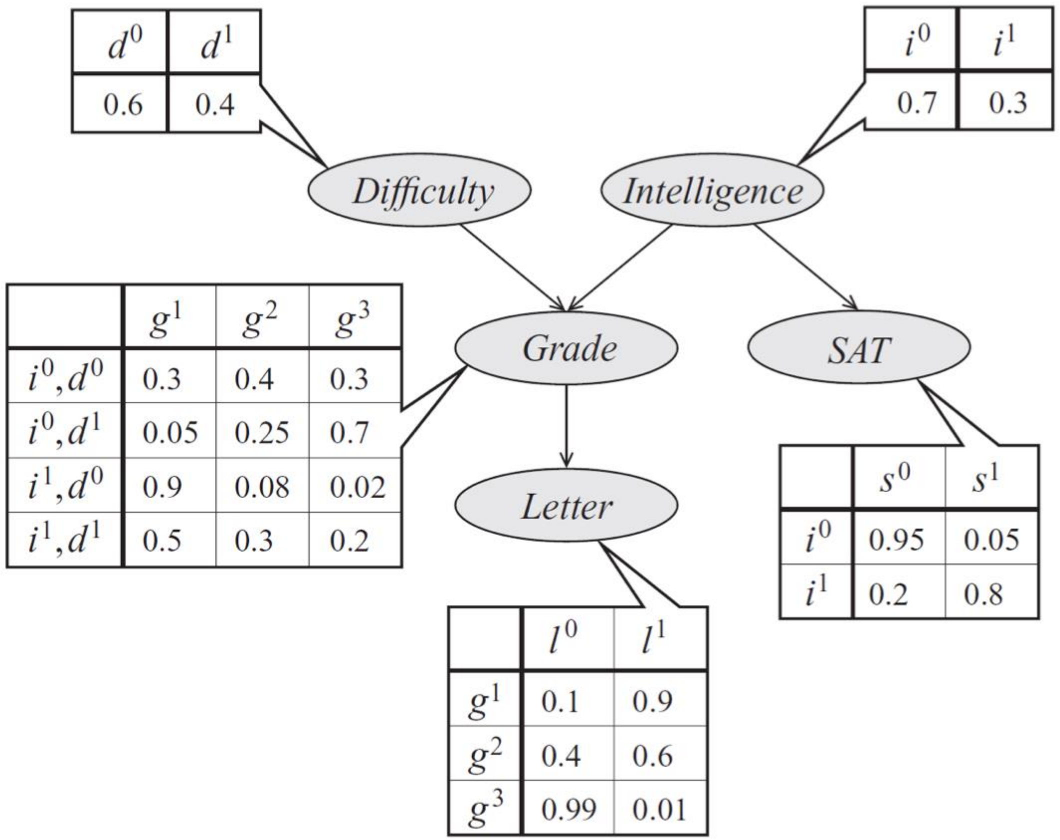
LernMINT

Educational datasets

Dataset	#Inst.	#Inst. (cleaned)	#Attributes (cat./bin./num.)	Class	IR (+:-)	Protected attributes	Target/ Class	Period	Location
Student-Math	395	395	4/13/16	Binary	2.04:1	Gender, age	Final grade	2005-2006	Portugal
Student-Por	649	649	4/13/16	Binary	5.49:1	Gender, age	Final grade	2005-2006	Portugal
OULAD	32,593	21,562	7/2/3	Multi	2.12:1	Gender	Final result	2013-2014	England
PISA	5,233	3,404	1/18/5	Binary	1.40:1	Male	Reading score	2009	The US
MOOC	416,921	393,465	9/4/8	Binary	1:27.0	Gender	Certified	2012-2013	The US
Law School	20,798	20,798	3/3/6	Binary	8.07:1	Male, race	Pass exam	1991	The US
Student aca.	131	131	17/5/0	Multi	3.85:1	Gender	ESP	2006-2013	India
xAPI-Edu-Data	480	480	9/4/4	Multi	2.78:1	Gender	Grade's level	2015	Jordan

Bayesian network

- Bayesian network (BN): probabilistic graphical model that measures the conditional dependence structure of a set of random variables based on the Bayes theorem
- Each node corresponds to a random variable
- Each edge represents the conditional probability for the corresponding random variables



A simple Bayesian network example

Source: https://jihongju.github.io/2018/11/11/pgm-lecture-note-01/

Bayesian network (cont.)

❖ The structure of a BN

The joint probability distribution of the attributes

$$P(A_1, A_2, \cdots, A_d) = \prod_{i=1}^{d} P(A_i \mid A_{pa_i})$$

$A_1, A_2, \dots, Ad$: attributes
A_{pa_i} : parent of A_i
Y : class attribute

❖ Position of the class attribute Y

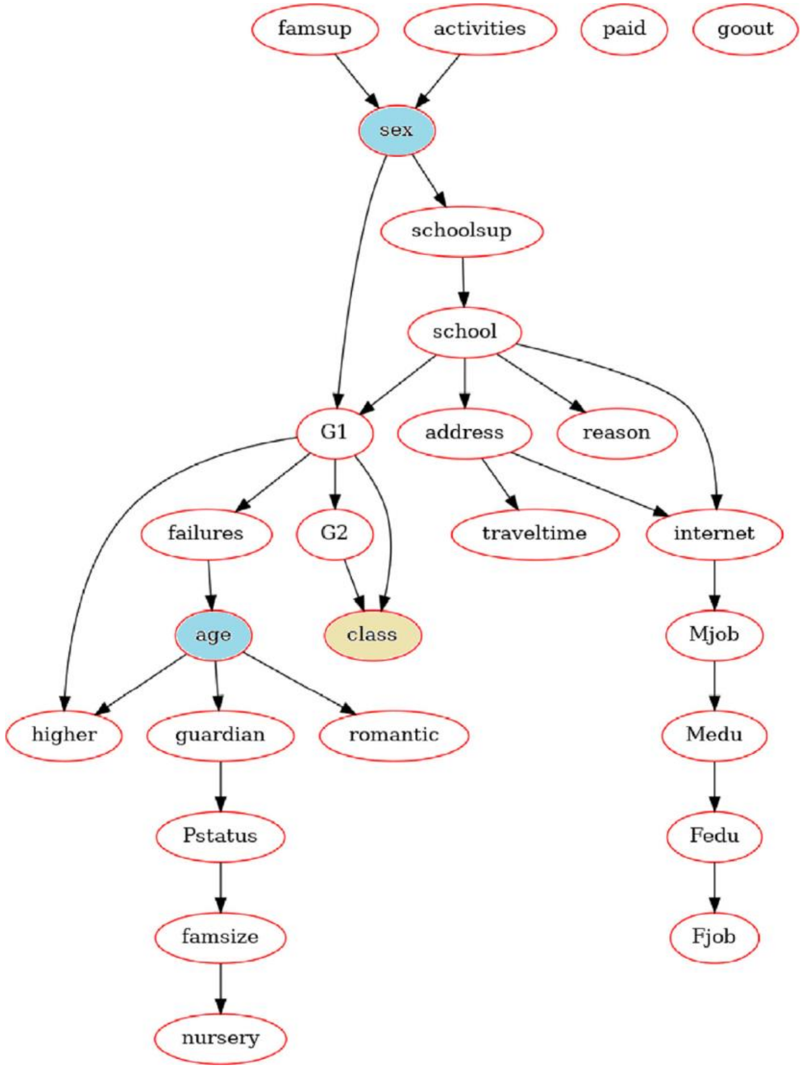
- In fact, Y can be in any position (root-, internal- or leaf-node)
- Our learning objective: the class attribute as a leaf node

$$\max_{\mathcal{M}^*}\{P(X \mid \mathcal{M}) - \gamma\widehat{\mathcal{M}}\}$$
$$\text{subject to } Y \in L$$

M : BN model,
$\widehat{M}$: set of parameters
M^* : optimal BN model
L : set of leaf nodes

Student performance dataset

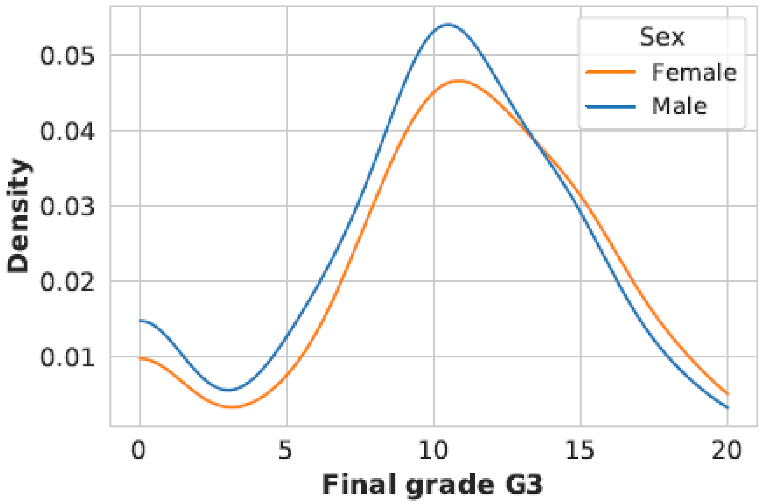
- Students' achievement in the secondary education of Portuguese schools (Mathematics, Portuguese subjects)
- Regression task: predict the final year grade of the students (G3 attribute, class = {Low, High} ({<10, ≥10})
- Protected attributes: age, sex
- The class label attribute is conditionally dependent on the grade G2 in both subsets



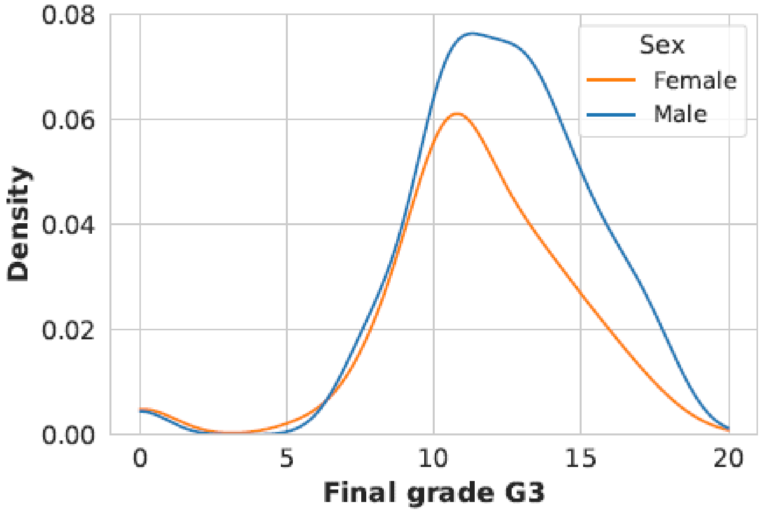
Student-Por: Bayesian network

Student performance dataset (cont.)

❖ The male students tend to receive high scores in the Portuguese subject, while the scores of Math are relatively evenly distributed across both sexes



(a) Mathematics subject

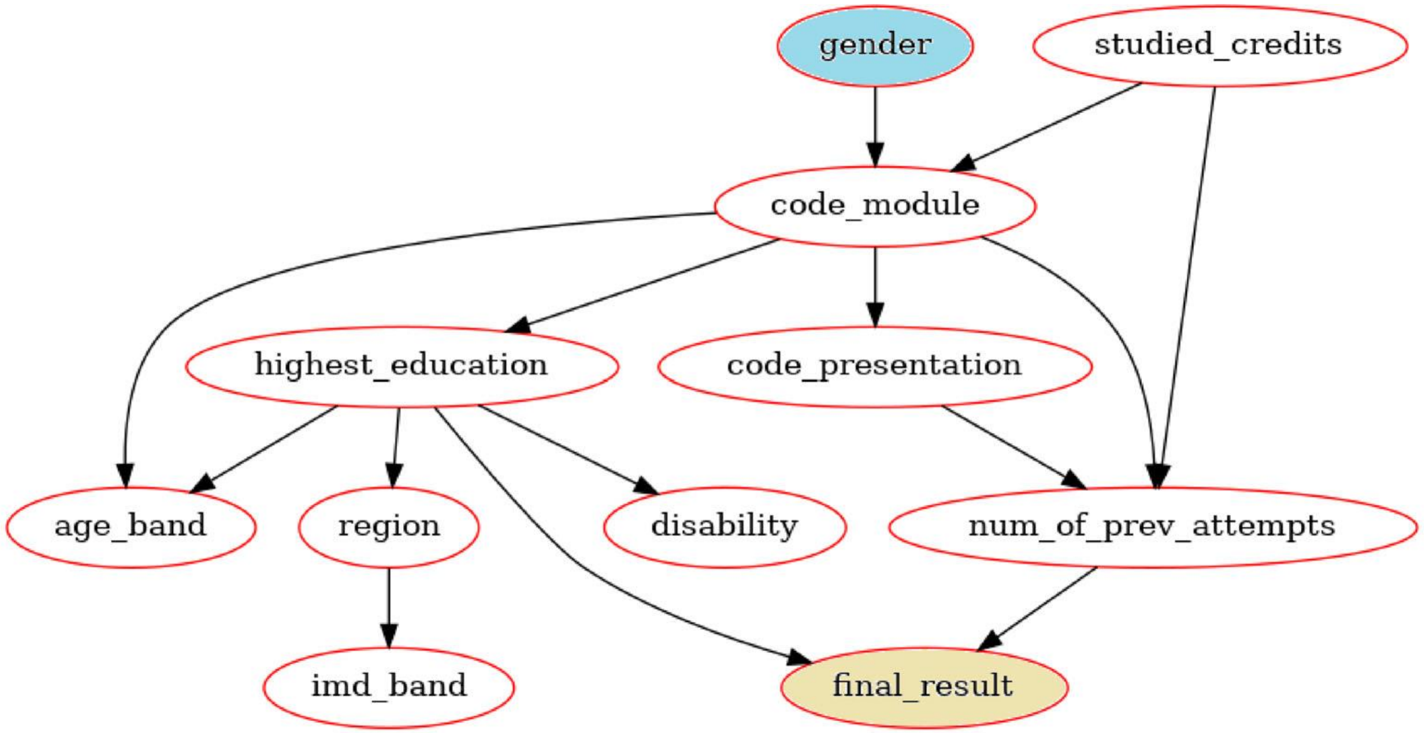


(b) Portuguese subject

Student-Por: Distribution of the final grade G3 with respect to sex

OULAD dataset

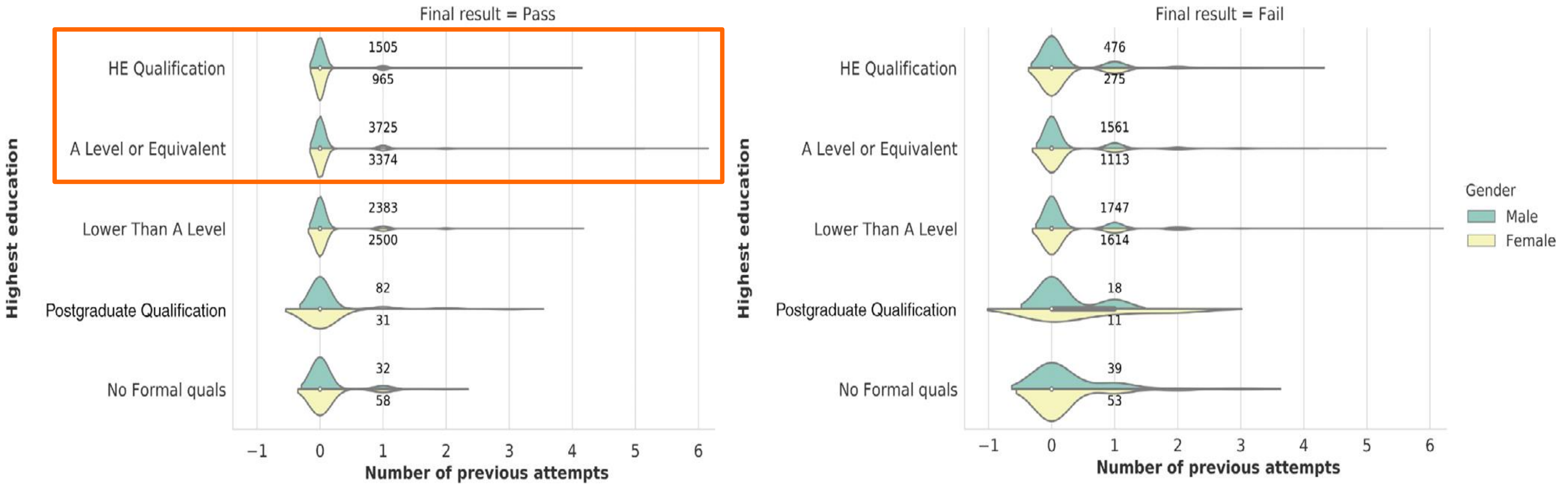
- OULAD dataset was collected from the OU analysis project
- The goal is to predict the success of students
- Protected attribute: gender



OULAD: Bayesian network

OULAD dataset (cont.)

- The ratio of male students having the highest education is "A level or equivalent" or "higher education (HE) qualification" is around 1.5 times higher than that of female students (gender-bias)

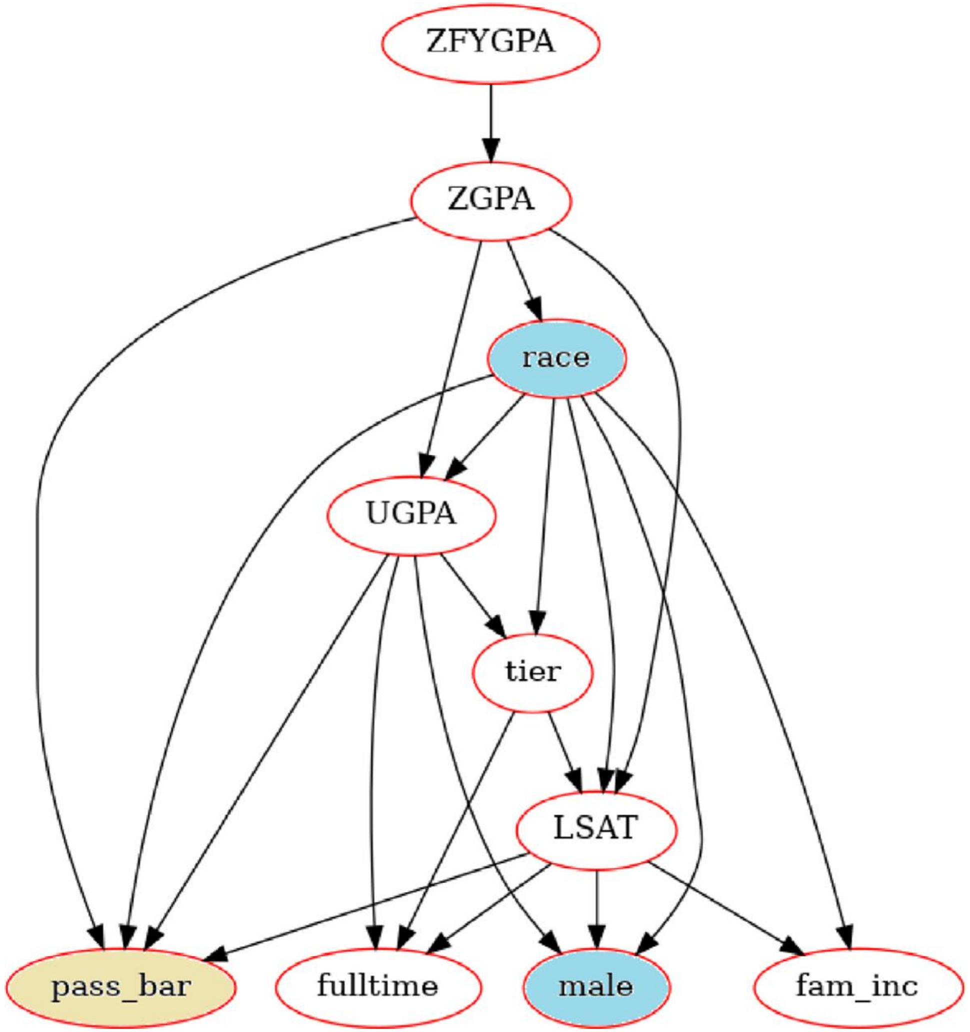


OULAD: Distribution of the number of previous attempts, the highest education and the final result with respect to gender

LernMINT

Law school dataset

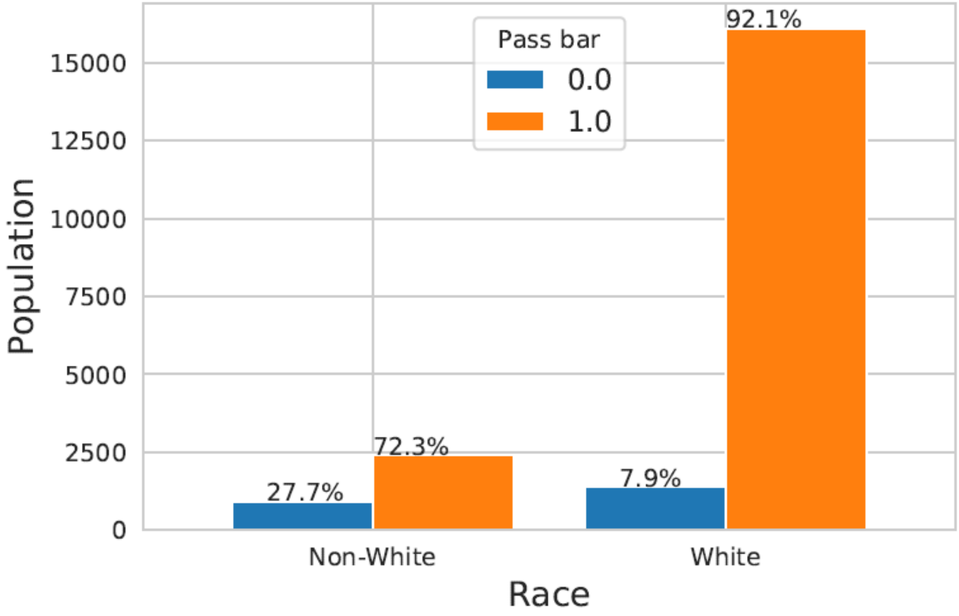
- The prediction task is to predict whether a candidate would pass the bar exam or predict a student's first-year average grade
- Protected attributes: male, race
- The bar exam's result is conditionally dependent on the law school admission test (LSAT) score, undergraduate grade point average (UGPA) and Race



Law school: Bayesian network

Law school dataset (cont.)

- White students have a higher chance to pass the bar exam (92.1%) than non-white students (72.3%)



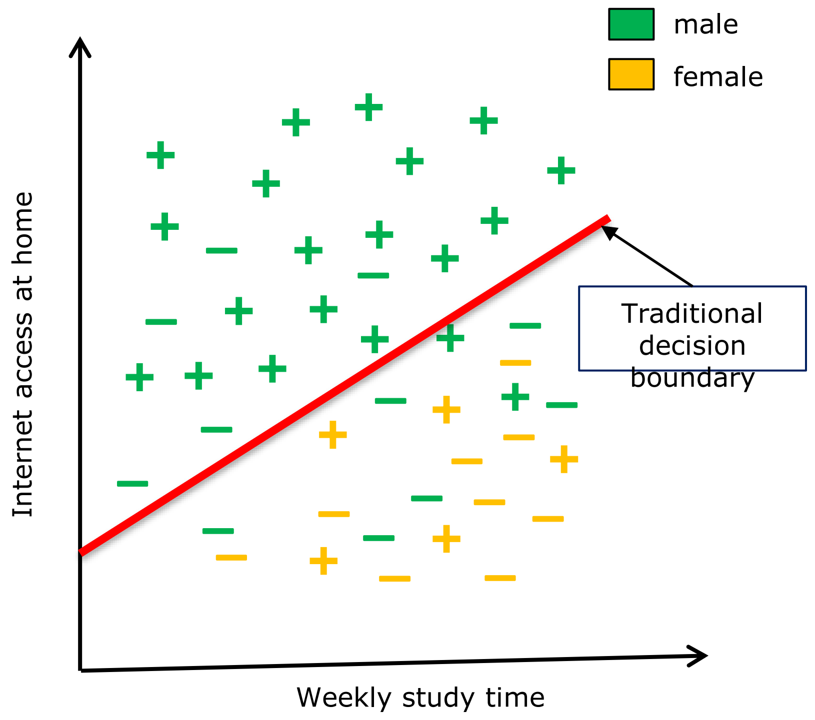
Law school: The percentage of students that passed w.r.t. Race

LernMINT

Evaluation of fairness measures in student performance prediction problems

Problem definition

- Student performance prediction problem is considered as a binary classification task:
  - X: a binary classification dataset
  - Class attribute $Y = \{+, -\}$, e.g., $Y = \{pass, fail\}$
  - $\mathcal{P}$: binary protected attribute, $\mathcal{P} \in \{p, \bar{p}\}$, e.g., $Gender \in \{female, male\}$
    - p: the discriminated group (protected group), e.g., *female*
    - $\bar{p}$: the non-discriminated group (non-protected group), e.g., *male*
  - Predicted outcome $\hat{Y} = \{+, -\}$



Fairness measures

The most prevalent group fairness notions

Measures	Proposed by	Published year	#Citations
Statistical parity	Dwork et al.	2012	2,367
Equal opportunity	Hardt et al.	2016	2,575
Equalized odds	Hardt et al.	2016	2,575
Predictive parity	Chouldechova et al.	2017	1,430
Predictive equality	Corbett-Davies et al.	2017	878
Treatment equality	Berk et al.	2018	626
Absolute Between-ROC Area	Gardner et al.	2019	84

The number of citations is reported by Google Scholar on 1^{st} August 2022.

Fairness measures (cont.)

- **Statistical parity (SP ∈ [−1, 1])**
  The difference in the predicted outcome ($\hat{Y}$) between any two groups

$$SP = P(\hat{Y} = +|\mathcal{P} = \overline{p}) - P(\hat{Y} = +|\mathcal{P} = p)$$

- **Equal opportunity (EO ∈ [0, 1])**
  The classifier should give similar results for students of both genders with actual "pass" class

$$P(\hat{Y} = +|\mathcal{P} = p, Y = +) = P(\hat{Y} = +|\mathcal{P} = \overline{p}, Y = +)$$

- **Equalized odds (EOd ∈ [0, 2])**
  Predicted true positive and false positive probabilities should be the same between male and female groups

$$P(\hat{Y} = +|\mathcal{P} = p, Y = y) = P(\hat{Y} = +|\mathcal{P} = \overline{p}, Y = y), \qquad y \in \{+, -\}$$

LernMINT

Fairness measures (cont.)

- Predictive parity (PP∈ [0, 1])

  The probability of a student predicted to “pass” actually having “pass” class should be the same, for both male and female

$$P(Y = +|\hat{Y} = +, \mathcal{P} = p) = P(Y = +|\hat{Y} = +, \mathcal{P} = \overline{p})$$

- Predictive equality (PE ∈ [0, 1])

  The probability of students with an actual “fail” class being incorrectly assigned to the “pass” class should be the same for both male and female

$$P(\hat{Y} = +|Y = -, \mathcal{P} = p) = P(\hat{Y} = +|Y = -, \mathcal{P} = \overline{p})$$

- Treatment equality

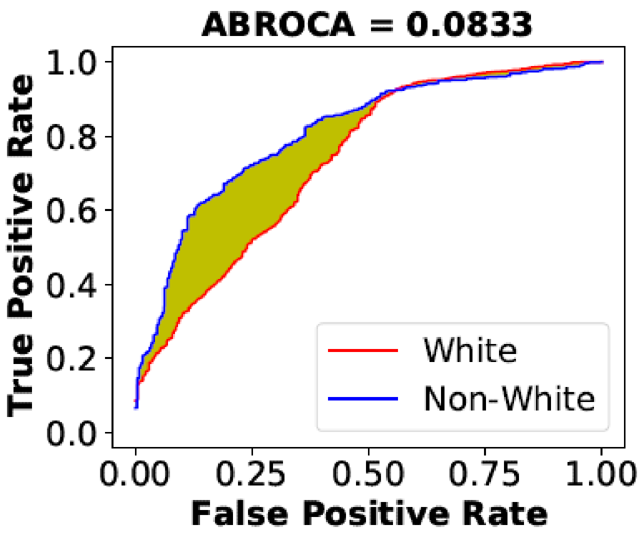
  The ratios of false negatives and false positives are the same for both male and female

$$\frac{FN_{prot.}}{FP_{prot.}} = \frac{FN_{non-prot.}}{FP_{non-prot.}}.$$

Fairness measures (cont.)

- Absolute Between-ROC Area (ABROCA ∈ [0, 1])
  - Measures the divergence between the protected (ROC_{p}) and non-protected group ($ROC_{\overline{\mathrm{p}}}$) curves across all possible thresholds $t \in [0,1]$ of FPR and TPR

$$\int_0^1 \mid ROC_p(t) - ROC_{\overline{\mathrm{p}}}(t) \mid \, dt.$$



Law school dataset: ABROCA measure (SVM classifier)

LernMINT

Evaluation setup

Datasets

Datasets	#Instances (cleaned)	#Attributes (cat./bin./num.)	Protected attribute	Class label	IR (+:-)
Law school (Law)	20,798	3/3/6	Race	Pass the bar exam	8.07:1
PISA	3,404	1/18/5	Male	Reading score	1.40:1
Studden aca. (S.Aca)	131	17/5/0	Gender	ESP	3.85:1
Student-Por (S.Por)	649	4/13/16	Gender	Final grade	5.49:1
xAPI-Edu-Data (xAPI)	480	9/4/4	Gender	Grade level	2.78:1

- Binarizing class labels:
  - PISA: reading score {<500, ≥500} ~ {low, high}
  - Student aca.: ESP (end semester percentage) {pass, good-and-higher}
  - Student-Por: final grade {<10, ≥10} ~ {fail, pass}
  - xAPI-Edu-Data: grade level {Low,Medium−High}
- 70% of data for training and 30% for testing (single split)

Evaluation setup (cont.)

- Predictive models
  - Traditional models
    - Decision Tree (DT)
    - Naive Bayes (NB)
    - Multi-layer Perceptron (MLP)
    - Support Vector Machines (SVM)
  - Fairness-aware models
    - Agarwal's: reduces the fair classification to a sequence of cost-sensitive classification problems with the lowest (empirical) error subject to the desired constraints (Agarwal et al., 2018)
    - AdaFair: updates the weights of the instances in each boosting round by considering a cumulative notion of fairness (Iosifidis and Ntoutsi, 2019)

LernMINT

Experimental results

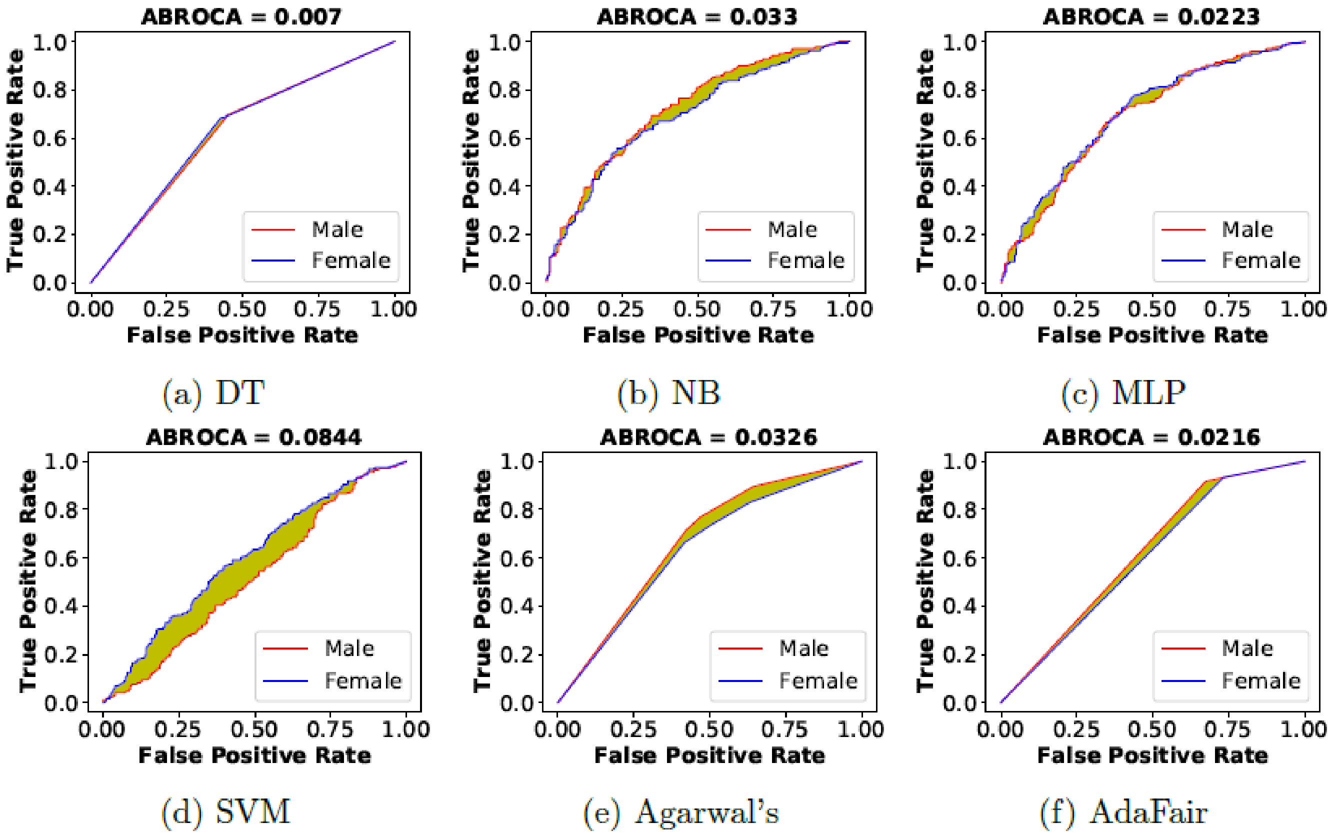
- ML model's performances variate over each value of the protected attribute

Measures	DT	NB	MLP	SVM	Agarwal's	AdaFair
Accuracy	0.9333	0.8974	0.9077	0.9231	0.8923	**0.9487**
Balanced accuracy	**0.8639**	0.8595	0.7840	0.7441	0.8565	0.8240
Statistical parity	-0.0382	-0.0509	-0.0630	**0.0151**	-0.0209	-0.0255
Equal opportunity	0.0125	0.0174	0.03	0.0183	0.0176	**0.0092**
Equalized odds	0.1316	0.2198	**0.1252**	0.3279	0.2200	0.1877
Predictive parity	**0.0456**	0.0591	0.0601	0.0944	0.0577	0.0639
Predictive equality	0.1190	0.2024	**0.0952**	0.3095	0.2024	0.1786
Treatment equality	2.0	7.5	**0.3333**	0.5	9.75	**0.3333**
ABROCA	0.0575	0.0686	0.0683	**0.0231**	0.0762	0.0887

Student-Por: performance of predictive models

Experimental results (cont.)

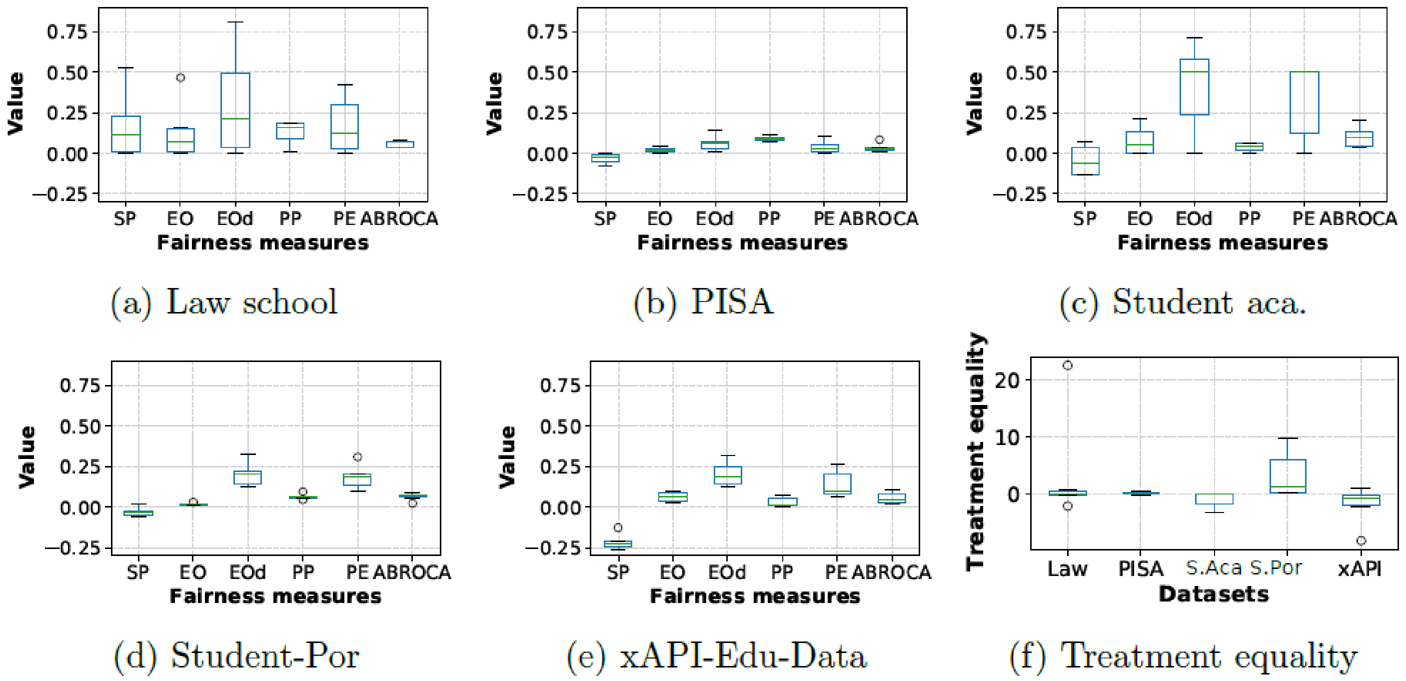
- ABROCA is the measure with the lowest variability across predictive methods and datasets



PISA: ABROCA slice plots

Experimental results (cont.)

- Equal opportunity and predictive parity also have a slight variation across methods and datasets.
- Equalized odds can represent two measures equal opportunity and predictive equality
- Treatment equality has a very wide range of values (the value may not be bounded)



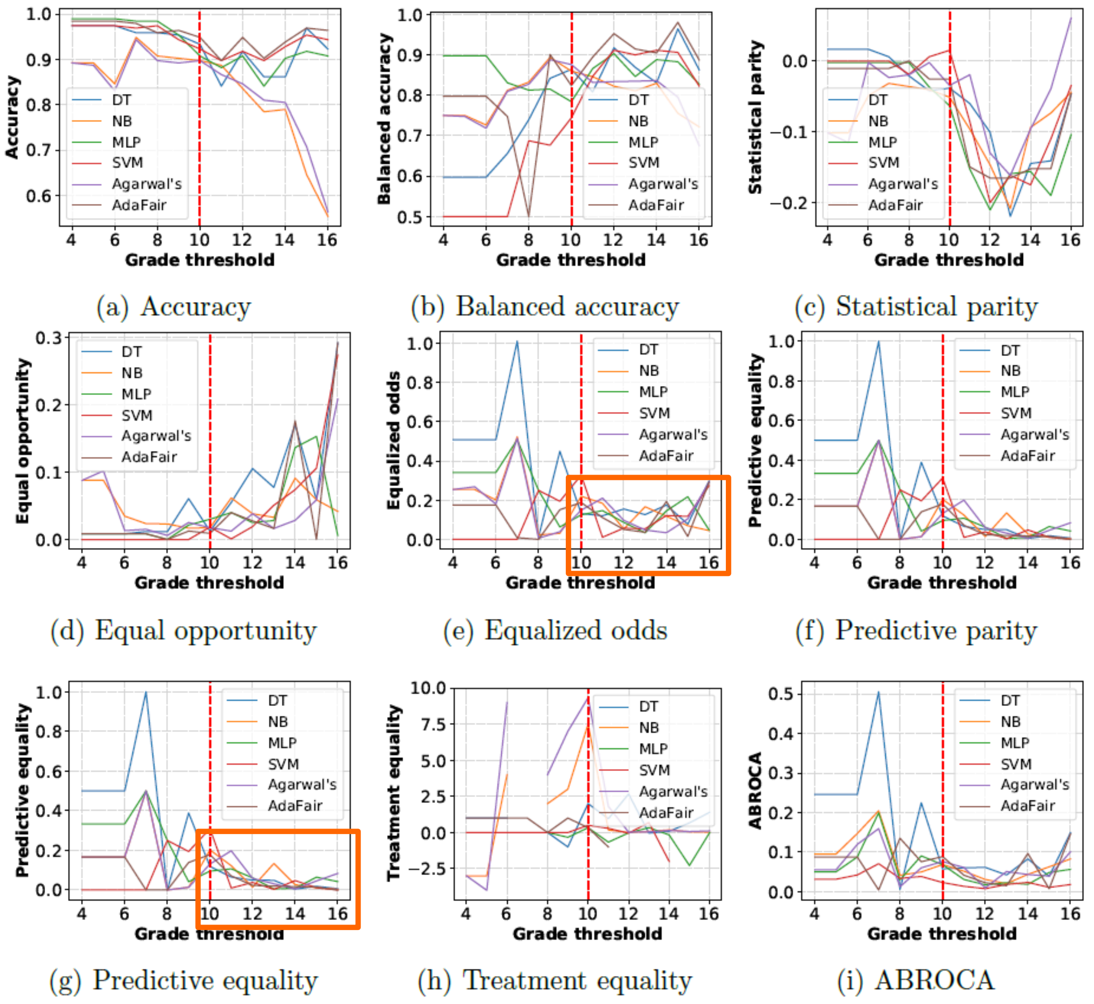
(a) Law school (b) PISA (c) Student aca.

(d) Student-Por (e) xAPI-Edu-Data (f) Treatment equality

Variation of fairness measures

Experimental results (cont.)

- Effect of varying grade threshold
  - All fairness measures are affected by the grade threshold
  - The predictive models tend to be fairer (equalized odds, predictive equality, and ABROCA) when the grade threshold is gradually increased

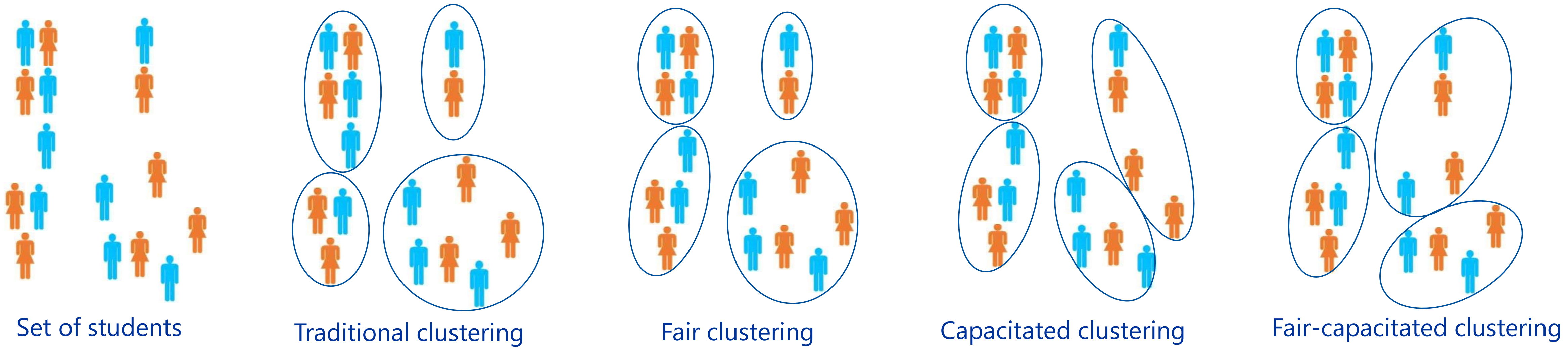


Accuracy and fairness interventions with varying grade threshold on Student-Por dataset

Fair-capacitated clustering

New problem: fair-capacitated clustering

- Clustering algorithm:
  - Considers the fairness w.r.t. protected attributes (fair clustering)
  - Considers the size of the clusters (capacitated clustering problem - CCP)



Problem definition

- (t,k,q) – fair-capacitated clustering
  - partitioning the data X by a clustering $\mathcal{C} = \{C_1, C_2, \dots, Ck\}$
  - $|\mathcal{C}| = k$ clusters
  - $|C_j| \leq q$ (capacity constraint)
  - balance($\mathcal{C}$) ≥ t (fairness constraint)
  - minimize the objective function $\mathcal{L}(X, \mathcal{C}) = \sum_{s_j \in S} \sum_{x \in C_j} dist(x, s_j)$
    - In which,

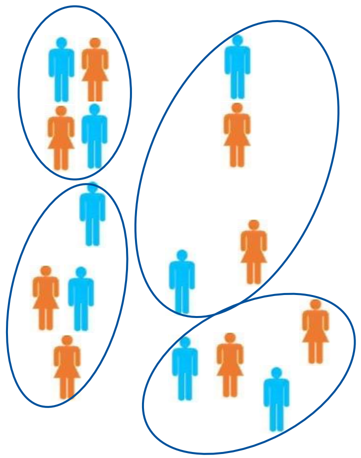
$$balance(\mathcal{C}) = min_{j=1}^{k} balance(C_j)$$

$$balance(C_j) = \min\left(\frac{|\{x \in C_j \mid \psi(x) = p\}|}{|\{x \in C_j \mid \psi(x) = \overline{p}\}|}, \frac{|\{x \in C_j \mid \psi(x) = \overline{p}\}|}{|\{x \in C_j \mid \psi(x) = p\}|}\right)$$

k - number of clusters

q - maximum cluster capacity

t - minimum balance threshold

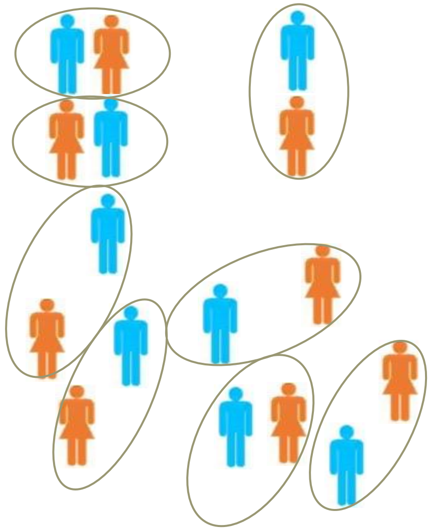


Fair-capacitated clustering

Fairlet decomposition

Fairlets

- Minimal sets that satisfy the balance requirement (Chierichetti et al., 2017)
- A clustering $\mathcal{F} = \{F_1, F_2, .., Fl\}$ of X is a fairlet decomposition if:
    - Each point $x \in X$ belongs to exactly one cluster F_i
    - $|F_i| \leq f + m$ for each F_i (the size of each fairlet is small)
    - balance(F_i) $\geq t$ (minimum balance threshold)
- Each F_i is a fairlet



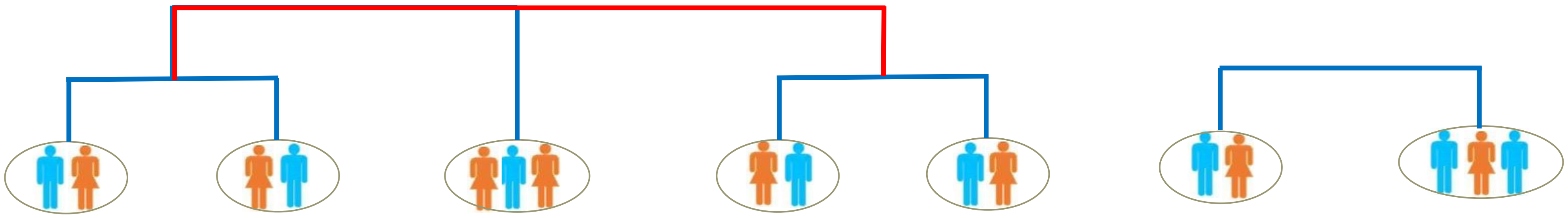
Proposed methods

- 2-phases algorithm
  - Phase 1: Compute the fairlet decomposition
    - Input: A set of instances X
    - Output: A clustering $\mathcal{F} = \{F_1, F_2, .., Fl\}$: a fairlet decomposition
  - Phase 2: Cluster the centers of fairlets into k groups
    - Input: The fairlet decomposition $\{F_1, F_2, .., Fl\}$
    - Output: The fair-capacitated clustering
    - Methods:
      - Hierarchical-based approach
        - Consider the cardinality of final clusters in the merging step
      - Partitioning-based approach
        - Re-formulate the assignment step of k-Medoids algorithm as a **Knapsack** problem

The knapsack problem

Fair-capacitated hierarchical clustering

- Hierarchical-based approach
  - Each point is a fairlet
  - Consider the cardinality of final clusters in the merging step



Example of hierarchical-based approach with the maximum capacity $q = 7$

Fair-capacitated partitioning-based clustering

k-Medoids fair-capacitated algorithm

- For each instance (fairlet) F_{i} we assign a value $v_{\mathrm{i}} = e^{-\frac{1}{\lambda} * d(F_i, s_{\mathrm{j}})}$
  - $s_j \in S$ cluster centers (medoids)
  - $d(F_{\mathrm{i}}, s_{\mathrm{j}})$ Euclidean distance between F_i and s_j
- Formulate the cluster assignment step as a 0-1 knapsack problem

$$\text{maximize} \sum_{\mathrm{i}=0}^{l} v_i y_i$$

$$\text{subject to} \sum_{\mathrm{i}=1}^{l} w_{\mathrm{i}} y_{\mathrm{i}} \le q \quad \text{and } y_i \in \{0,1\}$$

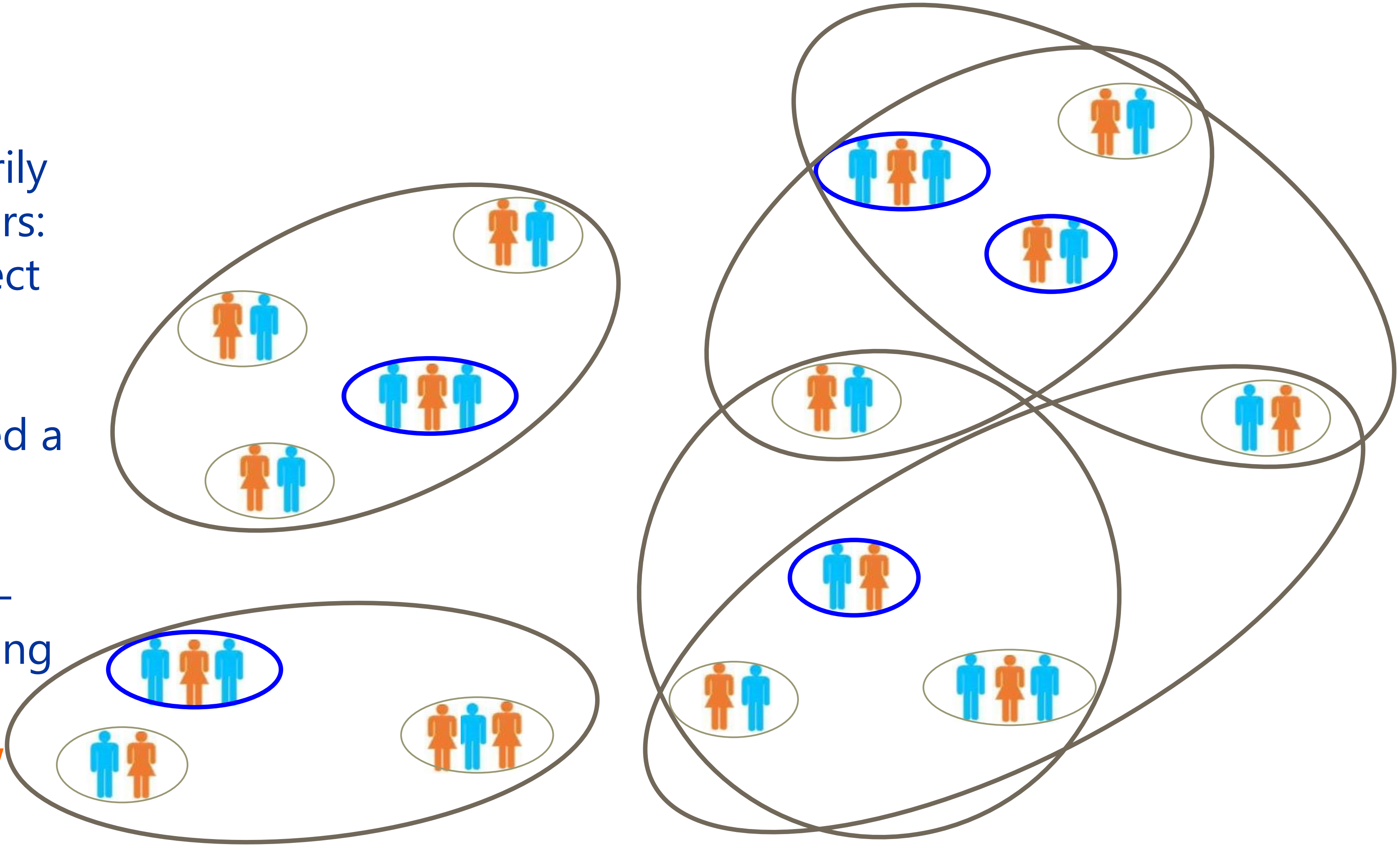
$$y_j = \begin{cases} 1 \text{ if } F_{\mathrm{i}} \text{ is assigned to a cluster} \\ 0 \text{ if } F_i \text{ is not assigned to a cluster} \end{cases}$$

$\mathcal{F} = \{F_1, \dots, F_l\}$: fairlets
$W = \{w_1, \dots, w_l\}$: weights
$w_i = |F_i|$

- We assign the most 'suitable' fairlets to each medoid s_j by the solution of 0-1 knapsack problem

Fair-capacitated partitioning-based clustering (cont.)

- k-Medoids fair-capacitated algorithm
  - Select *k* medoids arbitrarily
  - Assign fairlets to *k* clusters:
    - For each medoid select a set of fairlets by knapsack problem (capacity is not exceed a given threshold)
  - For each medoid
    - Try to swap with non-medoid if the clustering cost is reduced
    - Assign fairlets to new medoid



Example of k-Medoids fair-capacitated algorithm with maximum capacity $q = 9$

LernMINT

Evaluation setup

Dataset

Dataset	#Instances (cleaned)	#Attributes (cat./bin./num.)	Protected attribute	Balance score
Student-Math	395	4/13/16	Gender (F: 208, M: 187)	0.899
Student-Por	649	4/13/16	Gender (F: 383; M: 266)	0.695
PISA	3,404	1/18/5	Male (1: 1,697; 0: 1,707)	0.994
OULAD	4,000	7/2/3	Gender(F: 2,000; M: 2,000)	1
MOOC	4,000	9/4/8	Gender (F: 2,000; M: 2,000)	1

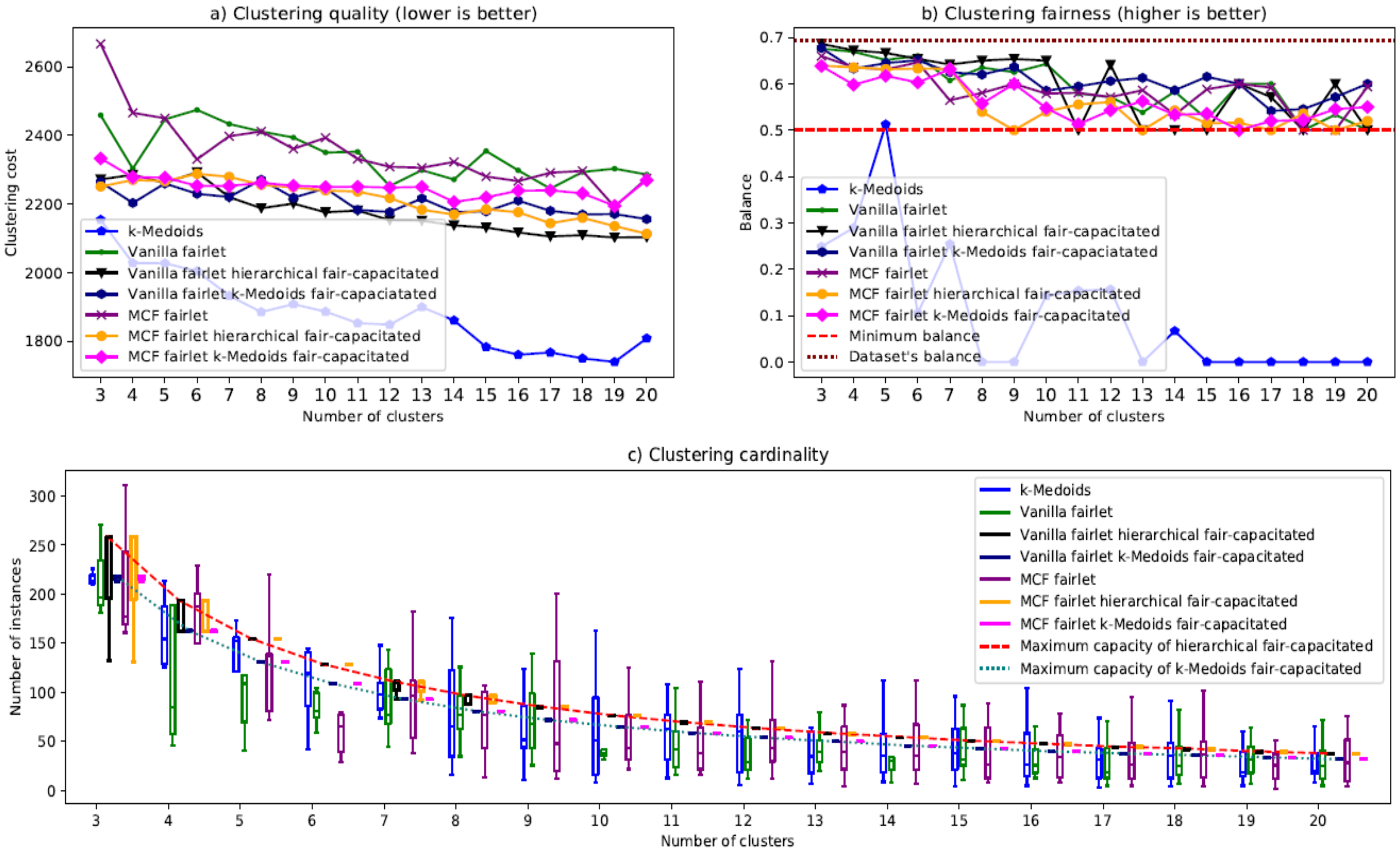
Experimental setup

- Baseline
  - k-Medoids
  - Vanilla fairlet
  - MCF fairlet
- Measurement
  - Clustering cost $\mathcal{L}(X, \mathcal{C}) = \sum_{s_j \in S} \sum_{x \in C_j} dist(x, s_j)$
  - Balance score $balance(\mathcal{C}) = min_{j=1}^{k} balance(C_j)$
  - Cardinality $q = \left\lceil \frac{n * \varepsilon}{k} \right\rceil$

LernMINT

Experimental results

- Clustering cost
  Our methods outperform competitors
- Fairness
  Well maintain the fairness above the minimum threshold
- Cardinality
  - Well maintain the capacity of clusters within the maximum cardinality
  - k-Medoid based method is better than Hierarchical based approach

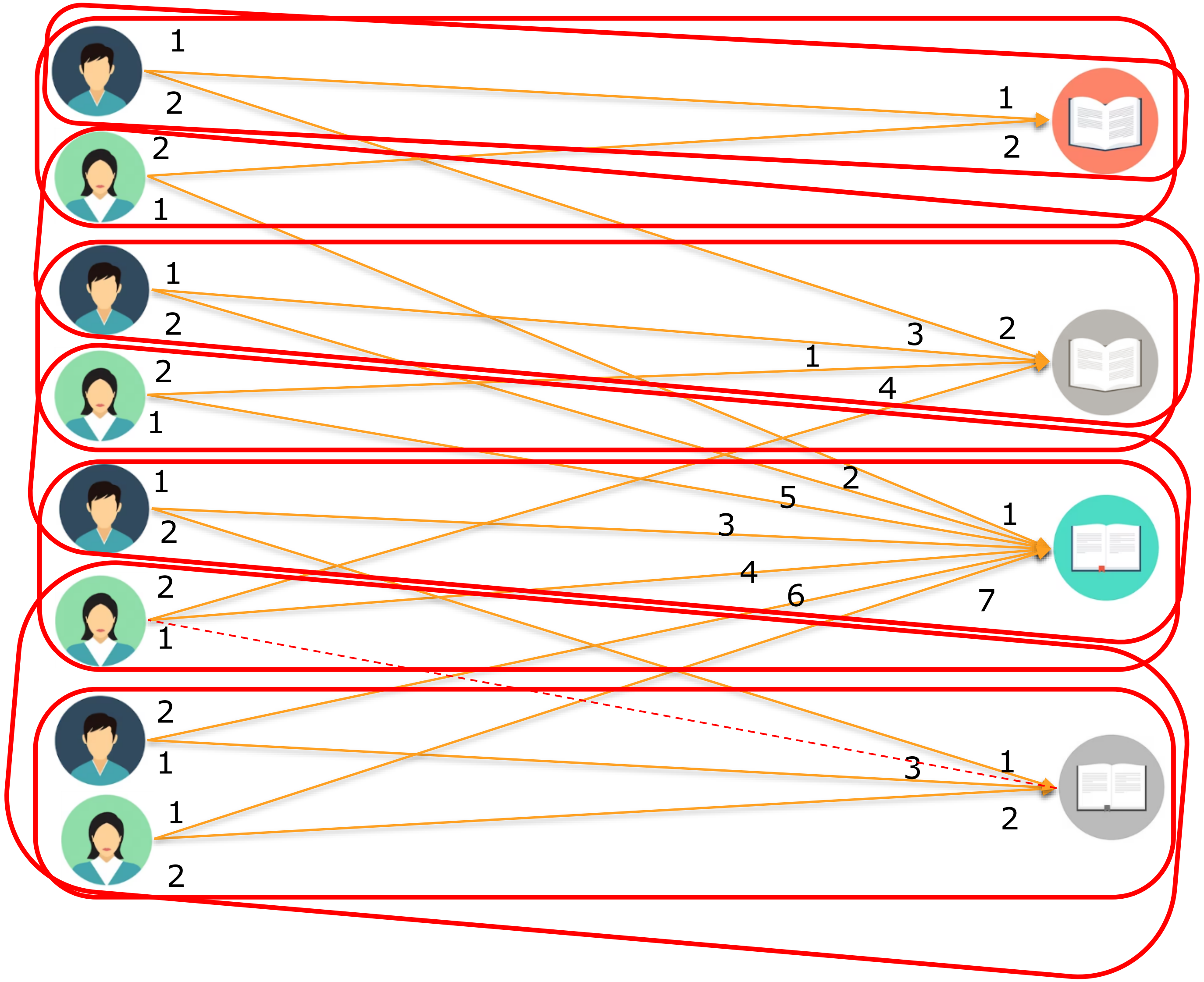


Performance of different methods on Student-Por dataset

Multi-fair capacitated students-topics grouping problem

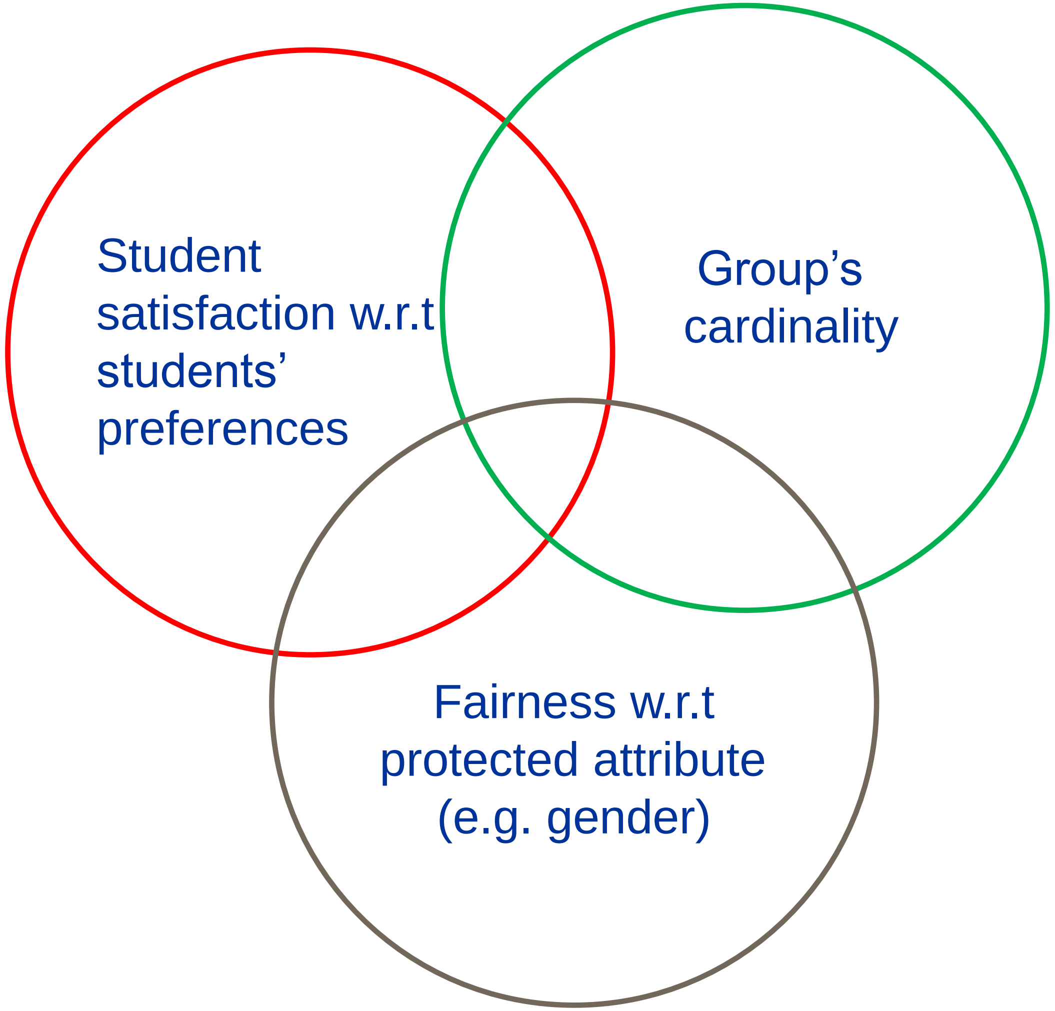
Students-topics grouping problem

Students



Topics

Students-topics grouping problem (cont.)



Problem definition

- $X = \{x_1, \dots, x_n\}$: n students, $T = \{t_1, \dots, tm\}$: a set of m topics

$h = 3$ preferences ($n = 5$ students)

	$wish_1$	$wish_2$	$wish_3$
x_1	t_1	t_3	t_2
x_2	t_3	t_1	t_4
x_3	t_2	t_3	t_1
x_4	t_4	t_2	t_3
x_5	t_1	t_4	t_3

a) Matrix $wishes_{n \times h}$

Students choose h topics as their wishes

$m = 4$ topics ($n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	3	1	1.5	0
x_2	1.5	0	3	1
x_3	1	3	1.5	0
x_4	0	1.5	1	3
x_5	3	0	1	1.5

b) Matrix $V_{n \times m}$

V: level of interest in the topic

$m = 4$ topics ($n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	4	1	3	0
x_2	3	0	5	3
x_3	2	2	4	0
x_4	0	3	2	1
x_5	1	0	1	2

c) Matrix $W_{n \times m}$

W: time-weight matrix based on registration time

$m = 4$ topics ($n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	2.0	0.67	1.1	0
x_2	1.25	0	2.0	1.08
x_3	0.83	1.67	1.3	1.0
x_4	0	1.5	0.73	1.25
x_5	1.25	0	0.53	1.0

d) Matrix $welfare_{n \times m}$

$$welfare_{ij} = \alpha v_{ij} + \beta w_{ij}$$

- Protected attribute, e.g., gender, $\psi(x_i) = \{p, \bar{p}\}$, i.e. $\{female, male\}$

Problem definition (cont.)

- The goal is to divide all students into k groups $\mathcal{G} = \{G_1, \ldots, G_k\}$, $k \leq m$, which maximizes the objective function:

$$\mathcal{L}(X, \mathcal{G}) = \prod_{r=1}^{k} \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r}$$

$\mathcal{L}(X, \mathcal{G})$ is the Nash social welfare (Nash equilibrium) function*

  - The group assignment is satisfactory, i.e., maximizing the objective function (students' satisfaction)
  - $balance(G_r)$ is maximized: fairness constraint w.r.t protected attribute
  - $C^l \leq |G_r| \leq C^u$: capacity constraint

where: $J = \{j_1, \ldots, j_k\} = \{j \mid x_i \in G_r, welfare_{ij} > 0\}$, r = 1..k

$$y_{ij_r} = \begin{cases} 1 \; if \; x_i \; is \; assigned \; to \; topic \; t_{j_r} \\ 0 \qquad\qquad\qquad if \; not \end{cases}$$

$$balance(G_r) = \min\left(\frac{\{x \in G_r | \psi(x) = p\}}{\{x \in G_r | \psi(x) = \bar{p}\}}, \frac{\{x \in G_r | \psi(x) = \bar{p}\}}{\{x \in G_r | \psi(x) = p\}}\right)$$

Multi-fair capacitated (MFC) grouping problem

* Fluschnik et al. (2019), Fair knapsack. In *AAAI, 2019*

LernMINT

Proposed methods

- **Greedy heuristic approach** — Student's preferences

  Assign students to the most preferred topic among their preferences

- **Knapsack-based approach** — Group's cardinality

  Search the most suitable students for each topic by a maximal knapsack problem

- **MFC knapsack approach** — MFC constraints

  Search the most suitable students for each topic by a new MFC knapsack satisfying constraints of the MFC problem

Greedy heuristic approach

- 2-step approach
  - Assign students to groups
    - Assign students to their most preferred topic
    - If many students choose the same topic, we assign the student with the highest *welfare* value to the topic
  - Group adjustment
    - To satisfy constraints (fairness w.r.t. protected attribute, cardinality).
    - If there are ungrouped students, we will try to assign them to existing groups

Knapsack-based approach

- Select suitable students for a group by a *maximal knapsack* problem
  - For each topic $t_{j_r} \in T$, r is the index of k selected topic $J = \{j_1, j_2, \ldots, j_k\}$, select a subset of students (G_r):

$$\text{maximize} \sum_{i=1} welfare_{ij_r} * y_{ij_r}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^l \end{cases}$$

where $y_{ij_r} = 1$ if x_i is assigned to topic t_{j_r}, else $y_{ij_r} = 0$

  - value ~ welfare, weight ~ capacity

The knapsack problem

LernMINT

MFC knapsack approach

MFC knapsack algorithm

Search the group of suitable student w.r.t. MFC constraints: select a subset G_r:

$$\text{maximize} \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^l \\ balance(G_r) \text{ is maximized} \end{cases}$$

where

$$y_{ij_r} = \begin{cases} 1, & \text{if student } x_i \text{ is assigned to topic } t_{j_r} \\ 0, & \text{otherwise} \end{cases}, \forall r \in [k]$$

LernMINT

MFC knapsack approach (cont.)

- 2-step approach
  - Assign students to groups
    - Select suitable candidates among unassigned students by the result of a group fairness MFC knapsack problem
    - Use dynamic programming to solve the MFC knapsack problem (inspired by knapsack problem with group fairness constraints of Patel et al. (2021)
  - Group adjustment
    - Apply the same procedure as in the greedy heuristic approach

Evaluation setup

Dataset

Dataset	#Instances	#Attributes	Protected attribute	Balance score
Real data science	24	23	Gender (F: 8, M: 16)	0.5
Student-Math	395	33	Gender (F: 208, M: 187)	0.899
Student-Por	649	33	Gender (F: 383; M: 266)	0.695

Measures

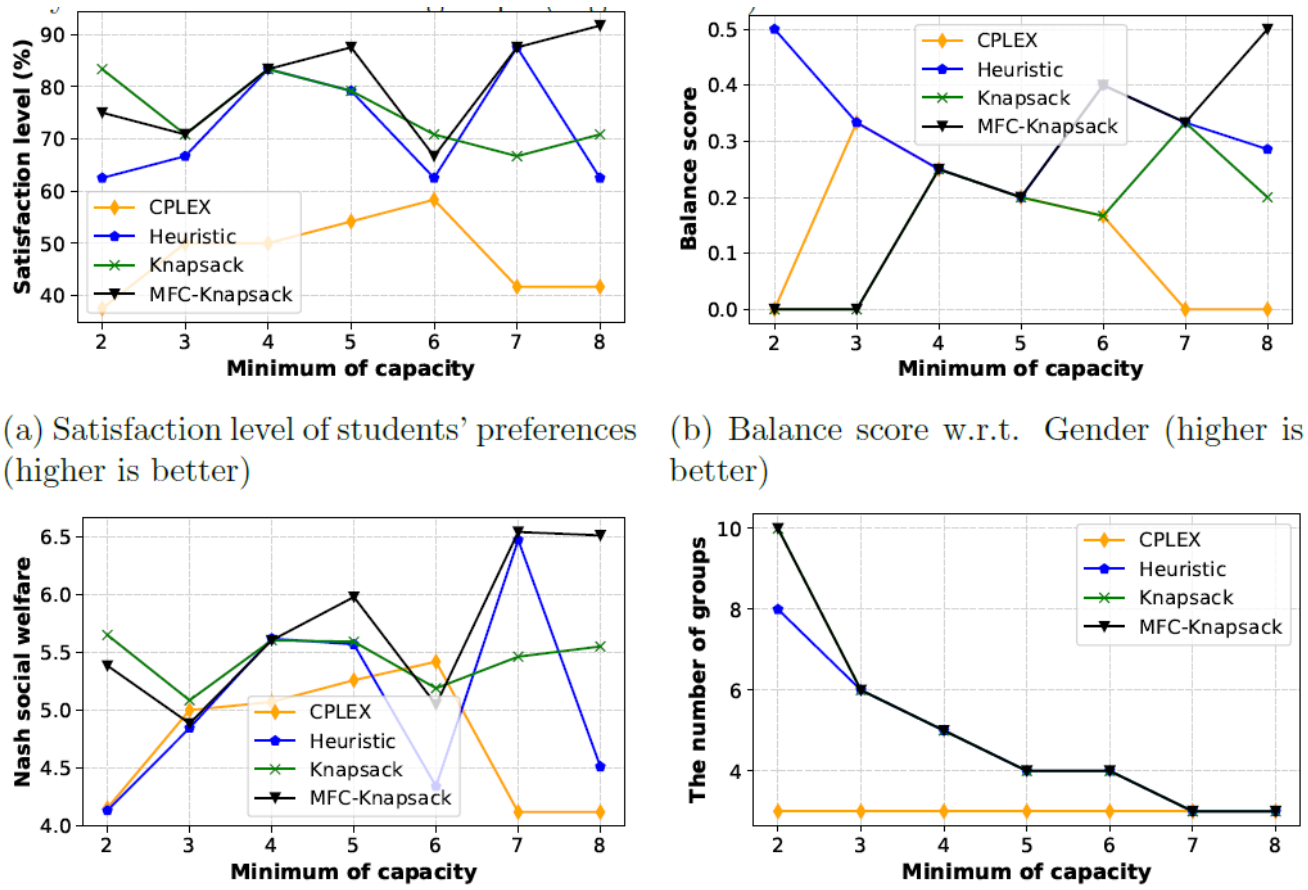
- Nash social welfare $Nash = log_k \mathcal{L}(X, \mathcal{G})$
- Balance score $balance(\mathcal{G}) = \min_{\forall G_r \in \mathcal{G}} balance(G_r)$
- Satisfaction level $Satisfaction = \frac{\mid \{i \mid wishes_{io} = k, i \in groups_k, o \in [h]\} \mid}{n}$

Baseline

- The CPLEX integer programming model (Magnanti et al, 2018)

Experimental results

- The MFC knapsack is better:
  - In terms of the Nash social welfare and satisfaction level
  - When a group has at least 4 people
- CPLEX fails to assign students while maintaining only a constant number of groups



(a) Satisfaction level of students' preferences (higher is better)

(b) Balance score w.r.t. Gender (higher is better)

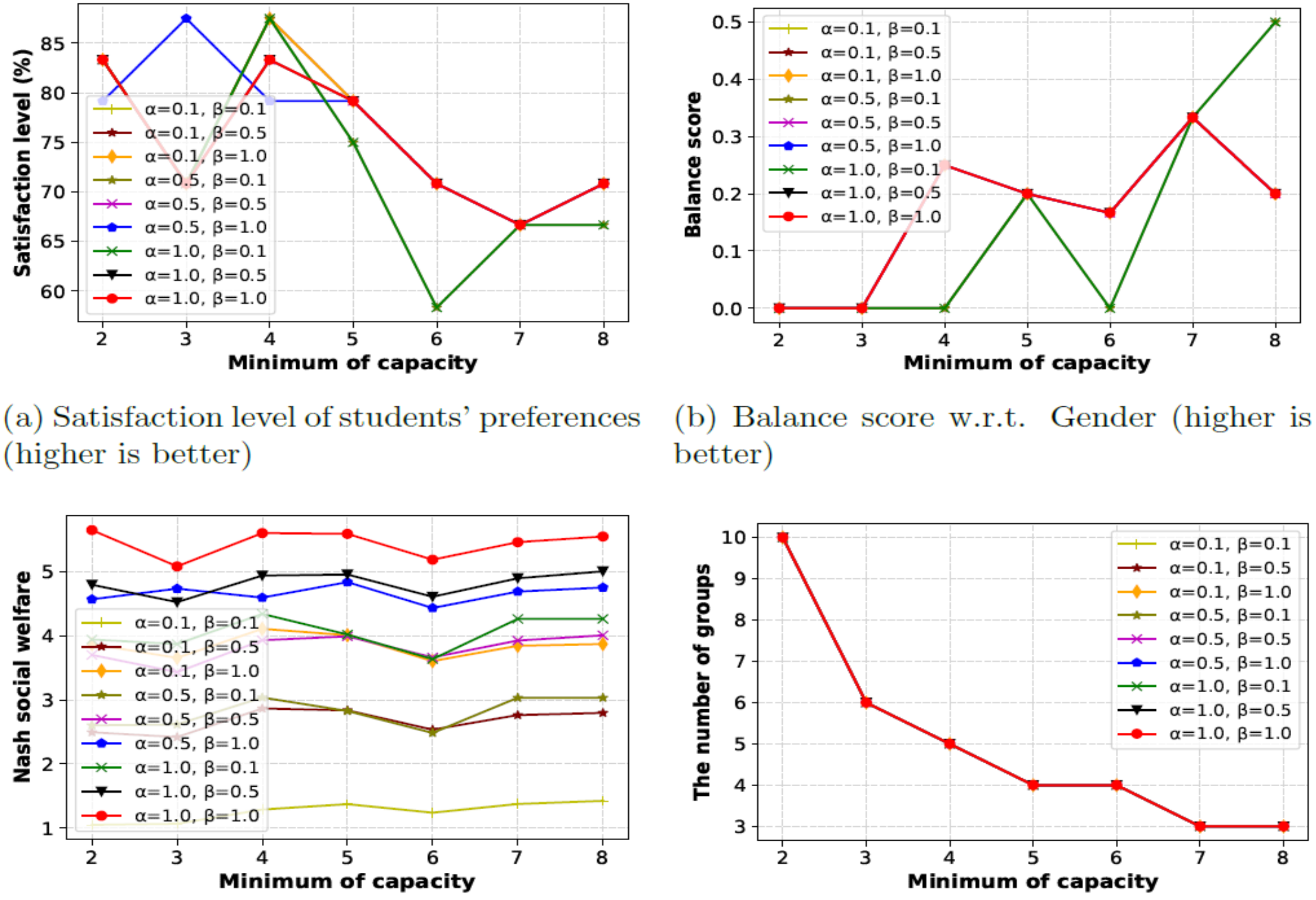
(c) Nash social welfare (higher is better)

(d) Number of groups

Performance of methods on the real data science dataset

Experimental results (cont.)

- In all datasets:
  The knapsack-based model shows the best performance with
  α = 1.0 and β = 1.0



(a) Satisfaction level of students' preferences (higher is better)

(b) Balance score w.r.t. Gender (higher is better)

(c) Nash social welfare (higher is better)

(d) Number of groups

Real data science: Impact of α, β parameters on the knapsack-based model

LernMINT

Conclusions

- A bias-aware analysis of 7 well-known educational datasets
  Bias appears in most datasets w.r.t. protected attributes (gender, race, etc.)
- Evaluation of 7 prevalent group fairness measures in student performance prediction problems
  - Varying behavior of fairness measures across datasets and predictive models
  - Equal opportunity, equalized odds, and ABROCA fairness measures
- We introduce the fair-capacitated clustering problem
  Two approaches: hierarchical-based and k-medoids fair-capacitated approach
- We introduce the MFC students-topics grouping problem
  Three approaches: greedy heuristic, knapsack-based, and MFC knapsack approach

LernMINT

Limitations

- We only consider a single protected attribute
- The experiments are performed on the offline settings
- The computational complexity of the hierarchical-based approach is high ($O(n^3)$)
- It is needed to provide theoretical guarantees for proposed algorithms

Outlook

- Generating synthetic educational datasets
- The issue of fairness on multiple protected attributes
- Evaluating the prevalent fairness measures for fairness-aware clustering models in EDM
- Developing new fairness measures for fair clustering
- Considering the theoretical aspect of the proposed algorithms
- Deploying the implementation of an explainable fair clustering algorithm to achieve the clarification of the resulting clustering

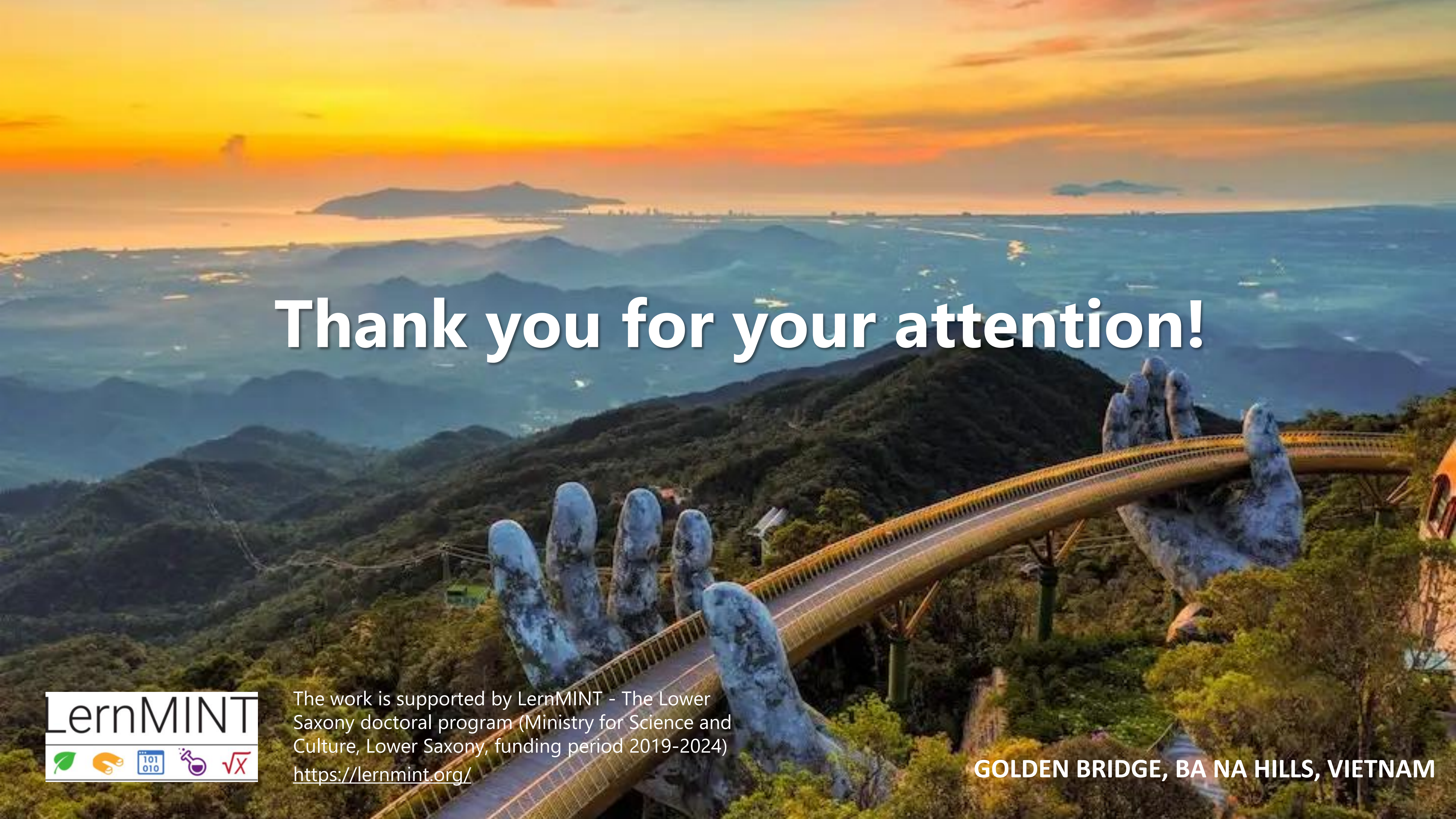
Publications

1. Tai Le Quy, Gunnar Friege, and Eirini Ntoutsi. A review of clustering models in educational data science towards fairness-aware learning. In *Educational Data Science: Essentials, Approaches, and Tendencies – Proactive Education based on Empirical Big Data Evidence*. Springer, 2023.
2. Tai Le Quy, Arjun Roy, Iosifidis Vasileios, Zhang Wenbin, and Eirini Ntoutsi. A survey on datasets for fairness-aware machine learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(3), 2022 (SJR Q1, IF 2021: 7.558. Top cited article of the WIREs Data Mining and Knowledge Discovery journal in 2021-2022).
3. Tai Le Quy, Gunnar Friege, and Eirini Ntoutsi. Multi-fair capacitated students-topics grouping problem. In *Proceedings of the 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (PAKDD 2023), 2023 (rank A CORE).
4. Tai Le Quy, Arjun Roy, Gunnar Friege, and Eirini Ntoutsi. Fair-capacitated clustering. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*, pages 407–414, 2021 (rank B CORE)
5. Tai Le Quy, Thi Huyen Nguyen, Gunnar Friege, and Eirini Ntoutsi. Evaluation of group fairness measures in student performance prediction problems. In *Proceedings of the International Workshops of ECML/PKDD 2022*, pages 119–136. Springer, 2023 (rank A CORE).

LernMINT

Publications (cont.)

6. Huyen Giang Thi Thu, Thuy Nguyen Thanh, and Tai Le Quy. Dynamic sliding window and neighborhood LSTM-based model for stock price prediction. *SN Computer Science*, 3(3):1–14, 2022 (SJR Q2).
7. Huyen Giang Thi Thu, Thuy Nguyen Thanh, and Tai Le Quy. A neighborhood deep neural network model using sliding window for stock price prediction. In *Proceedings of the 2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 69–74. IEEE, 2021 (Web of Science indexed).
8. Bahman Askari, Tai Le Quy, and Eirini Ntoutsi. Taxi demand prediction using an LSTM-based deep sequence model and points of interest. In *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1719–1724. IEEE, 2020 (rank B CORE).
9. Tai Le Quy, Wolfgang Nejdl, Myra Spiliopoulou, and Eirini Ntoutsi. A neighborhood-augmented LSTM model for taxi-passenger demand prediction. In *Proceedings of the International Workshop on Multiple-Aspect Analysis of Semantic Trajectories at ECML/PKDD 2019*, pages 100–116. Springer, Cham, 2019. (rank A CORE).
10. Tai Le Quy, Sergej Zerr, Eirini Ntoutsi, and Wolfgang Nejdl. Data augmentation for dealing with low sampling rates in NILM. In *4th International Workshop on Non-Intrusive Load Monitoring (NILM)*, 2018.
11. Tai Le Quy and Eirini Ntoutsi. Towards fair, explainable and actionable clustering for learning analytics. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*, pages 847–851, 2021 (doctoral consortium section, rank B CORE).



Thank you for your attention!

The work is supported by LernMINT - The Lower Saxony doctoral program (Ministry for Science and Culture, Lower Saxony, funding period 2019-2024)
https://lernmint.org/

GOLDEN BRIDGE, BA NA HILLS, VIETNAM