FAIRNESS-AWARE MACHINE LEARNING IN EDUCATIONAL DATA MINING

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

DOKTOR DER NATURWISSENSCHAFTEN

Dr. rer. nat.

genehmigte Dissertation
von

M. Sc. Tai Le Quy

geboren am 03 September 1985, in Hanam, Vietnam

2024

Referentin: Prof. Dr. Eirini Ntoutsi
Korreferent: Prof. Dr. Gunnar Friege
Vorsitz: Prof. Dr. Johannes Krugel
Tag der Promotion: 16.10.2023

DECLARATION OF AUTHORSHIP

I, Tai Le Quy, declare that this thesis, titled "Fairness-aware Machine Learning in Educational Data Mining" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

ABSTRACT

Fairness is an essential requirement of every educational system, which is reflected in a variety of educational activities. With the extensive use of Artificial Intelligence (AI) and Machine Learning (ML) techniques in education, researchers and educators can analyze educational (big) data and propose new (technical) methods in order to support teachers, students, or administrators of (online) learning systems in the organization of teaching and learning. Educational data mining (EDM) is the result of the application and development of data mining (DM), and ML techniques to deal with educational problems, such as student performance prediction and student grouping. However, ML-based decisions in education can be based on protected attributes, such as race or gender, leading to discrimination of individual students or subgroups of students. Therefore, ensuring fairness in ML models also contributes to equity in educational systems. On the other hand, bias can also appear in the data obtained from learning environments. Hence, bias-aware exploratory educational data analysis is important to support unbiased decision-making in EDM. In this thesis, we address the aforementioned issues and propose methods that mitigate discriminatory outcomes of ML algorithms in EDM tasks. Specifically, we make the following contributions:

- We perform *bias-aware exploratory analysis of educational datasets* using Bayesian networks to identify the relationships among attributes in order to understand bias in the datasets. We focus the exploratory data analysis on features having a direct or indirect relationship with the protected attributes w.r.t. prediction outcomes.
- We perform a *comprehensive evaluation of the sufficiency of various group fairness measures* in predictive models for student performance prediction problems. A variety of experiments on various educational datasets with different fairness measures are performed to provide users with a broad view of unfairness from diverse aspects.
- We deal with the *student grouping problem in collaborative learning.* We introduce the *fair-capacitated clustering* problem that takes into account cluster fairness and cluster cardinalities. We propose two approaches, namely hierarchical clustering and partitioning-based clustering, to obtain fair-capacitated clustering.
- We introduce the *multi-fair capacitated (MFC) students-topics grouping problem* that satisfies students' preferences while ensuring balanced group cardinalities and maximizing the diversity of members regarding the protected attribute. We propose three approaches: a greedy heuristic approach, a knapsack-based approach using vanilla maximal 0-1 knapsack formulation, and an MFC knapsack approach based on group fairness knapsack formulation.

In short, the findings described in this thesis demonstrate the importance of fairness-aware ML in educational settings. We show that bias-aware data analysis, fairness measures, and fairness-aware ML models are essential aspects to ensure fairness in EDM and the educational environment.

Keywords: fairness-aware machine learning, educational data mining, fair clustering, fairness measures, educational dataset, fair-capacitated clustering, multi-fair capacitated.

ZUSAMMENFASSUNG

Fairness ist eine wesentliche Anforderung jedes Bildungssystems in einer Vielzahl von Bildungsaktivitäten. Der breite Einsatz von künstlicher Intelligenz (KI) und maschinelles Lernen (ML) in der Bildung ermöglicht es Forschenden, (große) Bildungsdaten zu analysieren und neue (technische) Methoden vorzuschlagen, um Lehrende oder Administratoren von (Online-)Lernsystemen bei der Organisation des Lehrens und Lernens zu unterstützen. Educational Data Mining (EDM) ist das Ergebnis der Anwendung und Entwicklung von Data Mining (DM) und ML zur Lösung von Bildungsproblemen, wie die Vorhersage von Lernendenleistungen oder die Gruppierung von Lernenden. Entscheidungen im Bildungsbereich, die auf ML basieren, können jedoch auf geschützten Merkmalen wie Rasse oder Geschlecht beruhen, was zur Diskriminierung einzelner Lernenden oder Untergruppen von Lernenden führen kann. Daher trägt die Gewährleistung von Fairness in maschinellen Lernmodellen auch zur Gerechtigkeit in Bildungssystemen bei. Außerdem können Verzerrungen auch in den Daten auftreten, die aus der Lernumgebung gewonnen werden. Daher ist eine auf Verzerrungen ausgerichtete explorative Analyse von Bildungsdaten wichtig, um eine unvoreingenommene Entscheidungsfindung im EDM zu unterstützen. In dieser Arbeit befassen wir uns mit den oben genannten Problemen und schlagen Methoden vor, die die diskriminierenden Ergebnisse von Algorithmen des maschinellen Lernens bei EDM-Aufgaben abschwächen. Im Einzelnen leisten wir die folgenden Beiträge:

- Wir führen eine vorurteilsbewusste explorative Analyse von Bildungsdatensätzen mithilfe von Bayes'schen Netzwerken durchgeführt, um die Beziehungen zwischen den Attributen zu identifizieren und die Verzerrungen in den Datensätzen zu verstehen. Wir konzentrieren uns bei der explorativen Datenanalyse auf Merkmale, die eine direkte oder indirekte Beziehung zu den geschützten Attributen in Bezug auf die Vorhersageergebnisse aufweisen.
- Wir führen eine umfassende Bewertung der Eignung verschiedener Gruppenfairnessmaße in Vorhersagemodellen für Lernendenleistungen durchgeführt. Eine Vielzahl von Experimenten mit verschiedenen Bildungsdatensätzen und unterschiedlichen Fairness-Maßen wird durchgeführt, um den Nutzenden einen breiten Überblick über Unfairness unter verschiedenen Aspekten zu geben.
- Wir befassen uns mit dem Problem der Gruppierung von Lernenden beim kollaborativen Lernen. Wir führen das Problem des fair-capacitated Clustering ein, das die Fairness von Clustern und die Kardinalität von Clustern berücksichtigt. Wir schlagen zwei Vorgehensweisen vor, nämlich hierarchisches Clustering und partitionierungsbasiertes Clustering, um ein fair-capacitated Clustering zu erreichen.
- Wir stellen das multi-fair capacitated (MFC) Studierenden-Themen-Gruppierungsproblem vor, das die Präferenzen der Studierenden berücksichtigt und gleichzeitig ausgewogene Gruppenkardinalitäten sicherstellt und die Diversität der Mitglieder hinsichtlich des geschützten Attributs maximiert. Wir schlagen drei Ansätze vor: einen gierigen heuristischen Ansatz, einen Knapsack-basierten Ansatz, der die Maximal 0-1 Knapsack-Formulierung verwendet, und einen MFC Knapsack-Ansatz, der auf der Gruppenfairness-Knapsack-Formulierung basiert.

Im Allgemeinen zeigen die in dieser Arbeit beschriebenen Ergebnisse die Bedeutung einer fairnessbewussten ML in Bildungsumgebungen. Wir zeigen, dass bias-sensitive Datenanalyse, Fairness-Maßnahmen und fairness-sensitive ML-Modelle wesentliche Aspekte sind, um Fairness im EDM und in der Bildungsumgebung im Allgemeinen zu gewährleisten.

Schlagwörter: Fairness-bewusstes maschinelles Lernen, Data-Mining im Bildungsbereich, faires Clustering, Fairness-Maßnahmen, Bildungsdatensatz, fair-capacitated Clustering, multi-fair capacitated.

ACKNOWLEDGMENTS

Firstly, I would like to express my special gratitude to my supervisor Prof. Dr. Eirini Ntoutsi. She has guided, supported, and encouraged me with all her enthusiasm and helped me overcome difficulties during my Ph.D. throughout the years. I would like to thank Prof. Dr. Gunnar Friege for his great contributions not only to research works but also for the active and effective support from the LernMINT scholarship program. I am also very grateful to Prof. Dr. Wolfgang Nejdl for offering the ideal working environment at the L3S Research Center. I also thank Prof. Dr. Johannes Krugel for agreeing to consider and evaluate my thesis.

Secondly, I sincerely thank Arjun Roy, Nguyen Thi Huyen, Dr. Vasileios Iosifidis, and Dr. Sergej Zerr for their excellent contributions to my work. I would like to thank my office mates and all my colleagues and staff at the L3S Research Center and LernMINT Ph.D. program for making the workplace an exciting atmosphere.

I am very thankful for the financial support from the Ministry of Science and Culture of Lower Saxony (MWK), Germany, and the Ph.D. program "LernMINT: Data-assisted teaching in the MINT subjects". Besides, I would like to thank the 911 scholarship program of the Vietnamese government for supporting me during the first years of my Ph.D.

Last but not least, I would like to thank my mother, Nguyen Thi Nham, my father, Le Quy Tien, with all my family members, and Sebastian Paterski for their unconditional love, support, and tremendous patience.

I dedicate this to you all.

FOREWORD

During my Ph.D. journey, I had the opportunity to work and publish in the areas of unsupervised learning, fairness-aware machine learning, educational data mining, time-series analysis and prediction, and energy disaggregation.

The main contributions of this dissertation are published in the following research papers[1]:

Book chapters

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 [114].

Journal articles

1. 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 [119] (SJR Q1, IF 2021: 7.558. Top cited article of the WIREs Data Mining and Knowledge Discovery journal in 2021-2022).

Conference papers

1. 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) [113].

2. 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[118] (rank B CORE).

[1] The works presented in this thesis are inspired by the project "Dealing with bias and discrimination in learning analytics models", which is a part of the Ph.D. program "LernMINT: Data-assisted teaching in the MINT subjects", supported by the Ministry of Science and Culture of Lower Saxony, Germany (https://lernmint.org/).

Workshop papers

1. 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 [116] (rank A CORE).

In detail, the following chapters are based on the aforementioned publications:

- Chapter 2 presents the basic technical background of EDM and fairness-aware ML [114].
- Chapter 3 performs the bias-aware exploratory analysis of popular educational datasets used in the experiments of the following chapters [119].
- Chapter 4 evaluates the group fairness measures of predictive models in student performance prediction problems [116].
- Chapter 5 proposes the fair-capacitated clustering problem, which deals with the fair clustering and the cardinality constraint in student grouping problem [118].
- Chapter 6 proposes the multi-fair capacitated grouping problem, which deals with the students-topics grouping problem concerning students' preferences [113].

Moreover, I also worked, published, and co-authored several papers in time-series analysis and energy disaggregation:

Journal articles

1. 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 [76] (SJR Q2).

Conference papers

1. 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[75] (Web of Science indexed).

Workshop papers

1. 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 [19] (rank B CORE).

2. 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.[115] (rank A CORE).

3. Tai Le Quy, Sergej Zerr, Eirini Ntoutsi, and Wolfgang Nejdl. Data augmentation for dealing with low sampling rates in NILM. In *NILM*, 2018 [120]. (Accepted)

PhD consortium

1. 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 [117] (rank B CORE).

Documentation of publications including source code, presentation slide, etc., is listed in the appendix **Resources**.

Contents

List of Figures

List of Tables

List of Abbreviations

ABROCA	Absolute Between-ROC Area
Acc	Accuracy
AI	Artificial Intelligence
attr.	Attribute
BA	Balanced accuracy
bin.	Binary
BIRCH	Balanced iterative reducing and clustering using hierarchies
BN	Bayesian network
cat.	Category
CCP	Capacitated clustering problem
CLARA	Clustering in LARge Applications
CLARANS	Clustering Large Applications based on RANdomized Search
DM	Data Mining
DT	Decision Tree
EDM	Educational Data Mining
EDS	Educational Data Science
EO	Equal opportunity
EOd	Equalized odds
ESP	End semester percentage
FN	False Negative
FNR	False negative rate
FP	False Positive
FPR	False positive rate
FYA	First-year average grade
HE	Higher education
ints.	Instance
IPUMS	Integrated Public Use Microdata Series
IR	Imbalance ratio

LMS	Learning management system
LSAC	Law School Admission Council
LSAT	Law school admission test
LSTM	Long short-term memory
MCF	Minimum cost flow
MFC	Multi-fair capacitated
ML	Machine Learning
MLP	Multilayer perceptron
MOOC	Massive open online course
NB	Naïve Bayes
NCES	United States National Center for Education Statistics
non-prot.	Non-protected
num.	Numeric
OU	Open University
OULAD	Open University Learning Analytics
PAM	Partitioning around medoids
PE	Predictive equality
PISA	Program for International Student Assessment
PP	Predictive parity
PPV	Positive predictive value
prot.	Protected
ROC	Receiver Operating Characteristics
RQ	Research question
S.Aca	Student academics
S.Por	Student Portuguese
SP	Statistical parity
SPSS	Statistical Package for the Social Sciences
SVM	Support vector machine
TE	Treatment equality
TN	True Negative
TNR	True negative rates
TP	True Positive
TPR	True positive rate
UCI	University of California Irvine
UGPA	Undergraduate grade point average
VLE	Virtual learning environment
xAPI	Experience API
w.r.t.	with respect to

1 Introduction

1.1 Motivation

Fairness is a fundamental concept of education whereby all students must have an equal opportunity to study or be treated fairly regardless of their socioeconomic status, assets, gender, or race. Fairness in education systems is reflected in a wide range of education-related tasks, such as assessment and measurement of students' performance [58, 212], students' teamwork and group assignments [68, 162], graduate school admission [181], predicting student performance [200]. Education is one of the kernel demands for justice; therefore, having a fair education system is crucial to achieving justice in a society [141].

Educational Data Mining (EDM) "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" [180]. In the EDM community, researchers develop as well as apply Data Mining (DM), Machine Learning (ML), and statistical methods to get insights into the educational data and deal with education problems/tasks. The knowledge and valuable patterns discovered by EDM not only can support teachers and educators in administrating their classes either online or in person but also help learners improve their academic performance. A major benefit of EDM is that it is one of the best ways to analyze educational big data with methods coming from statistics, computer science, ML, DM, and other fields [200].

Three important aspects should be considered in EDM [165], namely objectives, data, and techniques. First, the objectives of EDM encompass both practical research objectives and the pursuit of research goals aimed at optimizing the learning experience, enhancing students' academic performance, and gaining insights into educational phenomena. Second, educational data are collected from different sources, such as educational activities in (online) classrooms and learning management systems (LMS). There is a variety of data types in educational data, as well as semantic information, relationships between different data sources, and a high level of meaningful hierarchy within the data, such as session level, student level, classroom level, and school level [165, 168]. Decisions or treatment may exhibit bias w.r.t. demographic attributes such as gender, race, age, and others, as these attributes encompass sensitive information. Third, the traditional DM and ML techniques can be applied directly or must be adapted due to the specificity of the educational problems. The DM and ML techniques applied in EDM, including but not limited to visualization, classification, clustering, regression, and association rules [156, 168].

In EDM, ML and DM methods have been applied in a wide variety of decision-making, and educational data mining tasks, e.g., student dropout prediction [54, 103], education admission decisions [181], forecasting on-time graduation of students [92, 128], student performance prediction [200]. The results of these ML models are the basis for building applications in EDM, such as student data analysis, learning support, and decision support systems. In fact, different ML-based decisions can be made based on protected attributes (i.e., the attributes for which the model is likely to exhibit bias), such as gender or race, leading to discrimination [151]. Hence, improving fairness w.r.t. the protected attributes in the results of ML models is imperative while maintaining the performance of the models. Put simply, the pursuit of fairness in ML models is directly linked to promoting equity in educational systems. As data are a vital part of ML and benchmark datasets a decisive factor for the success of AI[1], the first aspect of this thesis is to provide an exploratory analysis of educational datasets by using a Bayesian network to identify the relationships among attributes. Based on the Bayesian network, we provide a graphical analysis of the attributes for a deeper understanding of bias in datasets.

Along with the development of the EDM research community, there are more and more studies on ensuring the fairness of ML models, applied in education. These studies mainly focus on supervised learning models on students' data [24, 72, 163]. Recently, there have also been several surveys on algorithmic bias and fairness in education, covering different definitions of fairness and supervised learning models [22, 106]. In EDM, predicting students' academic performance is one of the key tasks; hence, the fairness of the predictive model should be taken into consideration. There are more than 20 different fairness measures introduced in the computer science research area [140, 188]. However, choosing proper measures can be cumbersome due to the dependence of fairness on context. Therefore, the second aspect of this thesis is to evaluate the prevalent group fairness measures in predictive models for student performance prediction problems.

Besides, clustering methods are widely employed in EDM [61] and educational data science (EDS) [138]. The main goal of clustering is to group/cluster instances, i.e., students, into groups of similar students; such as grouping allows for gaining insight, understanding student achievement [127], characterizing students' learning

[1] https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

behaviors [209]. Traditional clustering algorithms, however, focus solely on the similarity objective and do not consider the fairness of the resulting clusters w.r.t. protected attributes like gender or race, while students might learn better in diverse student groups, e.g., mixed-gender groups [77, 208]. Furthermore, conventional approaches fail to account for the cardinality constraint of clusters. Consequently, this may lead to the extraction of clusters with different sizes, diminishing the practicality and relevance of the partitioning for end-users, specifically teachers. Therefore, the third aspect of the thesis focuses on the development of clustering models and grouping models that prioritize fairness concerning both the protected attribute and the cardinality of the resulting clusters or groups. Our investigation centers around the problem of student grouping, considering both scenarios with and without the incorporation of students' preferences.

1.2 Research questions and contributions

In this thesis, we attempt to answer the following research questions (RQ):

- 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?

We tackle the research questions and make four main contributions to the field of fairness-aware ML in EDM. Figure 1.1 illustrates the contributions of the thesis.

First, we provide a *bias-aware exploratory analysis* of educational datasets using *Bayesian networks* to identify the relationships among attributes (RQ_1). Based on the Bayesian network, we provide a graphical analysis of the attributes for a deeper understanding of bias in the dataset. The Bayesian network illustrates the conditional (in)dependence between the protected attribute(s) and the class attribute; thus, it reduces the space and complexity of data analysis that needs to be performed to discover the fairness-related problems in the dataset. We then focus our exploratory analysis on features having a direct or indirect relationship with the protected attributes. In

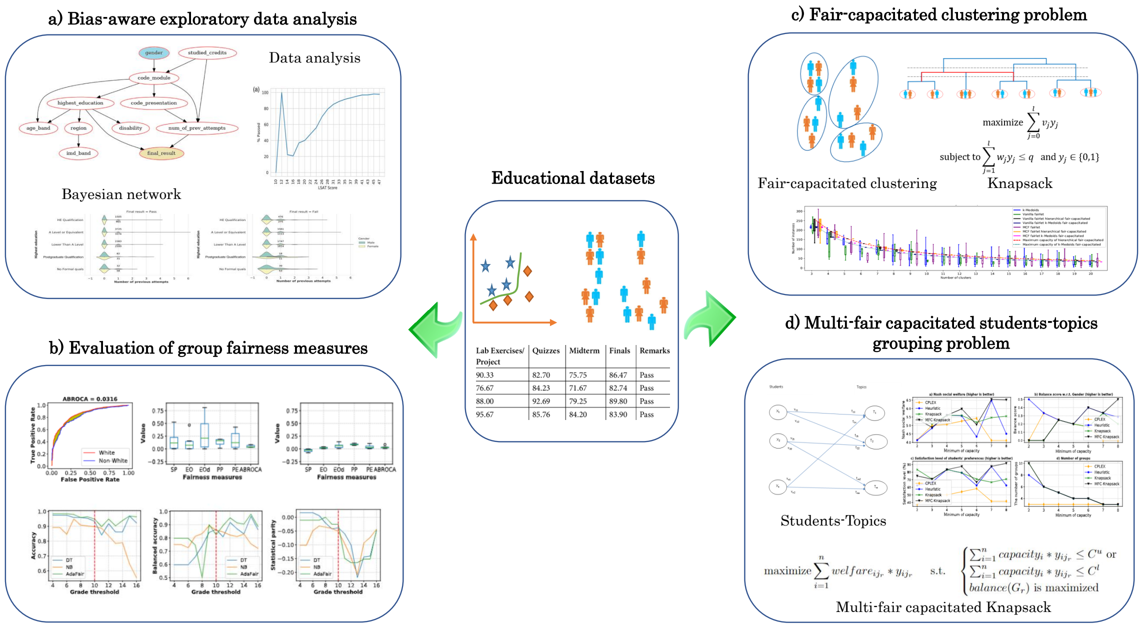


Figure 1.1. An overview of the main contributions

addition to our exploratory analysis, we supplement it with a quantitative evaluation of measures pertaining to both predictive performance and fairness performance (Figure 1.1-a).

Second, we provide a comprehensive study to *evaluate* the sufficiency of various group *fairness measures* in predictive models for student performance prediction problems (RQ_2). We conduct a variety of experiments on diverse educational datasets and evaluate them with different fairness measures. Our experiments provide users with a broad view of unfairness in educational contexts from diverse aspects. Besides, the results also guide the selection of suitable fairness measures to evaluate students' performance predictive models (Figure 1.1-b).

Third, we introduce the *fair-capacitated clustering problem* that partitions the data into clusters of similar instances while ensuring cluster fairness and balancing cluster cardinalities (RQ_3). We propose a two-step solution to the problem: i) we rely on fairlets to generate minimal sets that satisfy the fair constraint, and ii) we propose two approaches, namely hierarchical clustering and partitioning-based clustering, to obtain the fair-capacitated clustering. The hierarchical approach embeds the additional cardinality requirements during the merging step while the partitioning-based one alters the assignment step using a knapsack problem formulation to satisfy the additional requirements. Our experiments on four educational datasets show that our approaches deliver well-balanced clusters in terms of both fairness and cardinality while maintaining a good clustering quality (Figure 1.1-c).

Fourth, we introduce the *multi-fair capacitated* (MFC) grouping problem that fairly partitions students into non-overlapping groups while ensuring balanced group cardinalities (with a lower and an upper bound) and maximizing the diversity of members regarding the protected attribute (RQ_4). To obtain the MFC grouping, we propose three approaches: a greedy heuristic approach, a knapsack-based approach using vanilla maximal knapsack formulation, and an MFC knapsack approach based on group fairness knapsack formulation. Experimental results on a real dataset and a semi-synthetic dataset show that our proposed methods can satisfy students' preferences and deliver balanced and diverse groups regarding cardinality and the protected attribute, respectively (Figure d1.1-d).

1.3 Outline of the thesis

The remaining chapters of this thesis are organized as follows.

In Chapter 2, we provide the general technical background that is essential to achieve the goals conducted in this thesis. In particular, we first summarize the prevalent tasks in EDM, and then we outline the principle of fairness-aware ML w.r.t. both supervised learning and unsupervised learning. We also present detailed descriptions of vanilla clustering models and the knapsack problem, which are employed in the next chapters. The chapter is partially based on the publication:

- 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 [114].

Chapter 3 presents a bias-aware exploratory analysis of educational datasets which are further studied in the experiments of the next chapters. We provide a detailed illustration of the Bayesian network and graphical data analysis. The chapter is based on the publication:

- 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 Discovevry*, 12(3), 2022 [119].

In Chapter 4, we focus on student performance prediction problems, where we evaluate group fairness measures through comprehensive experiments. These experiments involve traditional predictive models as well as fairness-aware models, conducted on diverse educational datasets. Our experimental results show that the choice of the fairness measure is important, likewise for the choice of the grade threshold which determines whether a candidate passes or fails an exam. This chapter is based on the publication:

- 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 [116].

Chapter 5 investigates the student grouping problem. We introduce the fair-capacitated problem, which considers fairness in terms of the protected attribute and cluster cardinalities. We provide the experiments of our hierarchical and knapsack-based approaches on various educational datasets. The chapter is based on the paper:

- Tai Le Quy, Arjun Roy, Gunnar Friege, and Eirini Ntoutsi. Fair-capacitated clustering. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM21)*, pages 407–414, 2021 [118].

In Chapter 6, we focus on the students-topics grouping problem. We propose the multi-fair capacitated (MFC) grouping problem. The primary objective of this problem is to ensure fairness in terms of student satisfaction, the protected attribute, and groups' cardinality by taking into account students' preferences. We present the experimental results of three methods (greedy heuristic, a knapsack-based, and an MFC knapsack approach) on several real and semi-synthetic educational datasets. This chapter is based on the publication:

- 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)*. Springer, 2023 [113].

Finally, we conclude our work, emphasize the main contributions and discuss potential directions for future research in Chapter 7.

2 Background

In this chapter, we present the basic technical background in order to understand the next chapters of the thesis. We commence by showcasing the EDM tasks that can be effectively addressed through the application of DM/ML techniques. Then, we overview the fairness-aware ML models on both supervised learning and unsupervised learning approaches. Subsequently, we describe two vanilla clustering models and the knapsack problem, which have been modified and employed in our work in the later chapters.

2.1 Educational data mining tasks

Many EDM tasks in educational environments can be effectively addressed by DM/ML techniques. In the recent survey and review articles on EDM, researchers categorized EDM by different methods and properties [23, 61, 156, 165, 168]. Because EDM is the "application of data mining (DM) techniques to this specific type of dataset that comes from educational environments to address important educational questions" [166], therefore, first, we categorize and present seven groups of EDM tasks based on the DM/ML tasks and educational activities. For each group of EDM tasks, we divide it into several tasks. Figure 2.1 illustrates our proposed taxonomy of EDM tasks by summarizing the taxonomies introduced in the work of [23, 156, 165]. Next, we describe 17 EDM tasks in terms of their objective as well as the traditional statistics and DM/ML methods applied to such tasks in Table 2.1. Student data analysis, prediction and supporting learning and teaching activities are the most prevalent tasks in EDM [165]; therefore, in this thesis, we focus on three important EDM tasks, namely data analysis, student performance prediction, and student grouping.[1].

[1] We highlight considered tasks with rectangles with the yellow background in Figure 2.1.

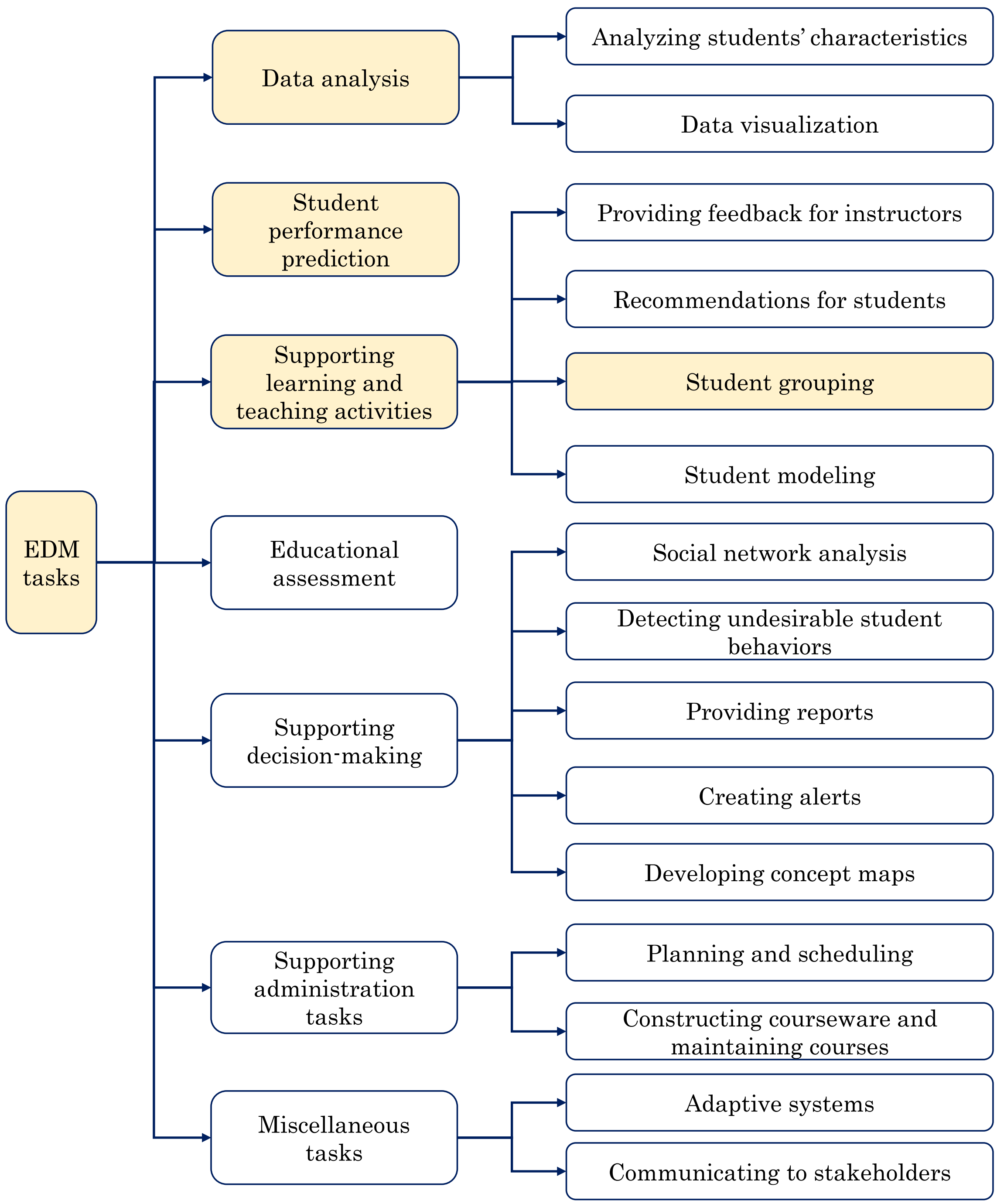


Figure 2.1. A taxonomy of EDM tasks

Table 2.1. A description of EDM tasks

Task	**Objective**	**Methods**
Analyzing students' characteristics	Discovering the patterns of students and highlighting valuable information [55, 56, 198].	Statistics, clustering
Data visualization	Presenting data in an easy-to-understand manner [34, 216].	Visualization tools
Student performance prediction	Predicting student grades, test scores, students' marks (pass or fail), the dropout phenomenon, etc. [167, 200].	Correlation analysis, Regression, classification, clustering, ensemble learning, deep learning
Providing feedback for instructors	Providing feedback to support teachers about students' progress, instructional resources, etc. [135, 177].	Association rules, clustering, classification
Recommendations for students	Providing personalized course and activity recommendations to students based on their individual characteristics [126, 207].	Association rules, clustering, classification
Grouping students	Creating clusters/groups of students based on their characteristics, preferences, etc., for collaboration work or for building personalized learning systems [158, 191, 199].	Clustering, classification
Student modeling	Developing cognitive models of students, encompassing the modeling of their knowledge, abilities, and skills [124, 125].	Association rules, regression, classification, clustering
Educational assessment	Assisting the educators in student evaluation/assessment by automatic evaluators [164, 179].	Clustering, classification, text mining
Social network analysis	Investigating relationships among students instead of individual characteristics [160, 161].	Visualization tools, clustering
Detecting undesirable student behaviors	Detecting students who have unusual behaviors such as erroneous actions, low motivation, cheating, etc. [32, 131].	Classification, clustering, association rules

Table 2.1 A description of EDM tasks (continued)

Task	Objective	Methods
Providing reports	Providing information that assists educators and administrators in their decision-making [170].	Clustering, visualization tools
Creating alerts	Providing real-time information or alerts to stakeholders [107].	Classification, clustering
Developing concept maps	Helping instructors/educators to process the construction of concept maps (a hierarchical structure of relationships between concepts) automatically [5].	Association rules, text mining
Planning and scheduling	Planning the courses and resource allocation, supporting the admission and counseling procedures, and developing curriculum [88, 190].	Association rules, classification, clustering
Constructing courseware and maintaining courses.	Automating the construction and development of learning content and courseware [71].	Classification, clustering, text mining
Adaptive systems	Personalizing systems to adapt to the students' behaviors/preferences [10].	Text mining
Communicating to stakeholders	Maintaining effective communication and collaboration among stakeholders [169].	Clustering, classification

2.1.1 Data analysis

Learner behavior can be analyzed and summarized using descriptive analytics. Researchers used statistical software, such as SPSS[2], to extract fundamental descriptive statistics information from educational datasets [165]. Because clustering is an important technique for analyzing student data, we perform a systematic literature review on clustering algorithms and their applicability and usability in EDM by considering 133 publications from January 2017 to June 2022 [114]. Consequently, the student data is investigated for the following tasks:

Analyzing students' behavior, interaction, engagement, motivation and emotion. Students' behavior is analyzed by using clustering techniques to discover the relationship between their behaviors and learning performance [55]. As a result,

[2]https://www.ibm.com/products/spss-statistics

managers can gain a thorough understanding of students and effectively manage them [56]. The students who participated more increased their interest in the course and achieved higher grades [63]. Therefore, the teacher can apply early interventions to prevent students from failing. Furthermore, student data are collected from various sources, from traditional classrooms to LMS, such as Moodle[3], and student behavior can change over time. Hence, it is necessary to investigate the behavior of students over time. In the massive open online course (MOOC) environment, students' motivation is essential [82]. In addition, teachers should also pay attention to students' emotions [192] because there is a relationship between emotions and performance [18] as well as emotions and motivation of students [146]. Students can obtain many emotional experiences through physical education activities, which has an impact on improving learning interest and relationships among teachers and students [79].

Analyzing student's performance. Students are often divided into groups based on their performance, and this helps to identify low-performing students so that educators can help them improve their performance [185, 189]. In addition, analyzing student learning outcomes is one of the effective methods to understand the factors affecting students' learning ability [28]. For example, procrastination is identified as an important indicator of students' performance, whereas non-procrastinators tend to have a higher performance [87].

Analyzing physical and mental health. Health and well-being are determined by an individual's physical condition, and their potential to contribute to society [201]. Based on these findings, colleges and universities can enhance their education and teaching plans, as well as refine their talent training programs, utilizing the results of the study. In terms of mental health analysis, psychological fitness is analyzed using the Fuzzy c-means algorithm [123]. As a result, the findings from these studies may support school counselors and student managers in providing better mental health services for students.

2.1.2 Student performance prediction

Student performance prediction (including test scores, students at-risk, dropout, etc.) is one of the most important tasks of EDM [200]. Various EDM techniques, such as correlation, regression, and classification, have been applied to predict students' performance [167]. In order to identify students at risk of academic failure early in the learning process, we can predict their performance and provide them with guidance. Researchers have tried to achieve various objectives, which can be grouped broadly into three categories: i) Predicting the performance of students in terms of the actual grade (considered as a regression problem); ii) Considering student performance prediction as a classification problem; iii) Predicting student at-risk or dropout. In practice, most studies use classification algorithms to predict students' performance, in which decision trees (DTs), Naïve Bayes (NB), and multilayer perceptron (MLP)

[3]https://moodle.org/

are the prevalent predictive models [200]. In addition, researchers also used clustering algorithms to pre-process the student data before applying the predictive models [85, 159]. Recently, there have been many studies on predicting student dropout and at-risk by applying clustering methods because student dropout is the main concern of many educational institutes [104], especially in the MOOC environment.

2.1.3 Student grouping

Collaborative work is a widely practiced activity in educational environments, and it is an essential factor in improving students' engagement in the classroom [64]. In the traditional classroom, students are grouped into homogeneous and heterogeneous groups based on their knowledge levels to capture comprehensive semantic information about the group [199]. Pratiwi et al. [158] proposed a clustering method to generate heterogeneous groups automatically based on dissimilarity between students. The resulting clusters are comparative with the teachers' manual grouping solutions. In the MOOC systems, the problem of grouping students w.r.t. preferences and interests was introduced by Akbar et al.[9]. They introduced a grouping method based on hierarchical k-means clustering and a weighting formula to satisfy the students' interests and ensure the division of students into teams with similar preferences. In addition, Wang et al. [191] developed their clustering model based on the enhanced particle swarm optimization algorithm to group students w.r.t. their knowledge state and interests. Moreover, classification models also have been applied to student grouping problems [184] with various methods (neural networks, random forests, and DTs).

2.2 Fairness-aware supervised learning

In this section, we provide an overview of fairness-aware supervised learning tasks with prevalent fairness notions (Section 2.2.2) and a taxonomy of fairness-aware classification models (Section 2.2.3).

2.2.1 Preliminary

First, we summarize the definition of important terms used in this thesis. According to Cambridge dictionary[4], *bias* is "the action of supporting or opposing a particular person or thing in an unfair way, because of allowing personal opinions to influence your judgment" and *discrimination* is "treating a person or particular group of people differently, especially in a worse way from the way in which you treat other people, because of their race, gender, sexuality, etc.". "*Algorithmic bias* describes systematic and repeatable errors in a computer system that create unfair outcomes, such

[4] https://dictionary.cambridge.org/

as privileging one arbitrary group of users over others"[5]. "*Protected attributes* are those qualities, traits or characteristics that, by law, cannot be discriminated against. Protected attributes include age, breastfeeding, gender identity, disability, lawful sexual activity, marital status, parental or carer status, pregnancy, physical features, race, religious belief, sex, sex characteristics, sexual orientation, industrial activity, employment activity, political belief or activity"[6].

Second, we describe the common notation used in this thesis in Table 2.2. These symbols are used for both supervised and unsupervised models. Throughout the thesis, other symbols will be introduced as required.

Table 2.2. Summary of notations

Symbol	**Description**
X	A dataset $X = \{x_1, x_2, \ldots, x_n\}$
x or x_i	A data point
n	The number of data points
d	The number of attributes (dimensions)
Y	The class attribute $Y = \{+, -\}$
$\hat{Y}$	The predicted outcome $\hat{Y} = \{+, -\}$
$\mathcal{P}$	The protected attribute $\mathcal{P} = \{p, \overline{p}\}$

In this thesis, we consider the binary classification problem. In Table 2.2, $\mathcal{P} = \{p, \overline{p}\}$ is the binary protected attribute, in which p is the discriminated group (referred to as *protected group*), and $\overline{p}$ is the non-discriminated group (referred to as *non-protected group*). For example, $\mathcal{P}$ = "gender" $\in \{female, male\}$ is the protected attribute; p = "female" could be the protected group, and $\overline{p}$ = "male" could be the non-protected group. We use the notation p_+ (p_-), $\overline{p}_+$ ($\overline{p}_-$) to denote the protected and non-protected groups for the positive (negative, respectively) class (Figure 2.2). We refer to the positive class as the target class, e.g., pass the exam. The goal of fairness-aware supervised learning is to find a map function $f : X \mapsto Y$ that minimizes the loss and mitigates the discriminatory outcomes simultaneously.

[5] https://guides.lib.fsu.edu/algorithm

[6] https://federation.edu.au/current-students/assistance-support-and-services/equity-and-inclusion/protected-attributes

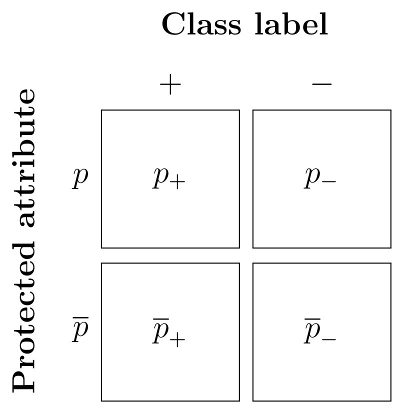


Figure 2.2. Different groups separated by the protected attribute and class label

2.2.2 Fairness notions for supervised learning

Measuring bias in ML models comprises the first step to bias elimination. Fairness depends on context; thus, a large variety of fairness measures[7] exists. Only in the computer science research area, more than 20 fairness measures have been introduced thus far [188, 214]. Nevertheless, there is no universally suitable fairness measure; instead, the notion of fairness depends on various factors such as context, application, space, time, etc.[69, 188]. In this section, we present the most frequent fairness notions, which will be experimented with in Chapter 3.

Statistical parity

Statistical parity (denoted as SP) is a well-known group fairness measure [62] whereby the output of any classifier satisfies statistical parity if the difference (bias) in the predicted outcome ($\hat{Y}$) between any two groups under study (i.e., p and $\overline{p}$) is up to a predefined tolerance threshold ε. Formally:

$$P(\hat{Y}|\mathcal{P}=p) - P(\hat{Y}|\mathcal{P}=\overline{p}) \leq \varepsilon \tag{2.1}$$

We use the violation of statistical parity [178, 213] to measure the bias of a classifier:

$$SP = P(\hat{Y}=+|\mathcal{P}=\overline{p}) - P(\hat{Y}=+|\mathcal{P}=p) \tag{2.2}$$

The value domain is $SP \in [-1, 1]$, with $SP = 0$ standing for no discrimination, $SP \in (0, 1]$ indicating that the protected group is discriminated, and $SP \in [-1, 0)$ meaning

[7] The fairness notion may be turned into measures by taking a difference or a ratio of the equation components [214]. Therefore, in this thesis, we use the terms "fairness notion" and "fairness measure" interchangeably.

that the non-protected group is discriminated (*reverse discrimination*). However, SP only necessitates a balanced representation of both groups in relation to the positive class, without ensuring that the selected instances are qualified or meet the necessary criteria [94].

Equalized odds

A predicted outcome $\hat{Y}$ is satisfied *equalized odds* (denoted as EOd) w.r.t. the protected attribute $\mathcal{P}$ and class label Y, if "$\hat{Y}$ and $\mathcal{P}$ are independent conditional on Y" [81].

$$P(\hat{Y} = +|\mathcal{P} = p, Y = y) = P(\hat{Y} = +|\mathcal{P} = \overline{p}, Y = y), \qquad y \in \{+, -\} \tag{2.3}$$

Therefore, we can measure the equalized odds as the following [94]:

$$EOd = \sum_{y \in \{+,-\}} |P(\hat{Y} = +|\mathcal{P} = p, Y = y) - P(\hat{Y} = +|\mathcal{P} = \overline{p}, Y = y)| \tag{2.4}$$

The value domain is $EOd \in [0, 2]$, with 0 standing for no discrimination (only when the Receiver Operating Characteristics - ROC curves of the two groups intersect) and 2 indicating the maximum discrimination. However, EOd measures primarily concentrate on achieving equal prediction outcomes between different groups, neglecting other dimensions of fairness such as individual treatment or disparate impact.

Absolute Between-ROC Area

Absolute Between-ROC Area (ABROCA) [72] is the first measure that can be represented visually, based on the ROC curve. ABROCA measures the divergence between the protected (ROC_p) and non-protected group ($ROC_{\overline{p}}$) curves across all possible thresholds $t \in [0, 1]$ of false positive rate (FPR) and true positive rate (TPR). The absolute difference between the two curves is measured to capture the case that the curves may cross each other. At the points where the two models intersect, they achieve equalized odds between the protected and non-protected groups.

$$\int_0^1 \mid ROC_p(t) - ROC_{\overline{p}}(t) \mid \, dt. \tag{2.5}$$

The value range is $ABROCA \in [0, 1]$. The lower value indicates a lower difference in the predictions between the two groups and, therefore, a fairer model. Figure 2.3 presents the ABROCA value of the SVM classifier on the Law school dataset[8]. The protected attribute is *Race* = {*White, Non-White*}. The predictive model is not fair because the ABROCA is quite high (0.0833).

[8]The Law school dataset is described in Section 3.3.5

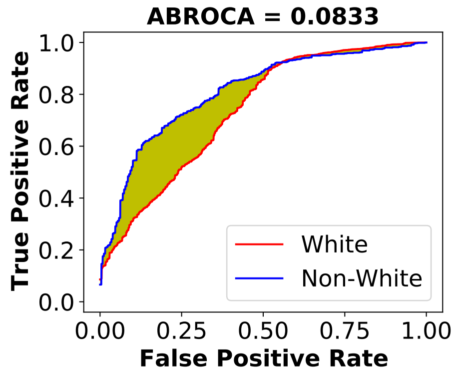


Figure 2.3. ABROCA measure on the Law school dataset based on an SVM classifier

2.2.3 Fairness-aware classification models

There are three approaches to mitigate bias in classification models: i) pre-processing approach, ii) in-processing approach, and iii) post-processing approach [140, 151].

Pre-processing approaches

In this direction, researchers focus on the data, the primary source of bias. They aim to generate a "balanced" dataset, i.e., a dataset that exists without one or other kind of bias, and then apply any ML algorithms to that. For example, the class labels could be altered [99], different weights could be assigned to instances [35], or the protected and non-protected groups could be balanced in the training set based on carefully chosen samples from each group. Recently, a method of augmenting the protected group via semi-synthetically generated instances that reside near the decision boundary of a classifier was introduced by Iosifidis and Ntoutsi [93]. Besides, Calmon et al. [37] proposed a method to change the instances to make the class label dependent on the protected attribute by using a probabilistic framework.

In-processing approaches

In-processing approaches reformulate the classification problem by explicitly incorporating the model's discrimination behavior in the objective function through regularization or constraints or by training on latent target labels [151]. This approach involves incorporating a model's discrimination behavior into the objective function by regularizing or constraining it. Zafar et al. [206] proposed an approach based on a constraint-based approach that can be incorporated into logistic regression and SVMs to handle disparate mistreatment. According to Agarwal's method [4], a fair classification can be reduced to a series of cost-sensitive classification problems with the lowest (empirical) error under the desired constraints. AdaFair, a sequential fair ensemble, was proposed by Iosifidis and Ntoutsi [94] that extends AdaBoost's weighted distribution approach by taking into account the cumulative fairness of the learner up until the current boosting round and accounts for class imbalance by optimizing for balanced error instead of an overall error.

Post-processing approaches

Unlike the above two approaches, post-processing approaches post-process the classification models once they have been learned from the data. There are two main methods, including altering the model's internal (white-box approaches) or its predictions (black-box approaches) [151]. An example of a white-box approach is the research of Calders et al. [36] where they tried to correct the probabilities in NB models to address fairness. But recent years have not seen any further development of white-box approaches. Kamiran et al. [100] used a black-box approach to proportionally promote or demote predictions near the decision boundary between protected and non-protected groups. The post-processing approach is also followed in fairness-aware unsupervised learning models which will be introduced in Section 2.4.2.

2.3 Clustering models

In this section, we introduce the clustering models which will be employed in the following chapters.

The goal of clustering is to group objects, e.g., students, into clusters where the objects in the same cluster are similar and the objects in different clusters are different. The clustering methods can be categorized into several main approaches, including i) partitioning-based approaches, ii) hierarchical-based approaches, iii) density-based approaches, iv) grid-based approaches, v) model-based approaches, and vi) constraint-based approaches.

In the next sections, we present two popular clustering models used in our work: i) hierarchical clustering (hierarchical-based approach) and ii) k-medoids (partitioning-based approach). Moreover, we overview the capacitated clustering problem (CCP). We extend the list of notions described in Table 2.2 with more symbols used for unsupervised learning, presented in Table 2.3. Because the objects or data points for clustering in the related work are mainly students, we use the terms "'objects", "data points" and "students" interchangeably in this thesis.

2.3.1 Hierarchical clustering

Hierarchical clustering creates a set of nested clusters organized as a hierarchical tree which can be visualized as a dendrogram [154]. An example of hierarchical clustering with a dendrogram on the student performance dataset[9] is visualized in Figure 2.4.

[9]The student performance dataset is described in Section 3.3.1

Table 2.3. Summary of notations in unsupervised learning

Symbol	Description
k	The number of clusters
$\mathcal{C}$	A hard clustering $\mathcal{C} = \{C_1, C_2, ..., C_k\}$
S	A set of cluster centers $S = \{s_1, s_2, ..., s_k\}$
x or x_i	A data point, $1 \leq i \leq n$
C_j	A cluster, $1 \leq j \leq k, C_j \in \mathcal{C}$
s or s_j	The cluster center (centroid, medoid)
$dist(,)$	A distance function
$\mathcal{L}(\mathcal{C}, X)$	An objective function of clustering $\mathcal{C}$ on dataset X

There are two main types of hierarchical clustering: *agglomerative* (bottom-up approaches) and *divisive* (top-down approaches). Hierarchical agglomerative clustering is a popular approach applied in related work [114]. BIRCH [210] - a version of hierarchical agglomerative clustering which is suitable for very large databases, is used by Dovgan et al. [60] for their analysis. In this section, we present the basic hierarchical agglomerative clustering method [101]. There are many approaches for computing the proximity of two clusters (linkage algorithms): single linkage, complete linkage, average linkage, centroid-distance, Ward's method [193], etc.

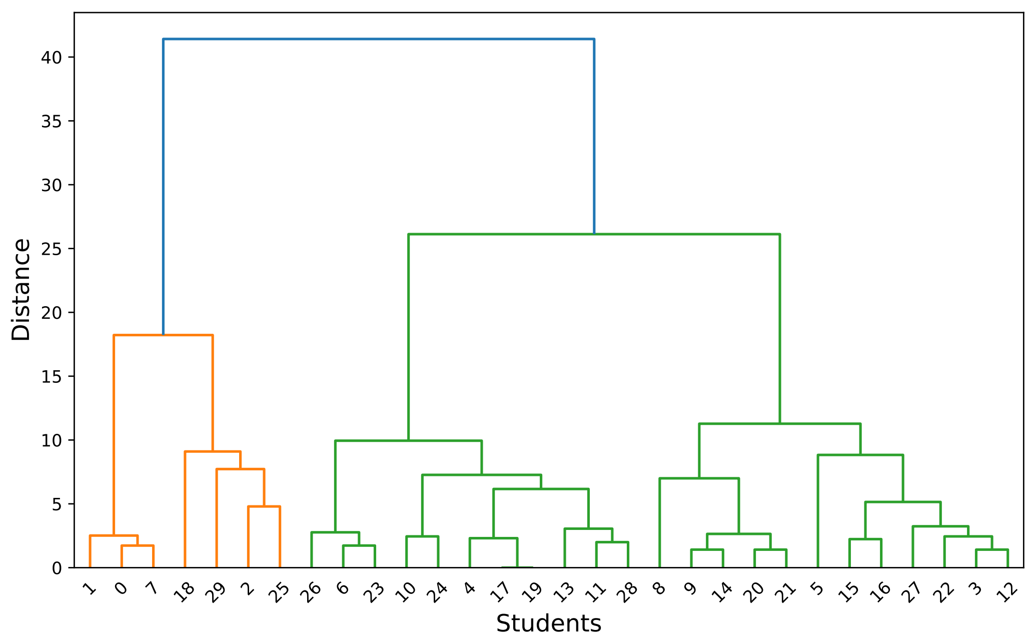


Figure 2.4. An example of hierarchical clustering on the first 30 instances of the student performance dataset. Colors indicate the possible clusters that can be formed based on the color threshold parameter, i.e., the cut threshold w.r.t. distance. In this example, *color threshold = 30*

Algorithm 1 describes the basic steps of hierarchical agglomerative clustering [3].

Algorithm 1: Hierarchical agglomerative clustering

Input: A dataset $X = \{x_1, x_2, \ldots, x_n\}$
Output: A tree $\mathcal{T}$

$\mathcal{A} \leftarrow \emptyset$ //Active set starts out empty;
for $i \leftarrow 1$ **to** n **do**
  $\mathcal{A} \leftarrow \mathcal{A} \cup \{\{x_i\}\}$ //Ech data point is consider as a cluster;
end
$\mathcal{T} \leftarrow \mathcal{A}$;
while $|\mathcal{A}| > 1$ **do**
  $\mathcal{G}_1^*, \mathcal{G}_1^* \leftarrow \text{argmin}_{\mathcal{G}_1, \mathcal{G}_1 \in \mathcal{A}}\, dist(\mathcal{G}_1, \mathcal{G}_2)$ //Two closest clusters;
  $\mathcal{A} \leftarrow (\mathcal{A} \backslash \{\mathcal{G}_1^*\}) \backslash \{\mathcal{G}_2^*\}$ //Remove each from active set;
  $\mathcal{A} \leftarrow \mathcal{A} \cup \{\mathcal{G}_1^* \cup \mathcal{G}_2^*\}$ //Add union to active set ;
  $\mathcal{T} \leftarrow \mathcal{T} \cup \{\mathcal{G}_1^* \cup \mathcal{G}_2^*\}$ //Add union to tree ;
end
return *Tree* $\mathcal{T}$;

Complexity: The time complexity of the basic agglomerative algorithm is $\mathcal{O}(n^3)$. The complexity can be reduced to $\mathcal{O}(n^2 \log n)$ by using the heap data structure.

2.3.2 k-medoids

k-medoids was introduced by Kaufman and Rousseeuw with the PAM (partitioning around medoids) algorithm [101]. k-medoids aim to partition n objects of dataset $X = \{x_1, x_2, \ldots, x_n\}$ into k clusters. The number of clusters k is given in advance. A clustering $\mathcal{C}$ is a partition of dataset X into k disjoint subsets, $\mathcal{C} = \{C_1, C_2, ..., C_k\}$, called *clusters* with $S = \{s_1, s_2, ..., s_k\}$ be the corresponding cluster centers. k-medoids selects the actual data points as cluster centers (denoted *medoids*). The goal of k-medoids is to minimize the clustering cost (Eq.2.6).

$$\mathcal{L}(\mathcal{C}, X) = \sum_{j=1}^{k} \sum_{x \in C_j} dist(x, s_j) \tag{2.6}$$

where $dist(x, s_j)$ is the distance from any point $x \in C_j$ to its medoid s_j. The PAM approach using a greedy search strategy is presented in Algorithm 2.

Figure 2.5 illustrates an example of k-medoids clustering on the student performance dataset. The medoids are marked by orange circles, while members of the two clusters are marked by red and dark blue colors.

Algorithm 2: k-medoids clustering

Input: A dataset $X = \{x_1, x_2, \ldots, x_n\}$
k: number of clusters
Output: A clustering

```
medoids ← select k of the data points arbitrarily ;
Assign each data point to the closest medoid ;
cost_best ← current clustering cost;
s_best ← null ;
o_best ← null ;
repeat
    for each medoid s in medoids do
        for each non-medoid o in X do
            consider the swap of s and o, compute the current clustering cost;
            if current clustering cost < cost_best then
                s_best ← s;
                o_best ← o;
                cost_best ← current clustering cost;
            end
        end
    end
    Update medoids by the swap of s_best and o_best ;
    Assign each data point to the closest medoid ;
until no improvements can be achieved by any replacement;
return clusters;
```

Complexity: The computation complexity of the original PAM algorithm per iteration is $\mathcal{O}(k(n-k)^2)$. The complexity can be reduced to $\mathcal{O}(n^2)$ with several improved algorithms, such as CLARA and CLARANS [175].

2.3.3 The capacitated clustering problem

The capacitated clustering problem (CCP) was first introduced by Mulvey and Beck [145]. The idea of CCP is to find a hard clustering $\mathcal{C}$ with a given capacity constraint, i.e., capacitated clusters (each cluster with a given capacity), that minimizes an objective function. The original formulation of the problem is stated as follows.

$$\text{minimize } \mathcal{L}(\mathcal{C}, X) = \sum_{x \in X} \sum_{s \in S} dist(x, s)\pi_{xs} \tag{2.7}$$

subject to

$$\sum_{s \in S} \rho_s = k \tag{2.8}$$

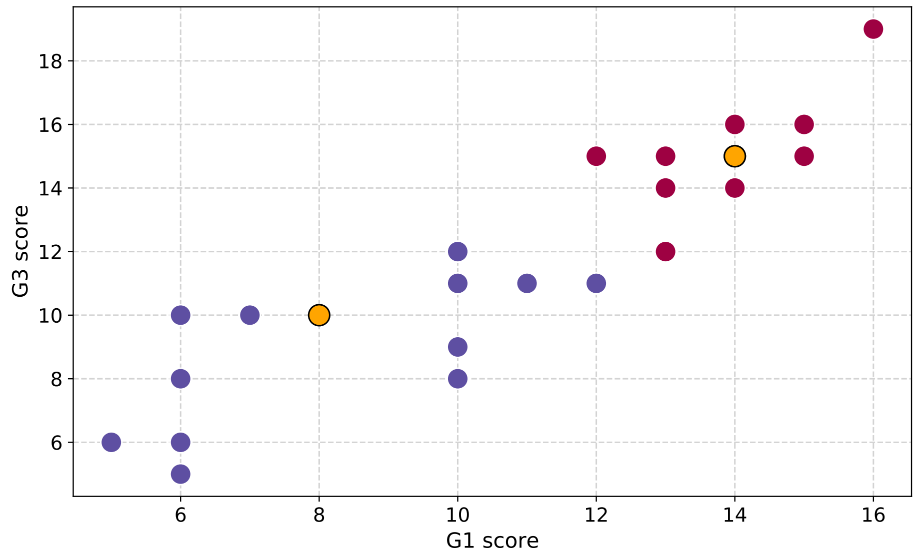


Figure 2.5. An example of k-medoids clustering on the first 30 instances of the student performance dataset

$$\sum_{s \in S} \pi_{xs} = 1, x \in S \tag{2.9}$$

$$\pi_{xs} \leq \rho_s, x \in X, s \in S \tag{2.10}$$

$$\pi_{xs}, \rho_s = \{0, 1\}, x \in X, s \in S \tag{2.11}$$

$$\sum_{x \in X} w_x \pi_{xs} \leq W_s, s \in S \tag{2.12}$$

where $|X| = n, |S| = m, m \geq k$, m is the number of median candidates, and k is the number of clusters. π_{xs} is a binary variable indicating whether or not point x is assigned to cluster s, if so $\pi_{xs} = 1$, otherwise $\pi_{xs} = 0$; variable $\rho_s = 1$ indicates point s as the median for cluster $s \in S$ (Eq. 2.11). Variable w_x specifies the weight of the data point x.

The goal of CCP is to find a clustering that minimizes the objective function (Eq. 2.7 and satisfies the constraints: i) the number of clusters is k (Eq. 2.8), ii) all points are assigned (Eq. 2.9), iii) a point is assigned to one median (Eq. 2.10), iv) the capacity of cluster $s \in S$, i.e., its cumulative weight, is less than the given maximum weight W_s (Eq. 2.12).

2.4 Fairness-aware unsupervised learning

This section summarizes the popular fairness notions for clustering and well-known fair clustering models.

2.4.1 Fairness notions in unsupervised learning

In general, the fairness notions depend on the application and specific context. There are 20 fairness notions used for fair clustering summarized in a survey of fairness in clustering [42]. In their review, the fairness notions are categorized into 4 types: *group-level, individual-level, algorithm agnostic*, and *algorithm specific*. Based on the popularity of the fairness notions [42], we summarize the following four fairness notions: *balance, bounded representation, social fairness cost, individual fairness.* These fairness notions are applied in hard clustering,i.e., each data point is only assigned to one cluster. We continue using the symbols listed in Table 2.2 and Table 2.3.

Balance

Balance (or balance score) is the most popular group-level fairness notion used in studies in fair clustering, which was introduced by Chierichetti et al. [44]. Given a clustering $\mathcal{C} = \{C_1, C_2, ..., C_k\}$ with k clusters, *balance* of a cluster C_j is the minimum ratio between the cardinality of the discriminated group and that of the non-protected group in the cluster, and *balance* of clustering is the minimum *balance* score of all clusters. If each cluster has a balance score of at least θ as defined by the balance requirement parameter θ, then clustering is fair. Let $\psi : X \rightarrow \mathcal{P}$ denote the demographic group to which the point belongs, i.e., "male" or "female".

The *balance* fairness measure can be applied in any fair clustering model. The balance score of a cluster is measured by Eq. 2.13, while the balance score of a clustering is computed by the minimum of all balance scores of clusters in that clustering (Eq. 2.14).

$$balance(C_j) = \min\left(\frac{|\{x \in C_j \mid \psi(x) = p\}|}{|\{x \in C_j \mid \psi(x) = \overline{p}\}|}, \frac{|\{x \in C_j \mid \psi(x) = \overline{p}\}|}{|\{x \in C_j \mid \psi(x) = p\}|}\right) \tag{2.13}$$

$$balance(\mathcal{C}) = min_{j=1}^{k} balance(C_j) \tag{2.14}$$

Bounded representation

Bounded representation is a generalization of disparate impact for clustering and was introduced by Ahmadian et al. [8]. This group-level measure aims to reduce imbalances in cluster representations of protected attributes (for example, gender).

Let α be the user-given over-representation parameter; a cluster C_j is fair if the fractional representation of each group (protected, non-protected group) in the cluster is at most α.

$$\frac{|\ \{x \in C_j | \mathcal{P} = g\}\ |}{|\ C_j\ |} \leq \alpha, \text{where } g \in \{p, \overline{p}\}. \tag{2.15}$$

Then, a clustering $\mathcal{C}$ is fair if all clusters satisfy the representation constraint. Like the concept of *balance*, *bounded representation* notion can be applied by all fair clustering models. Furthermore, *bounded representation* is generalized with two parameters α and β in the study of Bera et al. [25], where for each group i of the protected attribute we have two parameters $\beta_i, \alpha_i \in [0, 1]$. The clustering solution is fair if each cluster satisfies two properties: i) the fraction of people from group i in any cluster is at most α_i and ii) the fraction of people from group i in any cluster is at least β_i.

Social fairness

Social fairness [74] aims to provide equitable costs for different clusters. In the k-means algorithm, the target is to minimize the objective function (recall Eq. 2.6):

$$\mathcal{L}(\mathcal{C}, X) = \sum_{j=1}^{k} \sum_{x \in C_j} dist(x, s_j)$$

We denote X_p and $X_{\overline{p}}$ as two subsets of dataset X, which contain value p and $\overline{p}$ of the protected attribute, respectively, $X = X_p \cup X_{\overline{p}}$. Then, the fair k-means objective is the higher average cost (*social fairness cost*):

$$\Phi(\mathcal{C}, X) = max \left\{ \frac{\mathcal{L}(\mathcal{C}, X_p)}{|\ X_p\ |}, \frac{\mathcal{L}(\mathcal{C}, X_{\overline{p}})}{|\ X_{\overline{p}}\ |} \right\} \tag{2.16}$$

The goal of the fair k-means is to minimize the social fairness cost $\Phi(\mathcal{C}, X)$. A disadvantage of this measure is that it can only be applied in partitioning-based clustering like k-means or k-medoids [42].

2.4.2 Fairness-aware clustering models

In the related work, we overview the popular fair clustering approaches, in particular, partitioning-based and hierarchical approaches.

Partitioning-based fair clustering

The first work on group-level fair clustering was introduced by Chierichetti et al. [44] with the aim to ensure an equal representation for each protected attribute in every cluster. They defined a new fairness notion, namely *balance*, which is described in Section 2.4.1. A two-phase approach is proposed: i) fairlet decomposition - clustering all instances into fairlets which are small clusters guarantying fairness constraint; ii) applying vanilla clustering methods (k-center, k-median) on those fairlets to obtain the final resulting fair clusters. Figure 2.6 illustrates the fair clustering method using the *fairlets* concept.

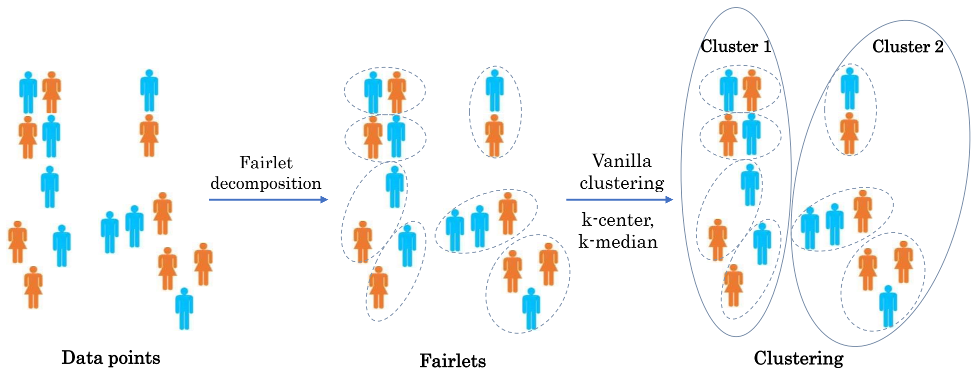


Figure 2.6. Fair clustering through fairlets

Next, Schmidt et al. [174] proposed a fair k-means clustering by using the *coresets* concept. Essentially, a coreset is a summary of a point set that approximates effectively the cost function of any potential solution. They extended the approach of Chierichetti et al. [44] to determine fairlets for the k-means algorithm by computing a minimum cost perfect matching between the *red* and *blue* points. They assigned edge weights based on the 1-means cost of clustering the two points with an optimal center. The experimental results showed that the new method is scalable. Ahmadian et al. [8] introduced a new fairness measure, namely *bounded representation* (see Section 2.4.1), by providing an upper bound constraint for fairness in resulting clusters, applied for k-center clustering. Jones et al. [97] proposed an algorithm with a linear time complexity and obtained a 3-approximation for the fair k-center for the data summarization problem. In addition, Ghadiri et al. [74] introduced the social fairness notion, which focuses on minimizing the clustering cost across groups of the protected attribute. They proposed a fair clustering version of the well-known Lloyd k-means and reported results through clustering cost. Abraham et al.[1] introduced a fair k-mean clustering model, namely *FairKM*, which combines the optimization of the classical clustering objective and a novel fairness loss term. Their model aims to achieve a trade-off between clustering quality (on the non-protected attributes) and cluster fairness (on the multiple protected attributes). Recently, Chakrabarti et al. [39] presented two individual fairness notions that guarantee each data point has a similar *quality of service*, and they proposed approximation algorithms for the k-center objective.

Hierarchical fair clustering

Ahmadian et al. [7] defined the fair hierarchical clustering for any fairness constraint, in which "a hierarchical clustering is fair if all of its clusters (besides the leaves) are fair". They extended the fairlet decomposition [44] for upper-bounded representation fairness and proved the results for three different objectives (revenue, value, and cost), computed based on the similarity and distance among data points. Their experimental results show that optimizing revenue and value with fairness considerations is relatively straightforward. However, the task of optimizing cost while ensuring fairness presents more inherent challenges. Recently, Chhabra and Mohapatra [43] propose fair algorithms for hierarchical agglomerative clustering that satisfy fairness constraints with multiple protected groups and distance linkage criteria as well as generalize to natural fairness notions for hierarchical agglomerative clustering. They define a fairness notion, namely *α-proportional fairness*, that satisfies the *ideal proportion* for each group, i.e., the desired proportion of points for each group in each cluster.

2.5 Knapsack problems

In this section, we present the essentials of knapsack problems, including problem definition and algorithms.

2.5.1 Problem definition

Knapsack problems have been studied for more than a century since the work of Mathews et al. [137] in 1897, and Dantzig [51] in 1930 in the field of combinatorial optimization. The knapsack problems are defined as: given a knapsack with *capacity* W and a set of n items; each item i has a *weight* w_i and a *value* v_i. The goal is to choose a subset of the items with the total value as large as possible while the total weight does not exceed the capacity W of the knapsack. All the coefficients w_i, v_i and W, $i = 1, \ldots, n$, are positive integers. Figure 2.7 demonstrates an example of a knapsack problem.

There are several variants of the knapsack problems [157] depending on the distributions of items and knapsacks. In this section, we introduce the most popular definitions of the knapsack problem.

Figure 2.7. An example of the knapsack problem: given a list of items, each item has a weight and value, the objective is to choose items in a way that maximizes the monetary amount, while also keeping the total weight at or below 7 kg

The 0-1 knapsack problem

This is the most common knapsack problem when each item may be chosen at most once. The *0-1 knapsack problem* is formalized as the maximization problem[10]:

$$\begin{gathered}\text{maximize} \sum_{i=1}^{n} v_i x_i \\ \text{subject to} \quad \sum_{i=1}^{n} w_i x_i \leq W \quad \text{and} \quad x_i \in \{0,1\}, \qquad i = 1, \ldots, n\end{gathered} \tag{2.17}$$

where x_i is a binary variable, $x_i = 1$ of item i is chosen for the knapsack, otherwise $x_i = 0$.

The bounded knapsack problem

In the *bounded knapsack problem*, each item type i has a bounded amount $c_i \geq 1$:

$$\begin{gathered}\text{maximize} \sum_{i=1}^{n} v_i x_i \\ \text{subject to} \quad \sum_{i=1}^{n} w_i x_i \leq W \quad \text{and} \quad x_i \in \{0,1,\ldots,c_i\}, \qquad i = 1, \ldots, n\end{gathered} \tag{2.18}$$

[10]It is also known as the *maximal 0-1 knapsack problem*.

The multi-constrained knapsack problem

The multi-constrained knapsack problem is the most general form of the knapsack problem, which is formulated as a general *integer programming problem*. The goal is to choose a set of items from n items to pack in m knapsacks of different capacities W_j ($j = 1, \ldots, m$) with a total value of all knapsacks as large as possible. All parameters v_i, w_i, and W_i are non-negative integers.

$$\begin{aligned} &\text{maximize} \sum_{i=1}^{n} v_i x_i \\ &\text{subject to} \quad \sum_{i=1}^{n} w_{ji} x_i \leq W_j, \quad j = 1, \ldots, m \\ &\text{and} \quad x_i \geq 0 \text{ integer}, \quad i = 1, \ldots, n \end{aligned} \tag{2.19}$$

2.5.2 Algorithms for knapsack problems

The knapsack problems are NP-hard [134] and can be solved by several exact and heuristic algorithms. The exact algorithms are divided into two groups: *branch and bound* approaches and *dynamic programming* approaches. We present the dynamic programming solution for the 0-1 knapsack problem since it will be employed in Chapter 5 and Chapter 6 of the thesis. The pseudo-code is described in Algorithm 3. We denote $M(i, w)$ as the maximum value that can be attained with the total weight not exceeding w using the first i items. In the initialization step (line 1), $M(0, w) = 0$ for all possible weight w. In the iteration step (line 2 and line 3), the maximum value of the first i items, $M(i, w)$, is computed based on the maximum value of the first $i - 1$ items and the value and weight of the i^{th} item.

Algorithm 3: Dynamic programming solution for the 0-1 knapsack problem

Input: A set of n items
$\{w_1, w_2, \ldots, w_n\}$: the weights of items
$\{v_1, v_2, \ldots, v_n\}$: the values of items
W: the maximum capacity of the knapsack

Output: The maximum total value of chosen items

1 $M(0, w) = 0$ //Initialization;
2 $M(i, w) = M(i - 1, w)$ if $w_i > w$ (the weight of the new item is higher than the current weight limit);
3 $M(i, w) = max\{M(i - 1, w), M(i - 1, w - w_i) + v_i\}$ if $w_i \leq w$;
4 **return** *M(n, W)*;

Complexity: The complexity of the dynamic programming approach is $\mathcal{O}(nW)$ in terms of computation time, and $\mathcal{O}(nW)$ in terms of space. The space complexity can be reduced to $\mathcal{O}(W)$ when we store only the recent two lines of the array M in the iteration step (line 2 and line 3 of Algorithm 3).

3 Bias-aware exploratory data analysis

As datasets play a foundational role in the advancement of ML and EDM research, in this chapter, we perform an exploratory analysis of well-known educational datasets. We first generate a BN to identify the relationship among attributes. We focus on the relationship between the protected attributes and the outcome attribute. Then, we provide a graphical analysis of the attributes to understand deeply the bias in the dataset. Finally, we evaluate the performance of the predictive model on datasets w.r.t. fairness measures and prediction outcomes.

3.1 Introduction

ML is essential in decision-making in almost all areas of our lives, including areas of high societal impact, like healthcare and education. In the educational domain, ML-based decision-making has been used in a wide variety of tasks, from student dropout prediction [72], forecasting on-time graduation of students [92] to educational admission decisions [144]. Along with the advantages, unfortunately, there is plenty of evidence regarding the discriminative impact of ML-based decision-making on individuals and groups of people based on *protected attributes* such as gender or race. Recently, the issue of bias and discrimination in ML-based decision-making systems has been receiving a lot of attention [151] as there are many recorded incidents of discrimination (e.g., recidivism prediction [112], grades prediction [29, 89]) caused by such systems against individuals or groups or people on the basis of *protected attributes* like gender, race, etc. Bias in education is not a new problem, rather there is already a long literature on different sources of bias in education [139] or students' data analysis [28] as well as studies on racial bias [194] and gender bias [136].

Datasets have a foundational role in the advancement of ML research [155]. The usage of sensitive information during the learning process is undesirable but hard to guarantee, even if known protected attributes are omitted from the analysis. The reason is the causal effects [132] of such attributes, including observable "proxy" attributes. As an example, the non-protected attribute "zip-code" was found to be a proxy for the protected attribute "race" [52] or "credit rating" can be used as a proxy for "safe driving" [195] attribute. Hence, even if protected attributes are not used, the resulting ML models can still be biased [17] due to the causal effects of such attributes. Although methods for detecting proxy attributes exist, e.g., [203] detects proxies in linear regression models by using a convex optimization procedure, eliminating all the correlated features might drastically reduce the utility of the data for the learning problem.

In this chapter, we analyze educational datasets by characterizing them according to their protected attributes and learning characteristics like cardinality, dimensionality, and class (im)balance. We provide a bias-aware exploratory analysis for each dataset by generating a BN and using the network to identify the relationships among attributes. Based on the BN, we provide a graphical analysis of the attributes for a deeper understanding of bias in the dataset. The BN illustrates the conditional (in)dependence between the protected attribute(s) and the class attribute; thus, it reduces the space and complexity of data analysis that needs to be performed to discover and clarify the fairness-related problems in the dataset. We then focus our exploratory analysis on features having a direct or indirect relationship with the protected attributes. We accompany our exploratory analysis with a quantitative evaluation of measures related to predictive and fairness performance. The results of our exploratory data analysis w.r.t. bias will be an important starting point for further in-depth problems, such as classification and clustering problems in the following Chapters.

The rest of the chapter is structured as follows: In Section 3.2, we describe the theory of the Bayesian network. The educational datasets used for this thesis are presented in Section 3.3 together with the results of their exploratory analysis. Section 3.4 demonstrates a quantitative evaluation of a predictive model on the datasets w.r.t. predictive performance and fairness. Finally, Section 3.5 summarizes the chapter with an outlook.

3.2 Bayesian network

A Bayesian network (BN) [86] is a directed and acyclic probabilistic graphical model that provides a graphical representation to understand the complex relationships between a set of random variables. In the case of a dataset, random variables correspond to the attributes of the feature space in which the data are represented. The graphical structure $\mathcal{M} : \{V, E\}$ of a BN contains a set of nodes V (random variables/attributes) and a set of directed edges E. Let $A_1, A_2, \cdots, A_d$ be the attributes defining the feature space $\mathcal{A}$ of a dataset X, such that $X \in \mathbb{R}^d$. For two attributes $A_i, A_j \in \mathcal{A}$, if there is a directed edge from A_i to A_j, then A_i is called the parent of A_j. The edges indicate conditional dependence relations, i.e., if we denote A_{pa_i} as the parents of A_i, the probability of A_i is conditionally dependent on the probability of A_{pa_i}. If we know the outcome (value) of A_{pa_i}, then the probability of A_i is conditionally independent of any other ancestor node. The structure of a BN describes the relationships between given attributes, i.e., the joint probability distribution of the attributes in the form of conditional independence relations. Formally:

$$P(A_1, A_2, \cdots, A_d) = \prod_{i=1}^{d} P(A_i \mid A_{pa_i}) \tag{3.1}$$

Learning the structure of a BN from the dataset X is an optimization problem [90], namely to learn an optimal BN model $\mathcal{M}^{\star}$ which maximizes the likelihood of generating X. A set of parameters of any BN model $\mathcal{M}$, denoted by $\widehat{\mathcal{M}}$, is the set of edges E which represents the conditional independence relationship between the attribute set V. Moreover, between the possible models M, the less complex one, i.e., the one with the least $\widehat{\mathcal{M}}$, should be selected.

Note that in a learned BN model $\mathcal{M}$, the position of the class attribute Y can be in any position (root-, internal- or leaf-node) since the objective is to maximize $P(X \mid \mathcal{M})$. However, we aim to investigate the factors (protected/non-protected attributes) that determine the class attribute's prediction probability. Therefore, we also employ a constraint on the class attribute as a leaf node in our learning objective. Formally the problem is defined as:

$$\begin{gathered} \max_{\mathcal{M}^{\star}} \{P(X \mid \mathcal{M}) - \gamma\widehat{\mathcal{M}}\} \\ \text{subject to } Y \in L \end{gathered} \tag{3.2}$$

where $Y \in \mathcal{A}$ is the class attribute, L is the set of leaf nodes, and γ is a penalty hyperparameter controlling the effect of the model's complexity in the final model selection. The aim of the learned model is to maximize $P(A_i \mid A_{pa_i})$ for each $A_i \in \mathcal{A}$ (Eq. 3.1 and Eq. 3.2).

A high conditional probability often refers to a strong correlation [50]. Attribute A_i is strongly correlated with A_j if a *direct edge* exists between A_i and A_j, for any pair of attributes $A_i, A_j \in \mathcal{A}$. Intuitively, the correlation is comparatively weaker with ancestors that are not immediate parents, i.e., *indirect edges*. In addition, for the attributes that do not have any incoming or outgoing edge (direct/indirect connection) with A_i, the correlation between them will be negligible. Consequently, if we find any direct/indirect edge from any protected attribute to the class attribute in our learned BN structure $\mathcal{M}^{\star}$ then we may infer that the dataset is biased w.r.t. the specific protected attribute.

When learning a BN, the continuous variables are often discretized because many BN learning algorithms cannot efficiently handle continuous variables [41]. Therefore, we need to discretize the continuous numeric data attributes into meaningful categorical attributes to keep the complexity of learning the BN model in a polynomial time. We describe the discretization procedure for each dataset in Section 3.3.

3.3 Educational datasets

In this section, we provide detailed descriptions of real-world educational datasets used in this thesis. For each dataset, we discuss the basic characteristics like cardinality, dimensionality, and class imbalance as well as typically used protected attributes in the literature. When available, we also provide temporal information regarding the data collection and the timespan of the datasets. A summary of the statistics of the different educational datasets[1] is provided in Table 3.1.

Table 3.1. Overview of real-world 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

We start our analysis with the BN structure learned from the data (see Section 3.2), which can help us understand the relationships between attributes of the dataset. In addition, the BN visualization already provides interesting insights into the dependencies between non-protected and protected attributes and their conditional dependencies in predicting the class attribute. We further provide an exploratory analysis of interesting correlations from the Bayesian graph (for both direct- and indirect-edges), particularly those related to the fairness problem (paths to and from protected attributes).

3.3.1 Student performance dataset

The student performance dataset [47] described students' achievement in the secondary education of two Portuguese schools in 2005 - 2006 with two distinct subjects: Mathematics and Portuguese.[2]. The regression task is to predict the final-year grades of the students.

[1] In this thesis, *sex*, and *gender* are used interchangeably with the same meaning. The IR values are reported on the cleaned datasets.

[2] https://archive.ics.uci.edu/ml/datasets/student+performance

Dataset characteristics: The dataset contains information about 395 (Mathematics subject, in short: *Student-Math*) and 649 (Portuguese subject, in short: *Student-Por*) students described by 33 attributes (4 categorical, 13 binary, and 16 numerical attributes). Characteristics of all attributes are described in Table 3.2. To simply the classification problem, we create a class label based on attribute *G3*, *class* = {*Low, High*}, corresponding to *G3* = {<10, ≥10}. The positive class is "High". The dataset is imbalanced with imbalance ratios (IR) of 2.04:1 (Mathematics subject) and 5.49:1 (Portuguese subject).

Table 3.2. Student performance: attributes characteristics

Attributes	**Type**	**Values**	**#Missing values**	**Description**
school	Binary	{GP, MS}	0	The student's school ('GP': Gabriel Pereira, 'MS': Mousinho da Silveira)
sex	Binary	{Male, Female}	0	Sex
age	Numerical	[15 - 22]	0	Age (in years)
address	Binary	{U, R}	0	The address type ('U': urban, 'R':rural)
famsize	Binary	{LE3, GT3}	0	The family size ('LE3': less or equal to 3, 'GT3': greater than 3)
Pstatus	Binary	{T, A}	0	The parent's cohabitation status ('T': living together, 'A': apart)
Medu	Numerical	[0 - 4]	0	Mother's education
Fedu	Numerical	[0 - 4]	0	Father's education
Mjob	Categorical	5	0	Mother's job
Fjob	Categorical	5	0	Father's job
reason	Categorical	4	0	The reason to choose this school
guardian	Categorical	3	0	The student's guardian (mother, father, other)
traveltime	Numerical	[1 - 4]	0	The travel time from home to school
studytime	Numerical	[1 - 4]	0	The weekly study time
failures	Numerical	[0 - 3]	0	The number of past class failures
schoolsup	Binary	{Yes, No}	0	Is there an extra educational support?
famsup	Binary	{Yes, No}	0	Is there any family educational support?
paid	Binary	{Yes, No}	0	Is there an extra paid classes within the course subject (Math or Portuguese)
activities	Binary	{Yes, No}	0	Are there extra-curricular activities?
nursery	Binary	{Yes, No}	0	Did the student attend a nursery school?
higher	Binary	{Yes, No}	0	Does the student want to take a higher education?
internet	Binary	{Yes, No}	0	Does the student have an Internet access at home?
romantic	Binary	{Yes, No}	0	Does the student have a romantic relationship with anyone?
famrel	Numerical	[1 - 5]	0	The quality of family relationships (1: very bad - 5: excellent)
freetime	Numerical	[1 - 5]	0	Free time after school (1: very low - 5: very high)
goout	Numerical	[1 - 5]	0	How often does the student go out with friends? (1: very low - 5: very high)
Dalc	Numerical	[1 - 5]	0	The workday alcohol consumption (1: very low - 5: very high)
Walc	Numerical	[1 - 5]	0	The weekend alcohol consumption (1: very low - 5: very high)
health	Numerical	[1 - 5]	0	The current health status (1: very bad - 5:very good)
absences	Numerical	[0 - 32]	0	The number of school absences
G1	Numerical	[0 - 19]	0	The first period grade
G2	Numerical	[0 - 19]	0	The second period grade
G3	Numerical	[0 - 19]	0	The final grade

Protected attributes: Typically, *sex* is considered as the protected attribute. In the work of [53, 102], they also select *age* as the protected attribute. Especially in the research of Kearns et al. [102], they consider attributes *romatic* (relationship) and *dalc, walc* (alcohol consumption) as the protected attributes. However, because of the unpopularity of these attributes, we did not consider those within the scope of this thesis.

- *sex* = {*male, female*}: the dataset is dominated by female students. The ratios of *male:female* are 208:187 (52.7%:47.3%) and 383:266 (59%:41%) for the Mathematics subject and Portuguese subject, respectively.

- *age* = {<*18*, ≥ *18*}: young students (less than 18 years old) are the majority with the ratios of "< 18":"≥ 18" are 284:111 (71.9%: 28.1%) and 468:181 (72.1%:27.9%) for the Mathematics subject and Portuguese subject, respectively.

Bayesian network: We perform a transformation of numerical variables: the number of school absences, *absences* = {0-5, 6-20, >20}; grade *G1* = {<10, ≥10}; *G2* = {<10, ≥10}. Due to the computation of the BN generator and the correlation coefficient with the class label (with a threshold of 0.02), we select 26 variables for the network. The BNs of the dataset on Portuguese and Mathematics subjects are visualized in Figure 3.1 and Figure 3.2, respectively.

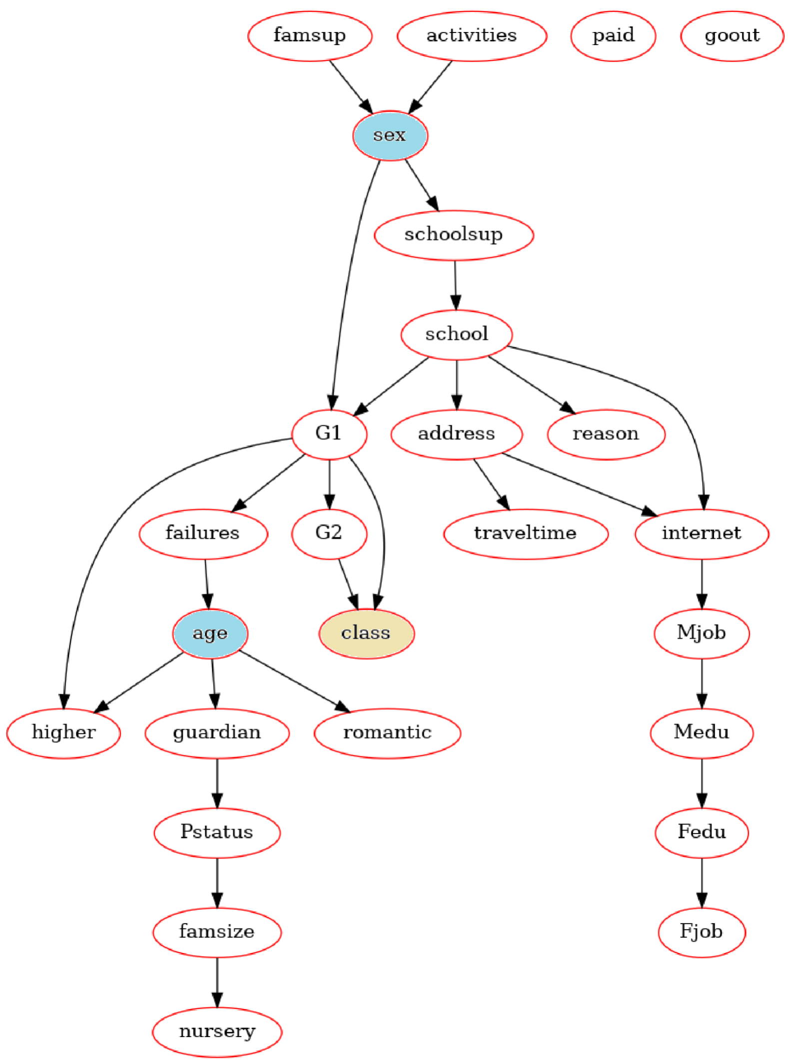


Figure 3.1. Student-Por: Bayesian network (class label: *class*, protected attributes: *age*, *sex*)

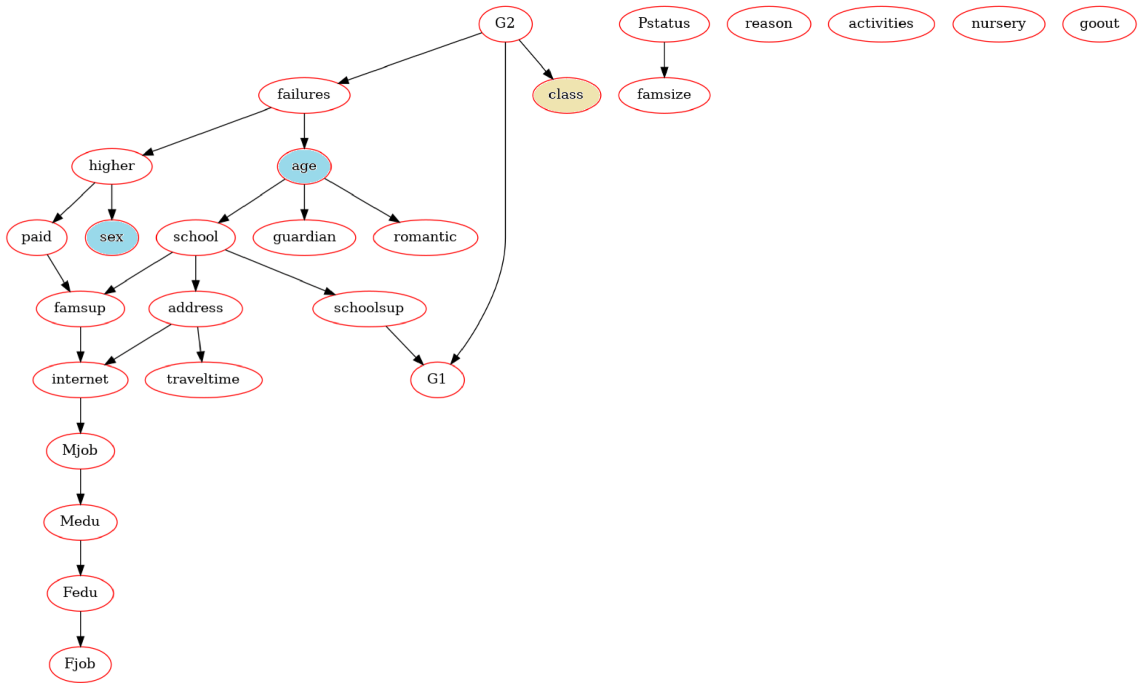
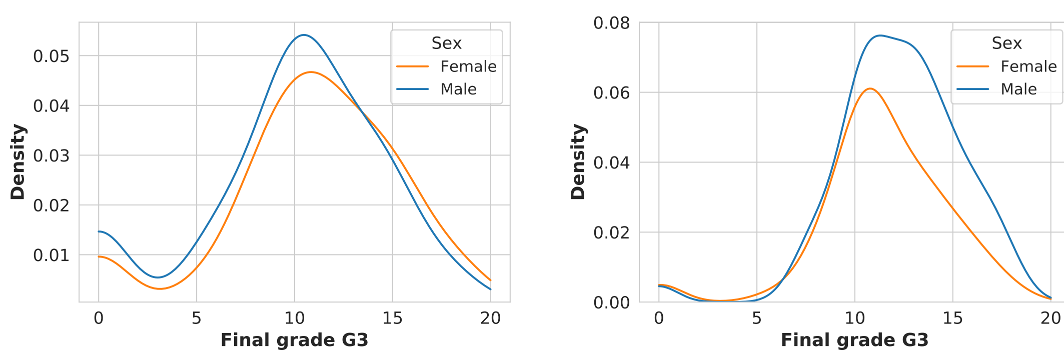


Figure 3.2. Student-Math: Bayesian network (class label: *class*, protected attributes: *age*, *sex*)



(a) Mathematics subject

(b) Portuguese subject

Figure 3.3. Student performance: Distribution of the final grade G3 w.r.t. *sex*

The *class label* attribute is conditionally dependent on the grade *G2* in both subsets (Mathematics and Portuguese subjects). This is explained by a very high correlation coefficient (above 90%) between *G2* and *G3* variables. In addition, we investigate the distribution of the final grade *G3* on *sex* because the attribute *sex* has an indirect relationship with the *class label*. Figure 3.3 reveals that male students tend to receive high scores in Portuguese, while Math scores are relatively evenly distributed across both sexes.

3.3.2 OULAD dataset

The Open University Learning Analytics (OULAD) dataset[3] was collected from the OU analysis project [110] of the Open University (England) in 2013 - 2014. The dataset contains information of students and their activities in the virtual learning environment (VLE) for 7 courses. The goal is to predict the success of students.

Dataset characteristics: The dataset contains information of 32,593 students characterized by 12 attributes (7 categorical, 2 binary and 3 numerical attributes). An overview of all attributes is illustrated in Table 3.3. Attribute *id_student* should be ignored in the analysis. Typically, in the related work, they consider the prediction task on the class label *final_result* = {*pass, fail*}. Therefore, we investigate the cleaned dataset with 21,562 instances after removing the missing values and rows with *final_result* = *"withdrawn"*. *Pass* is the positive class. The ratio of *pass:fail* is 14,655:6,907 (68%:32%). In other words, the dataset is imbalanced with the IR is 2.12:1 (positive:negative).

Table 3.3. OULAD: attributes characteristics

Attributes	Type	Values	#Missing values	Description
code_module	Categorical	7	0	The identification code of the module on which the student is registered
code_presentation	Categorical	4	0	The identification code of the presentation on which the student is registered
id_student	Numerical	[3,733 - 2,716,795]	0	A unique identification number for the student
gender	Binary	{Male, Female}	0	Gender
region	Categorical	13	0	The geographic region
highest_education	Categorical	5	0	The highest student education level
imd_band	Categorical	10	1111	Index of multiple deprivation (IMD) band of the student's place of residence
age_band	Categorical	3	0	The category of the student's age
num_of_prev_attempts	Numerical	[0 - 6]	0	The number times the student has attempted this module
studied_credits	Numerical	[30 - 655]	0	The total number of credits for the modules the student is currently studying
disability	Binary	{Yes, No}	0	Whether the student has declared a disability
final_result	Categorical	4	0	The student's final result (in the module presentation)

Protected attributes: *gender* = {*male, female*} is considered as the protected attribute, in the literature. *Male* is the majority group with the ratio *male:female* is 11,568:9,994 (53.6%:46.4%).

Bayesian network: The numerical attributes are encoded for generating the BN: *num_of_prev_attempts* = {0, >0}, *studied_credits* = {≤100, >100}. The network is depicted in Figure 3.4. The final result attribute is directly conditionally dependent on the highest education level (*highest_education*) and the number of times the student has attempted the module (*num_of_prev_attempts*) attributes, while *gender* has a more negligible effect on the outcome.

We perform the analysis on the relationship of the highest education, the number of previous attempts and the final result for each gender. As demonstrated in

[3] https://analyse.kmi.open.ac.uk/open_dataset

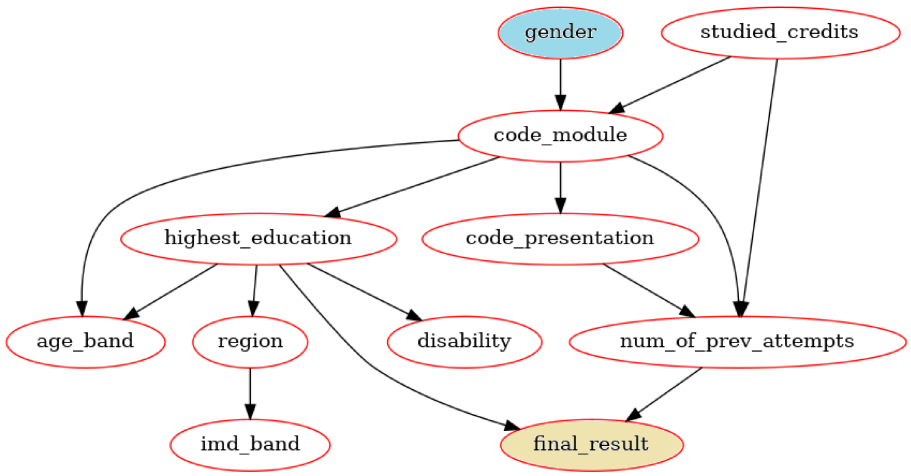


Figure 3.4. OULAD: Bayesian network (class label: *final_result*, protected attributes: *gender*)

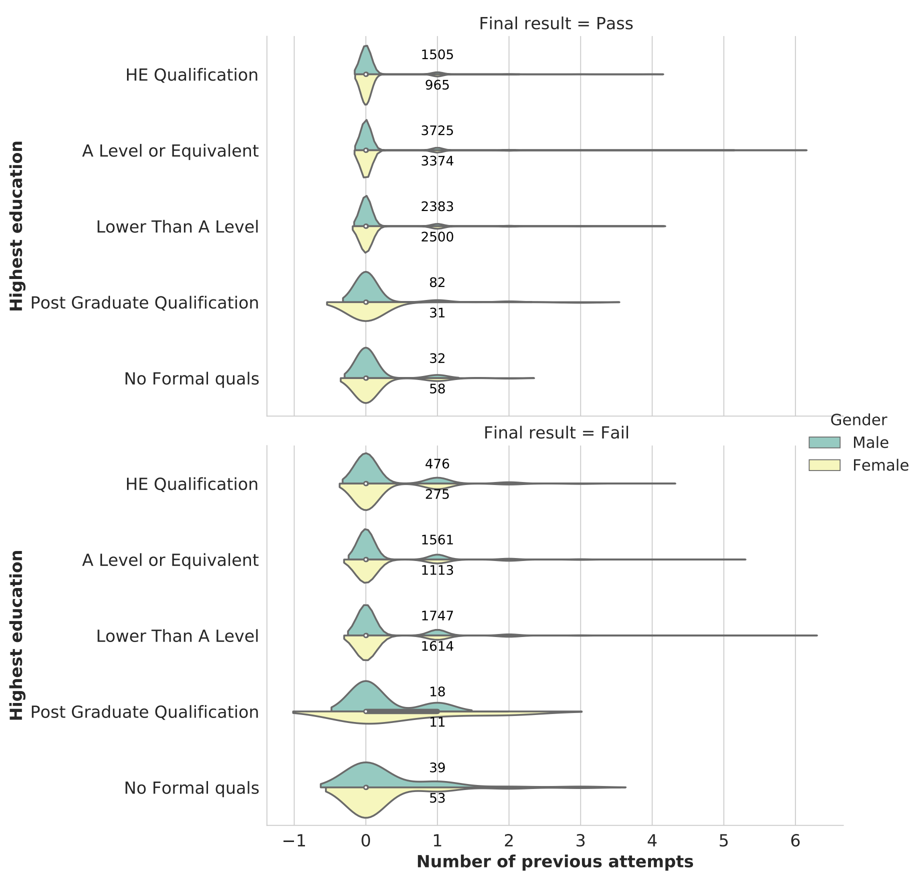


Figure 3.5. OULAD: Distribution of *the number of previous attempts*, *the highest education* and *the final result* w.r.t. *gender*

Figure 3.5, students have a higher probability of failure when attempting the exam many times in the past. 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.

3.3.3 PISA test scores dataset

The PISA test scores[4] dataset (in short: PISA) contains information about the demographics and schools of American students [66] taking the exam in 2009 from the Program for International Student Assessment (PISA) distributed by the United States National Center for Education Statistics (NCES).

Dataset characteristics: The dataset contains information of 5,233 students characterized by 24 attributes (1 categorical, 18 binary and 5 numerical attributes). An overview of all attributes is depicted in Table 3.4. In the original dataset, the ML problem could be grade prediction w.r.t. *readingScore* attribute. However, in this work, we consider the classification problem by creating a class label based on attribute *reading scores*, $class = \{Low, High\}$, corresponding to $reading\ scores = \{<500, \geq 500\}$. The positive class is "High". The dataset is imbalanced with an imbalance ratio of 1.40:1.

Table 3.4. PISA: attributes characteristics

Attributes	Type	Values	#Missing values	Description
grade	Numerical	8 - 12	0	The grade in school of the student
male	Binary	{0, 1}	0	Whether the student is male (1:male/0:female)
raceeth	Categorical	7	48	The race/ethnicity composite of the student
preschool	Binary	{0, 1}	77	Whether the student attended preschool
expectBachelors	Binary	{0, 1}	85	Whether the student expects to obtain a bachelor's degree
motherHS	Binary	{0, 1}	142	Whether the student's mother completed high school
motherBachelors	Binary	{0, 1}	585	Whether the student's mother obtained a bachelor's degree
motherWork	Binary	{0, 1}	129	Whether the student's mother works part-time or full-time
fatherHS	Binary	{0, 1}	370	Whether the student's father completed high school
fatherBachelors	Binary	{0, 1}	857	Whether the student's father obtained a bachelor's degree
fatherWork	Binary	{0, 1}	346	Whether the student's father works part-time or full-time
selfBornUS	Binary	{0, 1}	93	Whether the student was born in the US
motherBornUS	Binary	{0, 1}	94	Whether the student's mother was born in the US
fatherBornUS	Binary	{0, 1}	171	Whether the student's father was born in the US
englishAtHome	Binary	{0, 1}	98	Whether the student speaks English at home
computerForSchoolwork	Binary	{0, 1}	95	Has the student access to a computer at school?
read30MinsADay	Binary	{0, 1}	55	Whether the student reads for pleasure for 30 minutes/day
minutesPerWeekEnglish	Numerical	0 - 2025	289	The number of minutes spent per week in English class
studentsInEnglish	Numerical	1.0 - 90.0	363	This student's English class size
schoolHasLibrary	Binary	{0, 1}	201	Whether this student's school has a library
publicSchool	Binary	{0, 1}	0	Whether this student attends a public school
urban	Binary	{0, 1}	0	Whether this student's school is in an urban area
schoolSize	Numerical	100 - 6,694	231	The number of students in this student's school
readingScore	Numerical	242.64 - 772.46	0	The student's reading score, on a 1000-point scale

Protected attributes: In general, *raceeth* (race/ethnicity) and *male* can be consid-

[4] https://www.kaggle.com/econdata/pisa-test-scores

ered as protected attributes.

- *male* = {0, 1}. *Female* is the majority group. The ratio of *male* (1):*female* (0) is 1,697:1,707 (49.9%:50.1%).

- *raceeth* = {*White, Hispanic, Black, Asian, More than one race, American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander*}. We encode new attribute *race* = {*white, non-white*} based on the original *raceeth* attribute. *Non-white* is the minority group with the ratio *white:non-white* is 2,083:1,321 (61.2%:38.8%).

Bayesian network: To generate the BN, we encode the numerical attributes as follows: *minutesPerWeekEnglish*={135, 136 - 270, $>$ 270} (equivalent to 3, 6, 9 teaching units where each teaching unit is 45 minutes); *schoolSize*={$\leq$ 500, 501 - 1000, 1001 - 2000, $>$ 2000} (based on suggestions of Leithwood et al. [121]; *studentsInEnglish* = {$\leq$ 25, $>$ 25 } [30]. The BN of the PISA dataset is visualized in Figure 3.6.

We discover that the *class* attribute is conditionally dependent on several attributes: race/ethnicity (*raceeth)*, the degree of the student's father (*fatheBachelors*), the degree expectation of the student (*expectBachelors*), and the reading manner of the student (*read30MinsADay*). Besides, there is an indirect connection between the protected attribute *male* and the *class* attribute. Therefore, we investigate the relationship among protected attributes (*male, raceeth)* and the class label. We have found that 67.3% of *white* students (1,401/2,083) obtain high reading scores, while this ratio in *non-white* students is only 44.4% (Figure 3.7)-a. Interestingly, the ratio of female students having high reading scores is higher than that ratio of male students (63.2% and 53.5%, respectively), as visualized in Figure 3.7-b.

3.3.4 MOOC dataset

The massive open online course (MOOC) dataset covers students who enrolled in the 16 edX courses offered by the two institutions (Harvard University and the Massachusetts Institute of Technology) during 2012- 2013 [84]. The dataset contains aggregated records that represent students' activities and their final grades in the courses. The classification task is to predict whether a student will earn a certificate [196].

Dataset characteristics: The dataset contains information of 416,921 students characterized by 21 attributes (9 categorical, 4 binary and 8 numerical attributes). The class label *certified* = {0, 1} is used for the classification task [196]. The positive class is *1 - earned a certificate*. The dataset is imbalanced with an imbalance ratio 1:27.0 (positive:negative). Figure 3.8 demonstrates the distribution of students who earn the certificate across courses. Table 3.5 provides an overview of all attributes of the dataset.

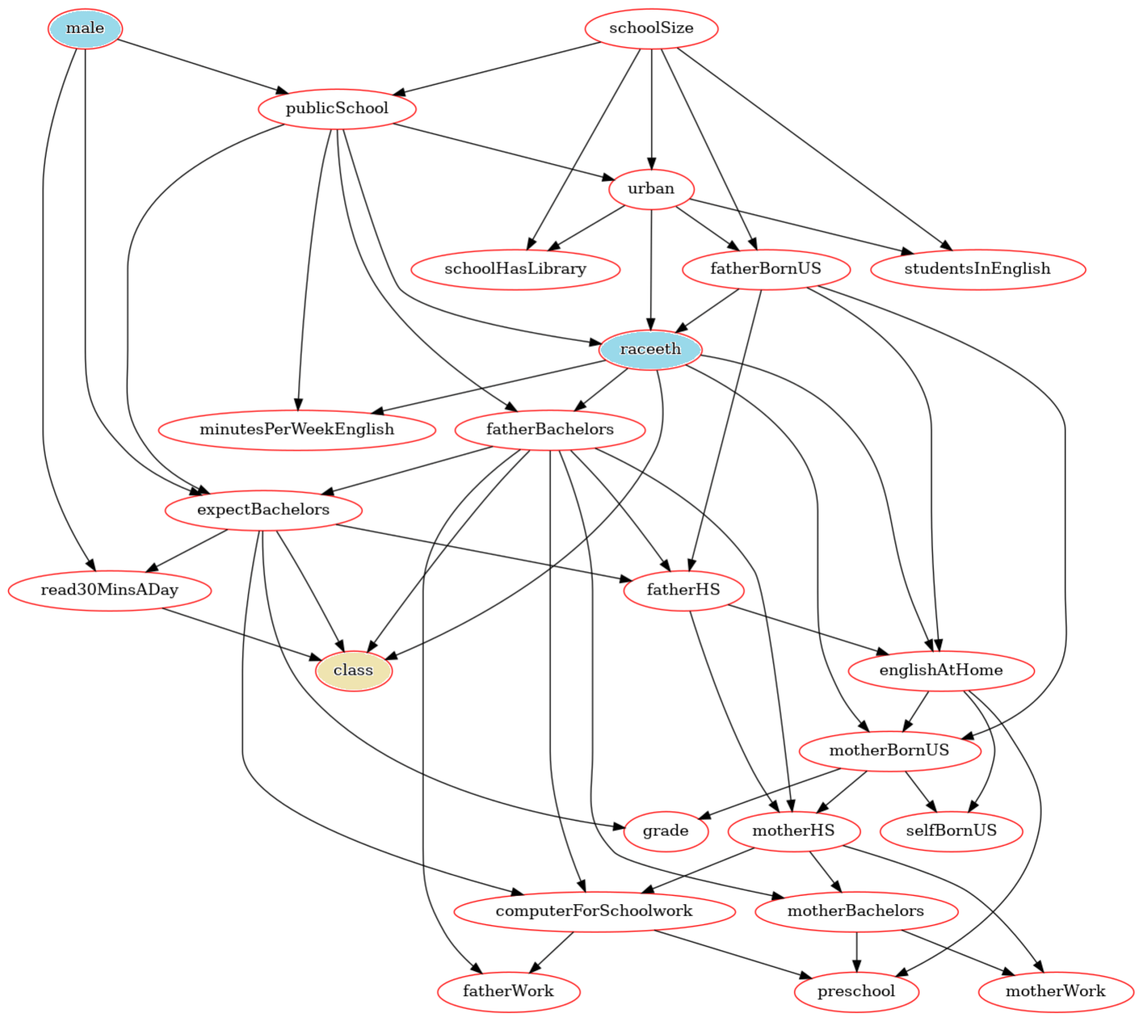
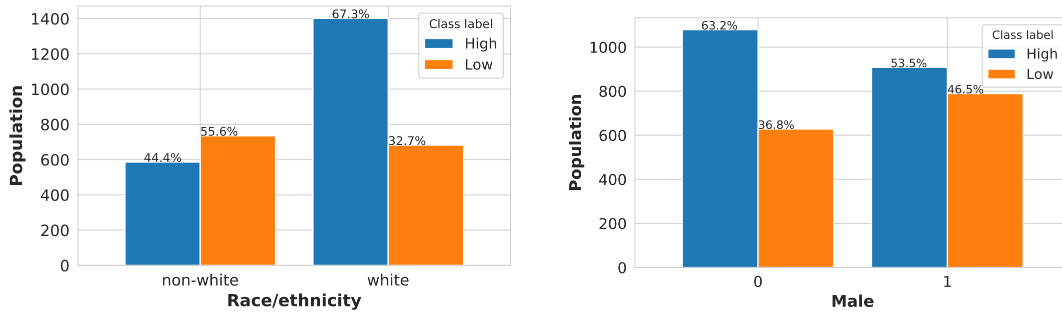


Figure 3.6. PISA: Bayesian network (class label: *class*, protected attributes: *male*)

Protected attributes: In the literature, *gender* is considered as the protected attribute. *Gender* = {m, f, o}. Since there are only 4 people with *gender* = *"o"*, we eliminate 4 corresponding instances from the dataset. In the cleaned version, *Male* is the majority group. The ratio of *male* (m):*female* (f) is 292:246:101,219 (74.3%:25.7%).

Bayesian network: In this thesis, we investigate a subset of students studying the course *6.002x* due to the computation complexity of the BN. We remove index-related attributes such as *course_id, userid_DI*. Then, we encode the numerical attributes as follows: $age = \{< 18, \geq 18\}$, $grade = \{< 0.5, \geq 0.5\}$, $nevents = \{\leq 500, 501 - 2000, > 2000\}$, $ndays\_act = \{\leq 7, 8 - 30, > 30\}$, $nforum\_posts = \{0, > 0\}$, $nplay\_video = \{\leq 5000, 5001 - 20000, > 20000\}$, $nchapters = \{\leq 1, > 1\}$. A new attribute *country* = {*US, non-US*} is computed based on attribute *final_cc_cname_DI*. The BN of the MOOC dataset is visualized in Figure 3.9 using 16 attributes.



(a) Protected attribute: Race/ethnicity (b) Protected attribute: Male

Figure 3.7. PISA: The percentage of students having high reading scores w.r.t. protected attributes

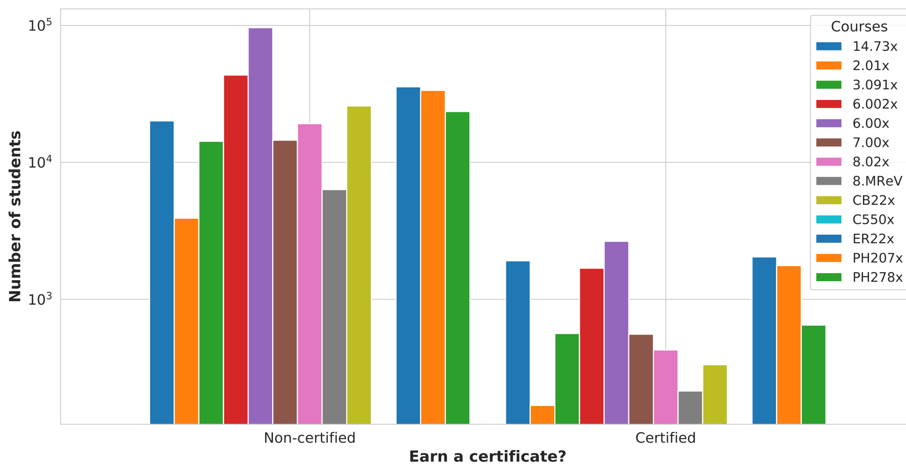


Figure 3.8. MOOC: The distribution of students who have obtained certificates in courses)

Figure 3.9 shows that the class label *certified* is conditionally dependent on the *grade* and *year*, and there is no connection between *certified* and the protected attribute *gender*. The number of chapters with which the student interacted (*nchapters*) has strong connections with other attributes. Therefore, we investigate the relation among *nchapters*, *year*, *gender*, and the class label *certified*. The result is visualized in Figure 3.10. The students who interacted with more than 10 chapters are more likely to get a certificate. Besides, most students, of both sexes, tend to read less than 5 chapters and spend less than a week interacting with lectures, which is revealed in Figure 3.11.

Table 3.5. MOOC: attributes characteristics

Attributes	Type	Values	#Missing values	Description
institute	Categorical	2	0	The online course holders (MIT, Harvard)
course_id	Categorical	13	0	Identifies institution,course name and semester
year	Numerical	2012 - 2013	0	The launch year of the course
semester	Categorical	3	0	The launch semester of the course
userid_DI	Categorical	318106	0	A random ID of the user
viewed	Binary	{0, 1}	0	Accessed course materials from the "Courseware" tab?
explored	Binary	{0, 1}	0	Accessed at least half of the chapters in the courseware?
certified	Binary	{0, 1}	0	Anyone who earned a certificate?
final_cc_cname_DI	Categorical	34	0	Country name
LoE_DI	Categorical	5	0	Highest level of education completed
gender	Categorical	3	0	Gender
grade	Numerical	0.0 - 1.0	0	The final grade of the course
start_time_DI	Categorical	411	0	The date of course registration
last_event_DI	Categorical	404	0	The date of the last interaction with course
nevents	Numerical	0 - 53,180	0	The number of interactions with the course (from tracking logs)
ndays_act	Numerical	0 - 205	0	The number of unique days students interacted with the course
nplay_video	Numerical	1 - 197,757	0	The number of play video events within the course
nchapters	Numerical	0 - 47	0	The number of chapters with which the student interacted
nforum_posts	Numerical	0 - 6	0	The number of posts to the *Discussion forum*
incomplete_flag	Binary	{0, 1}	0	Identifies records that are internally inconsistent
age	Numerical	1 - 82	0	Age

3.3.5 Law school dataset

The Law school[5] dataset [197] was conducted by a Law School Admission Council (LSAC) survey across 163 law schools in the United States in 1991. The dataset contains the law school admission records. The task is to predict whether a candidate would pass the bar exam or predict a student's first-year average grade (FYA).

Dataset characteristics: The dataset contains information of 20,798 students characterized by 12 attributes (3 categorical, 3 binary and 6 numerical attributes). The class label *pass_bar* = {0, 1} is used for the classification task. The positive class is *1 - pass*. The dataset is imbalanced with an imbalance ratio of 8.07:1 (positive:negative). An overview of all attributes is depicted in Table 3.6.

Protected attributes: In the literature, *race* and *male* are considered as the protected attributes.

- *male* = {0, 1}. *Male* is the majority group. The ratio of *male* (1):*female* (0) is 11,675:9,123 (56.1%:43.9%).
- *race* = {*white, black, Hispanic, Asian, other*}. As introduced in the related work, we encode *race* = {*white, non-white*} based on the original attribute. *Non-white* is the minority group with the ratio *white:non-white* is 17,491:3,307 (84%:16%).

[5] https://github.com/tailequy/fairness_dataset/tree/main/Law_school

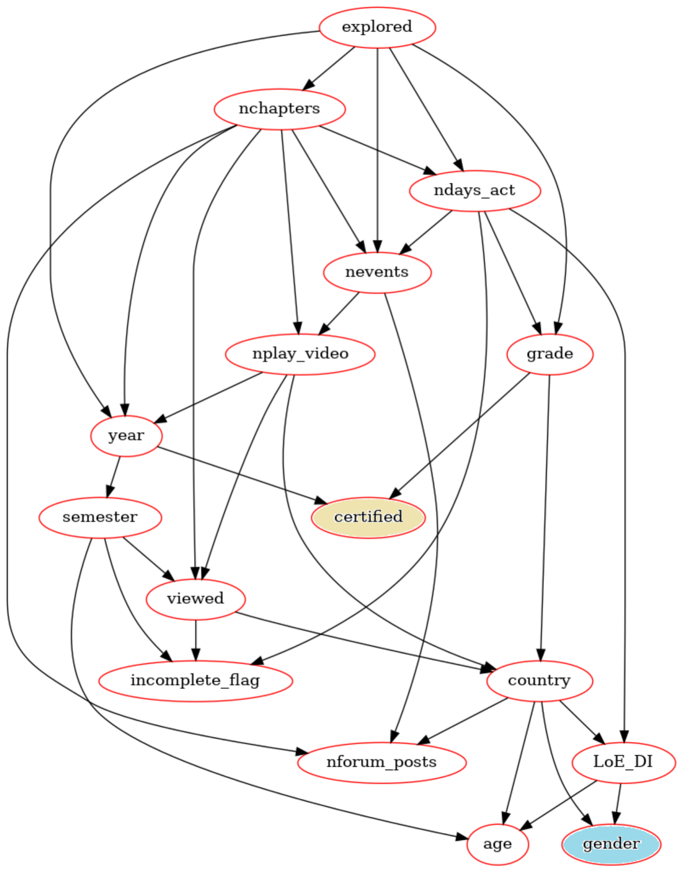


Figure 3.9. MOOC: Bayesian network (class label: *certified*, protected attributes: *gender*)

Table 3.6. Law school: attributes characteristics

Attributes	Type	Values	#Missing values	Description
decile1b	Numerical	[1.0 - 10.0]	0	The student's decile in the school given his grades in Year 1
decile3	Numerical	[1.0 - 10.0]	0	The student's decile in the school given his grades in Year 3
lsat	Numerical	[11.0 - 48.0]	0	The student's LSAT score
ugpa	Numerical	[1.5 - 4.0]	0	The student's undergraduate GPA
zfygpa	Numerical	[-3.35 - 3.48]	0	The first year law school GPA
zgpa	Numerical	[-6.44 - 4.01]	0	The cumulative law school GPA
fulltime	Binary	{1, 2}	0	Whether the student will work full-time or part-time
fam_inc	Categorical	5	0	The student's family income bracket
male	Binary	{0, 1}	0	Whether the student is a male or female
tier	Categorical	6	0	Tier
race	Categorical	6	0	Race
pass_bar	Binary	{0, 1}	0	Whether the student passed the bar exam on the first try

Bayesian network: To generate the BN, we encode the numerical attributes as follows: *decile1b* = $\{\leq 5, >5\}$, *decile3* = $\{\leq 5, >5\}$, *lsat* = $\{37, >37\}$, *ugpa* = $\{<3.3, \geq 3.3\}$, *zgpa* = $\{\leq 0, >0\}$, *zfygpa* = $\{\leq 0, >0\}$. The BN is visualized in Figure 3.12.

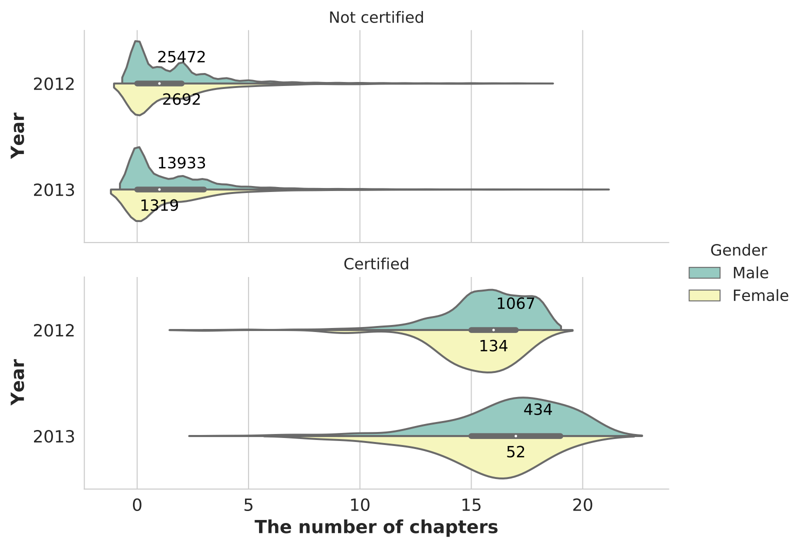
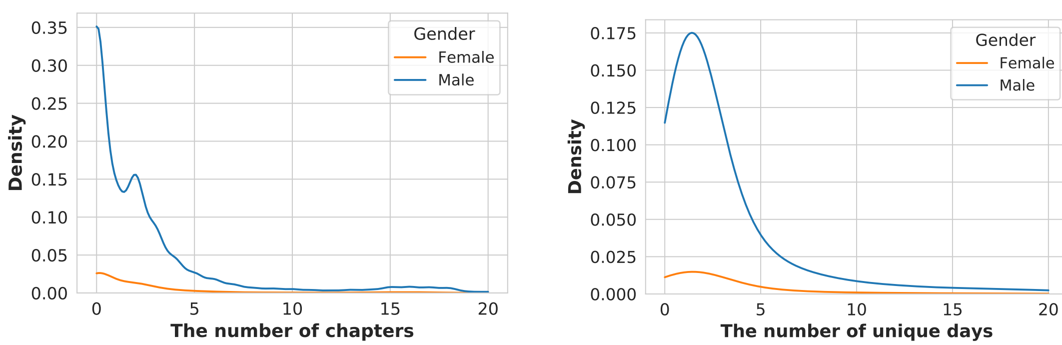


Figure 3.10. MOOC: Distribution of the number of chapters with which the student interacted, the launch year of the course, and the certificate status w.r.t. *gender*



(a) W.r.t. the number of chapters with which the student interacted

(b) W.r.t. unique days students interact the course

Figure 3.11. MOOC: The relation of *gender* with the number of chapters with which the student interacted, the number of unique days students interact with the course

It is easy to observe that the bar exam's result is conditionally dependent on the law school admission test (LSAT) score, undergraduate grade point average (UGPA) and *Race*. We discover that 92.1% of *white* students (16,114/17,491) pass the bar exam, while this ratio in *non-white* students is only 72.3%. In general, the percentage of students who passed the bar exam is increased in proportion to the LSAT and UGPA scores, which is depicted in Figure 3.13.

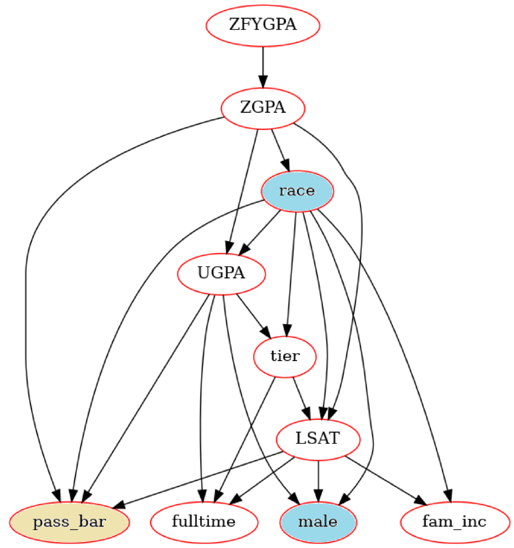
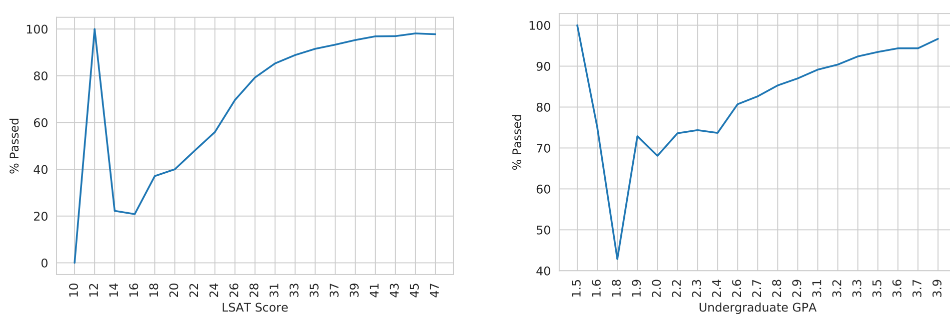


Figure 3.12. Law school: Bayesian network (class label: *pass_bar*, protected attributes: *male, race*)



(a) LSAT score (b) Undergraduate GPA

Figure 3.13. Law school: The percentage of students that passed the bar exam by LSAT and UGPA scores

3.3.6 Student academics performance dataset

The student academics performance dataset[6] [91] (in short: *student aca.*) consists of socio-economic, demographic, and academic information of students from three different colleges in India. The dataset was collected in 8 years (from July 2006 to July 2013).

Dataset characteristics: The dataset contains information of 131 students described by 22 attributes (17 categorical, 5 binary). The original class label is ESP (end semester percentage), $ESP = \{Best,\ Very\ good\ (Vg),\ Good,\ Pass\}$. However, in this thesis, we consider the binary classification problem. Hence, we create a new class

[6] https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance

label as a binary attribute based on the *ESP* attribute with values {*"pass"*, *"good-and-higher"*}, where "good-and-higher" is a positive class. The dataset is imbalanced with an imbalance ratio 3.85:1 (positive:negative). *Gender* (ge) = {F, M} (F: female, M: male) is the protected attribute. *Male* is the majority group with the ratio of *male*:*female* is 72:59 (55%:45%). An overview of all attributes is demonstrated in Table 3.7.

Table 3.7. Student aca. dataset: attributes characteristics

Attributes	Type	Values	#Missing values	Description
ge	Binary	{F, M}	0	Gender (Male - M, Female - F)
cst	Categorical	5	0	Caste (General, SC, ST, OBC, MOBC)
tnp	Categorical	4	0	Class X percentage (Best, Very good, Good, Pass)
twp	Categorical	4	0	Class XII percentage (Best, Very good, Good, Pass)
iap	Categorical	4	0	Internal assessment percentage (Best, Very good, Good, Pass)
esp	Categorical	4	0	End semester percentage (Best, Very good, Good, Pass)
arr	Binary	{Y, N}	0	Whether the student has back or arrear papers
ms	Categorical	1	0	Marital Status (Unmarried)
ls	Binary	{T, V}	0	Lived in town (T) or village (V)
as	Binary	{Paid, Free}	0	Admission category
fmi	Categorical	5	0	Family monthly income
fs	Categorical	3	0	Family size (Large, Average, Small)
fq	Categorical	6	0	Father qualification (IL, UM, 10, 12, Degree, PG)
mq	Categorical	6	0	Mother qualification (IL, UM, 10, 12, Degree, PG)
fo	Categorical	5	0	Father occupation (Service, Business, Retired, Farmer, Others)
mo	Categorical	5	0	Mother occupation (Service, Business, Retired, Farmer, Others)
nf	Categorical	3	0	Number of friends (Large, Average, Small)
sh	Categorical	3	0	Study hours (Good, Average, Poor)
ss	Binary	{Govt., Private}	0	Student school attended at class X level
me	Categorical	4	0	Medium
tt	Categorical	3	0	Home to college travel time (Large, Average, Small)
atd	Categorical	3	0	Class attendance percentage (Good, Average, Poor)

Bayesian network: In this dataset, because all attributes are categorical or binary, an encoding procedure is not needed. The BN is visualized in Figure 3.14. In the BN, the name of attributes is replaced by their descriptions. Interestingly, quite a few attributes of the dataset are independent as they are isolated in the BN. Besides, the class label *end semester percentage* is also not conditionally dependent on the protected attribute *gender*. The class label *end semester percentage* has a strong connection with the status of the *back or arrear papers*, and the *admission category*. We discover that 92.3% of students who do not have back or arrear papers archive above *Good* level of the end semester percentage, as shown in Figure 3.15. In addition, the admission type has a strong impact on the final exam with 89.5% of students who paid for the admission earning the high grade, while this ratio in the free admission group is only 65.5%.

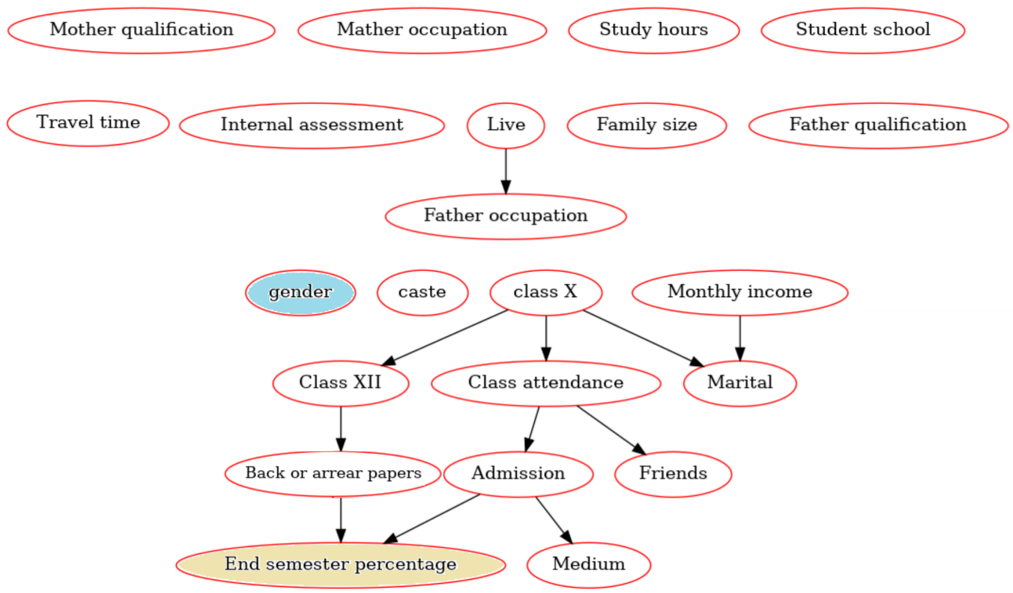
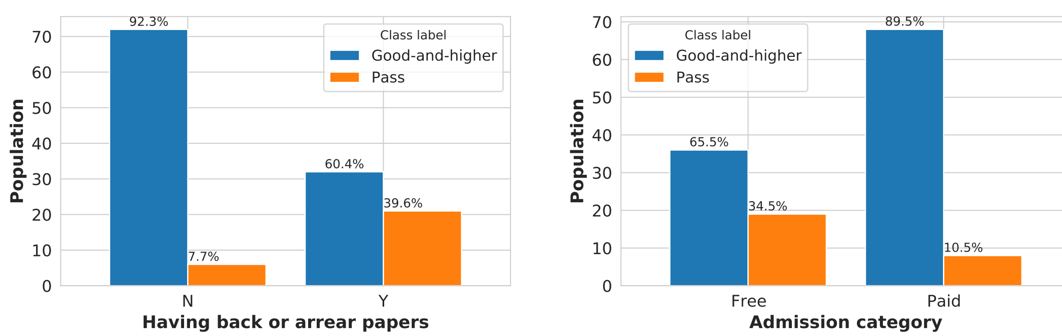


Figure 3.14. Student aca.: Bayesian network (class label: *End semester percentage*, protected attributes: *gender*)



(a) Whether students have back or arrear papers

(b) Admission category

Figure 3.15. Student aca.: Distribution of students w.r.t. the *end semester percentage*

3.3.7 xAPI-Edu-Data dataset

xAPI-Edu-Data[7] [12] is collected from *Kalboard360* LMS at the university of Jordan using experience API (xAPI) web service (a learner activity tracker tool). The dataset is collected through two educational semesters in 2015. The task is to predict students' academic performance [13].

Dataset characteristics: The dataset consists of 480 student records described by

[7] https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

17 attributes (9 categorical, 4 binary, 4 numerical). The attributes are categorized into three major groups: i) demographic attributes such as gender and nationality, ii) academic background attributes such as educational stage, grade level and section, iii) behavioral attributes such as a raised hand in class, opening resources, and school satisfaction. The class attribute is $Class = \{Low\ (L),\ Medium\ (M),\ High\ (H)\}$. To reduce the complexity of the classification problem, we transfer the class attribute into the binary attribute as $Class = \{Low,\ Medium\text{-}High\}$. The positive class is "Medium-High". The dataset (with new class attribute) is imbalanced with an imbalance ratio 2.78:1 (positive:negative). In this dataset, we consider the protected attribute $Gender = \{F,\ M\}$ (F: female, M: male). The majority group is *male* with the ratio off *male:female* is 305:175 (63.5%:36.5%). Table 3.8 provides an overview of all attributes of the dataset.

Table 3.8. xAPI-Edu-Data dataset: attributes characteristics

Attributes	Type	Values	#Missing values	Description
Gender	Binary	{M, F}	0	Gender
Nationality	Categorical	14	0	The nationality of student
PlaceOfBirth	Categorical	14	0	The place of birth of student
StageID	Categorical	3	0	Educational level (lower level, middle school, high school)
GradeID	Categorical	10	0	The grade of student
SectionID	Categorical	3	0	The classroom (A, B, C)
Topic	Categorical	12	0	Course topic (English, French, etc.)
Semester	Categorical	2	0	School year semester (first, second)
Relation	Categorical	2	0	Parent responsible for student (mom, father)
Raisedhands	Numerical	0 - 100	0	How many times the student raises his/her hand
VisitedResources	Numerical	0 - 99	0	How many times the student visits a course content
AnnouncementsView	Numerical	0 - 98	0	How many times the student checks new announcements
Discussion	Numerical	1 - 99	0	How many times the student participate on discussion
ParentAnsweringSurvey	Binary	{Yes, No}	0	Whether parent answered the surveys
ParentschoolSatisfaction	Binary	{Yes, No}	0	Whether the parents are satisfied
StudentAbsenceDays	Binary	$\{<7, >7\}$	0	The number of absence days
Class	Categorical	3	0	The grade's level (low, middle, high)

Bayesian network: We encode the numerical attributes as follows in order to generate the BN: $Raisedhands = \{\leq 30, 31{-}70, > 70\}$, $VisitedResources = \{\leq 30, 31{-}70, > 70\}$, $AnnouncementsView = \{\leq 30, 31{-}70, > 70\}$, $Discussion = \{\leq 30, 31{-}70, > 70\}$. In the BN (Figure 3.16), no relation between the class label and the protected attribute is detected. The class label is conditionally dependent on the number of absence days (*StudentAbsenceDays*) and how many times the student visits a course content (*VisitedResources*). We investigate the relationship of absence days, the number of visits to course resources, and the class label for each gender. As illustrated in Figure 3.17, the vast majority of people with average and high scores are male. In the group of people with absence days greater than 7, only 23 are women, while this number is 93 for men. Interestingly, those with low scores had significantly higher visits to the course material than those with high and middle scores.

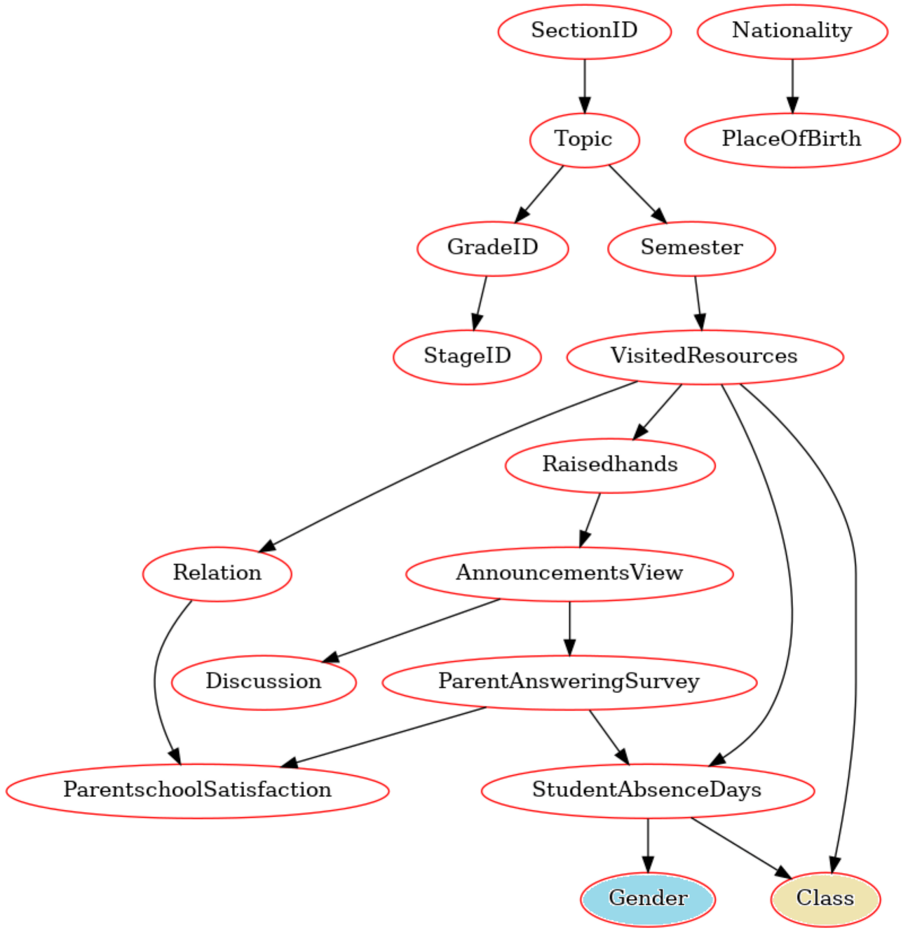


Figure 3.16. xAPI-Edu-Data: Bayesian network (class label: *Class*, protected attributes: *gender*)

3.4 Experimental evaluation

In this section, we present a short fairness-vs-predictive performance evaluation[8] using a popular classification method (namely, *logistic regression*).

3.4.1 Evaluation setup

Predictive model. As our classification model, we use *logistic regression* [48], a statistical model using a logistic function to model a binary dependent variable. We apply the logistic regression model to the binary classification problem to simplify the task.

Fairness measures. Based on the confusion matrix in Figure 3.18 (in which, *prot.*

[8] The source code is available at: https://github.com/tailequy/fairness_dataset

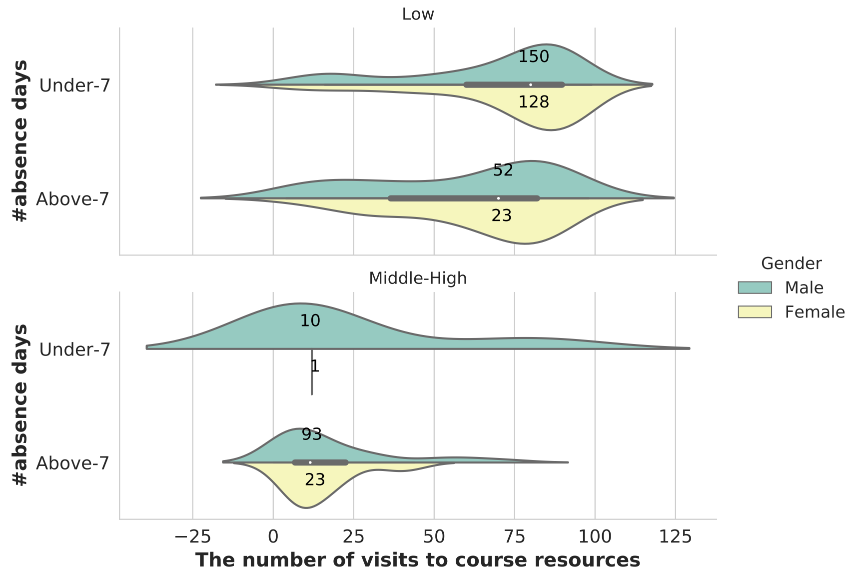


Figure 3.17. xAPI-Edu-Data: Distribution of *the number of absence days*, *the number of visits to course resources* and *class label* w.r.t. *gender*

and *non-prot.* stand for *protected, non-protected*, respectively), we report the performance of the predictive model on the following measures:

		Predicted class	
		Positive +	**Negative -**
Actual class	**Positive +**	True Positive (TP) $TP_{prot.} + TP_{non-prot.}$	False Negative (FN) $FN_{prot.} + FN_{non-prot.}$
	Negative -	False Positive (FP) $FP_{prot.} + FP_{non-prot.}$	True Negative (TN) $TN_{prot.} + TN_{non-prot.}$

Figure 3.18. The confusion matrix, including protected/non-protected groups

- Accuracy (Acc)

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.3}$$

- Balanced accuracy (BA)

$$BA = \frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \tag{3.4}$$

- True positive rate (TPR) on the protected group

$$TPR_{prot.} = \frac{TP_{prot.}}{TP_{prot.} + FN_{prot.}} \tag{3.5}$$

- TPR on non-protected group

$$TPR_{non-prot.} = \frac{TP_{non-prot.}}{TP_{non-prot.} + FN_{non-prot.}} \tag{3.6}$$

- True negative rate (TNR) on the protected group

$$TNR_{prot.} = \frac{TN_{prot.}}{TN_{prot.} + FP_{prot.}} \tag{3.7}$$

- TNR on non-protected group

$$TNR_{non-prot.} = \frac{TN_{non-prot.}}{TN_{non-prot.} + FP_{non-prot.}} \tag{3.8}$$

- Statistical parity (SP) (see Eq. 2.2)
- Equalized odds (EOd) (see Eq. 2.4)
- ABROCA (see Eq. 2.5)

Training/test set spliting. The ratio of the training set and test set in our experiment is 70%:30% (single split) applied for each dataset.

3.4.2 Experimental results

Table 3.9 describes the performance of the logistic regression model on all seven educational datasets. In general, the classification model has different responses to the data sets. The logistic regression model provides the best results in terms of fairness measures on the *OULAD* dataset; however, the predictive performance is very low in comparison with other datasets. The predictive model works well on the *student performance* and *xAPI-Edu-Data* datasets with the values of accuracy and balance accuracy measures above 90%. A trade-off between predictive performance and fairness is observed in the *MOOC* dataset with high accuracy and good fairness

Table 3.9. Predictive- and fairness-related performance of logistic regression model on seven educational datasets (the best values in **bold**)

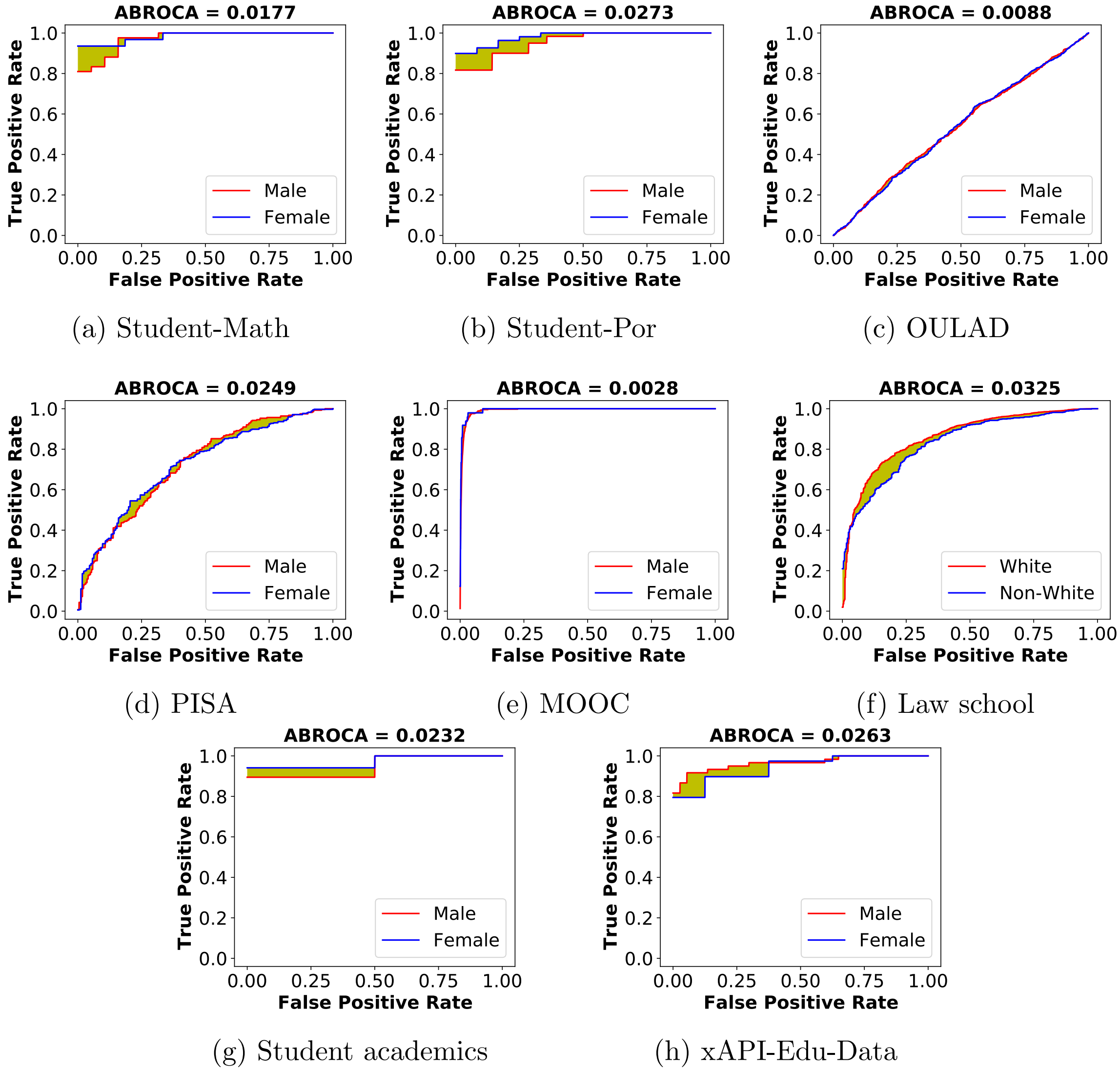
Dataset	Protected attribute	Group distribution (%) $[p_+, p_-, \overline{p}_+, \overline{p}_-]$	Acc	BA	SP	EOd	ABROCA	TPR prot.	TPR non-prot.	TNR prot.	TNR non-prot.
Student-Math	Gender	[33.7, 19.0, 33.4, 13.9]	0.9412	**0.9360**	0.2041	0.1616	0.0177	0.9354	0.9762	0.9630	0.8421
Student-Por	Gender	[51.3, 7.7, 33.3, 7.7]	0.9282	0.8447	-0.0682	0.0490	0.0273	0.9633	0.95	0.75	0.7143
OULAD	Gender	[32.1, 14.2, 35.9, 17.8]	0.6751	0.5	**0.0**	**0.0**	**0.0088**	**1.0**	**1.0**	0.0	0.0
PISA	Male	[31.7, 18.4, 26.7, 23.2]	0.6791	0.6511	-0.2090	0.3652	0.0249	0.8522	0.6895	0.4091	0.6116
MOOC	Gender	[0.4, 8.9, 3.3, 87.4]	**0.9826**	0.8380	0.0053	0.0807	0.0028	0.6122	0.6893	**0.9975**	0.9939
Law school	Race	[11.5, 4.4, 77.5, 6.6]	0.9072	0.6260	0.1937	0.5043	0.0325	0.9100	0.9955	0.5251	0.1063
Student aca.	Gender	[38.9, 6.1, 40.5, 14.5]	0.9	0.8333	-0.0852	0.5464	0.0232	0.9412	0.8947	0.5	**1.0**
xAPI-Edu-Data	Gender	[31.4, 5.0, 42.1, 21.5]	0.9167	0.9091	-0.2634	0.3530	0.0263	0.9487	0.9167	0.6250	0.9459

metrics. We believe that the experimental results can be considered as the baseline for the researchers' future studies.

In addition, we plot the ABROCA slicing of all datasets in Figure 3.19. In the Figure, the *red* ROC curve represents the non-protected group (e.g., Male) while the *blue* ROC is the curve of the protected group (e.g., Female). The best value of the ABROCA is seen in the *MOOC* dataset (ABROCA = 0.0088). The worst cases are the *Law school* and the *Student performance - Portuguese* datasets with ABORCA = 0.0325 and 0.0273, respectively.

3.5 Chapter summary

In this chapter, we investigate and perform the bias-aware analysis of the educational datasets that were collected in many countries around the world. The typical learning task is to predict students' outcomes or grades. All datasets are imbalanced with very different imbalance ratios in terms of class imbalance. Since the bias is observed in the datasets w.r.t. protected attributes, i.e., *gender, race*, fairness-aware algorithms need to take into account these attributes to achieve fairness in education. Besides, bias and discrimination are the common problems of almost all domains in reality. The definition of fairness, of course, is different across domains. It is not easy to evaluate the efficiency of fairness-aware algorithms because they must be based on such fairness notions. Therefore, it is crucial and necessary to select or define the appropriate fairness notions for each problem in each domain, eg. education, because there is no universal fairness notion for every problem. This remains a major challenge for researchers. To partially address this challenge, in Chapter 4, we will evaluate the prevalent group fairness measures in the student performance prediction problems. In addition, because clustering is an important technique for analyzing student data [114] we will investigate the student grouping problems clustering models w.r.t. protected attributes and student's preferences in Chapters 5 and 6.



(a) Student-Math (b) Student-Por (c) OULAD

(d) PISA (e) MOOC (f) Law school

(g) Student academics (h) xAPI-Edu-Data

Figure 3.19. ABROCA slice plot on seven educational datasets

In our exploratory data analysis, the BN of each dataset is generated from the dataset and all continuous numeric data attributes are discretized into meaningful categorical attributes. Therefore, the changes in this discretization process will yield different results on the same data set. The consistent and systematic application of effective data discretization methods [173, 202] might be a suitable approach to deal with this limitation. Moreover, the selection of the protected attributes is also a matter of consideration. In these educational datasets, *gender* is the most popular protected attribute, followed by *age* and *race*. The selection of one or more protected attributes for the experiment depends on many factors such as domain, problem, and the purpose of the experiment. In our preliminary experiments, for each dataset, we only demonstrate the performance of the predictive model w.r.t. one of the most popular protected attributes. In addition, the identification and handling of "proxy" attributes is also an issue that requires more research. In another aspect, an excellent understanding of well-known datasets can also inspire researchers to develop synthetic data generators to overcome the difficulties caused by the lack of benchmark educational datasets.

4 Evaluation of fairness measures in student performance prediction problems

As mentioned in Chapter 2, predicting students' academic performance is one of the key tasks of the EDM community. Moreover, as mentioned in Chapter 3, selecting and defining appropriate fairness measures for a specific problem in education is still a challenge. Therefore, in this chapter, we evaluate different group fairness measures for student performance prediction problems on various educational datasets and fairness-aware learning models. Our study demonstrates the significance of selecting an appropriate fairness measure, as well as the grade threshold when determining whether a student passes or fails the exam.

4.1 Introduction

One of the most important tasks in EDM that attracts great attention is the student performance prediction problem[200]. The early estimation of student learning outcomes can help detect and notify students at risk of academic failure. Besides, it supports institutional administrators in identifying key factors affecting students' grades and providing suitable interventions for outcome improvement. The performance prediction process relies on historical academic records and trains ML algorithms [70, 104, 215] on labeled data [47, 110, 197] to predict students' performance. In addition, endeavoring to reduce biases is important and decisive in the applicability of an ML model in education. As an example, a recent study has proposed approaches that aim at predicting calculated grades of students in England as a replacement for actual grades due to the cancellation of exams during COVID-19 [14]. Unfortunately, the proposal could not be implemented due to the presence of historical biases that have been exposed. This is because the data used to train the machine learning system no longer accurately represents the current reality. Therefore, fairness has become a crucial criterion in designing such systems.

A large variety of fairness measures have been introduced in the ML area. There are more than 20 different fairness measures introduced in the computer science research area [140, 188]. Although fairness is a fundamental concept in education, ensuring equal opportunities for all students in their studies and fair treatment [141], there is a lack of previous research examining the effectiveness of various fairness measures and the process of selecting them within educational contexts.

In this chapter, we conduct a comprehensive study to assess the adequacy of prevalent group fairness measures in student performance prediction problems. Our experiments provide users with a broad view of unfairness from diverse aspects in an educational context. Moreover, the results also guide the selection of suitable fairness measures to evaluate students' grade predictive models. We believe our contributions are crucial to alleviate the burden of choosing fairness measures for consideration and motivate further studies to improve the accuracy and fairness of student performance prediction models.

The rest of the chapter is organized as follows. In Section 4.2, we present some closely related work on fairness-aware ML and student performance prediction problems. Section 4.3 describes the most popular group fairness measures in ML. Next, we conduct quantitative evaluations of predictive models on educational datasets and discuss the choice of suitable fairness metrics in Section 4.4. Finally, we summarize the chapter in Section 4.5.

4.2 Related work

Extensive research efforts have been conducted to provide useful insights into students' performance analysis and prediction [200]. Various ML models were tested on different problem settings. Cortez et al. [47] presented an early study to predict the grades of secondary students in Portuguese and Mathematics classes. Their results showed that good predictive accuracy could be achieved when previous school period grades are available. Similarly, Berhanu et al. [26] employed a decision tree to predict students' performance using the agriculture college dataset. Some studies [20, 215] proposed diverse approaches to forecast students' grades in higher education. Besides, many studies were reviewed in multiple surveys [2, 147, 172, 176]. They highlighted the most frequently employed techniques such as DT, NB, SVM, and neural networks and dominant factors impacting predictive outcomes, i.e., cumulative grade point average, previous grades, classroom attendance, etc.

One of the most well-known fairness measures is *demographic parity*, so-called *statistical parity* [62]. It requires an equal probability of positive predictions in protected and non-protected groups and it fails to ensure individual fairness. To avoid this, Hardt et al. [81] proposed *equalized odds* metric. It measures whether a classifier predicts labels equally well for all values of attributes. Besides, many other popular metrics were introduced and used in fairness ML studies such as *predictive parity*, *predictive equality* [45], *treatment equality* [27], etc. Despite a substantial number of fairness measures, there is no metric that fits all circumstances [69, 140, 188].

Following the evolution of fairness measures, recent studies have attempted to evaluate fairness in an educational context [72]. Anderson et al. [15] conducted two post hoc fairness assessments for existing student graduation prediction models. Renzhe et al. [205] studied different combinations of student data sources for building highly predictive and fair models for predictions of college success. Jiang et al. [96] proposed several strategies to mitigate bias in the LSTM grade prediction model. They report experimental results on TPR, TNR, and accuracy measures.

4.3 Group fairness measures

In this section, we present the most prevalent group fairness notions used in ML. The list of fairness notions[1] is summarized in Table 4.1.

Table 4.1. An overview of group fairness measures

Measures	Proposed by	Year	#Citations
Statistical parity	Dwork et al. [62]	2012	2,367
Equal opportunity	Hardt et al. [81]	2016	2,575
Equalized odds	Hardt et al. [81]	2016	2,575
Predictive parity	Chouldechova et al. [45]	2017	1,430
Predictive equality	Corbett-Davies et al. [46]	2017	878
Treatment equality	Berk et al. [27]	2018	626
ABROCA	Gardner et al. [72]	2019	84

Student performance prediction problems refer to many ML tasks, including clustering, classification, and regression [211]. In this chapter, we consider the student performance prediction problem as a binary classification task [200]. We use the symbols described in Table 2.2 (Chapter 2), in which, class attribute $Y = \{+, -\}$, e.g., $Y = \{pass, fail\}$. Furthermore, we use a confusion matrix (Figure 4.1) to demonstrate the group fairness measures with an example of a dataset with 100 instances (class $Y = \{pass, fail\}$). The protected attribute is "gender", and the protected group is "female"; the distribution of "female":"male" is 46:54. Examples of fairness measures in the following subsections are computed based on this confusion matrix.

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

		Predicted class	
		Positive +	**Negative -**
Actual class	**Positive +**	True Positive (TP) $TP_{prot} + TP_{non-prot}$ **70** (32:38)	False Negative (FN) $FN_{prot} + FN_{non-prot}$ **10** (4:6)
	Negative -	False Positive (FP) $FP_{prot} + FP_{non-prot}$ **9** (4:5)	True Negative (TN) $TN_{prot} + TN_{non-prot}$ **11** (6:5)

Figure 4.1. The confusion matrix with an example

Details of *statistical parity (SP), equalized odds (EOd)* and *ABROCA* measures are already described in Section 2.2.2, Chapter 2. In our example: $SP = \frac{38+6}{54} - \frac{32+4}{46} \approx 0.0322$ and $EOd = |\frac{32}{32+4} - \frac{38}{38+6}| + |\frac{4}{4+6} - \frac{5}{5+5}| \approx 0.1253$. Therefore, we present these fairness measures: *equal opportunity, predictive parity, predictive equality* and *treatment equality* in the following subsections.

4.3.1 Equal opportunity

Equal opportunity (denoted as EO) is proposed by Hardt et al. [81], whereby a binary predicted outcome $\hat{Y}$ satisfies equal opportunity w.r.t. the protected attribute G and the class attribute Y if:

$$P(\hat{Y} = +|\mathcal{P} = p, Y = +) = P(\hat{Y} = +|\mathcal{P} = \overline{p}, Y = +) \tag{4.1}$$

In other words, the protected and non-protected groups should have equal true positive rates (TPR) [140, 188], $TPR = \frac{TP}{TP+FN}$ (i.e., the classifier should give similar results for students of both genders with actual "*pass*" class). A classifier with equal false negative rates (FNR), $FNR = \frac{FN}{TP+FN}$, will also have equal TPR [188]. Equal opportunity can be measured by:

$$EO = |P(\hat{Y} = -|Y = +, \mathcal{P} = \overline{p}) - P(\hat{Y} = -|Y = +, \mathcal{P} = p)| \tag{4.2}$$

The value range: $EO \in [0, 1]$; with 0 standing for no discrimination and 1 indicating maximum discrimination. In our example, $EO = |\frac{38}{38+6} - \frac{32}{32+4}| \approx 0.0253$.

4.3.2 Predictive parity

Predictive parity [45] (denoted as PP) is satisfied if both protected and non-protected groups have an equal positive predictive value (PPV) or *Precision*, $PPV = \frac{TP}{TP+FP}$, i.e., the probability of a student predicted to "*pass*" actually having "*pass*" class should be the same, for both male and female students.

$$P(Y = +|\hat{Y} = +, \mathcal{P} = p) = P(Y = +|\hat{Y} = +, \mathcal{P} = \overline{p}) \tag{4.3}$$

Therefore, we report the predictive parity measure as:

$$PP = |P(Y = +|\hat{Y} = +, \mathcal{P} = p) - P(Y = +|\hat{Y} = +, \mathcal{P} = \overline{p})| \tag{4.4}$$

where $PP \in [0, 1]$, with 0 standing for no discrimination and 1 indicating the maximum discrimination. $PP = \frac{32}{32+4} - \frac{38}{38+5} \approx 0.0052$, in our example.

4.3.3 Predictive equality

Predictive equality [46] (denoted as PE), also referred as false positive error (FPR) rate balance [45] ($FPR = \frac{FP}{TN+FP}$), aims to the equality of decision's accuracy across the protected and non-protected groups. In detail, 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 students.

$$P(\hat{Y} = +|Y = -, \mathcal{P} = p) = P(\hat{Y} = +|Y = -, \mathcal{P} = \overline{p}) \tag{4.5}$$

In practice, researchers report predictive equality measure by the difference of $FPRs$ [94]:

$$PE = |P(\hat{Y} = +|Y = -, \mathcal{P} = p) - P(\hat{Y} = +|Y = -, \mathcal{P} = \overline{p})| \tag{4.6}$$

The value range: $PE \in [0, 1]$, 0 and 1 indicate no discrimination and maximum discrimination, respectively. $PE =| \frac{4}{6+4} - \frac{5}{5+5} |= 0.1$, in our example.

4.3.4 Treatment equality

Treatment equality [27] (denoted as TE) is satisfied if the ratios of false negatives and false positives are the same for both protected and non-protected groups.

$$\frac{FN_{prot.}}{FP_{prot.}} = \frac{FN_{non-prot.}}{FP_{non-prot.}}. \tag{4.7}$$

In our paper, we report the treatment equality by the difference between two ratios described in Eq.4.7.

The metric becomes unbounded if $FP_{prot.}$ or $FP_{non-prot.}$ is zero[2]. In our example, $TE = -0.2$, because the ratios of FN and FP are 1 and 1.2 for female and male groups, respectively.

4.4 Evaluation

In this section, we evaluate the performances of predictive models w.r.t. accuracy and fairness measures on five datasets and investigate the effect of the grade threshold on fairness and predictive performance. The datasets are depicted in Section 4.4.1; the predictive models are described in Section 4.4.2 and Section 4.4.3 presented our experimental results.

4.4.1 Datasets

We evaluate the fairness measures on popular educational datasets [119, 142, 200], which are summarized in Table 4.2. All datasets are imbalanced, as shown in the IR column. Detailed descriptions of these datasets are provided in Chapter 3.

Table 4.2. An overview of five educational datasets used for the evaluation

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

4.4.2 Predictive models

We select four prevalent classifiers used for student performance prediction problems based on the survey of Xiao et al. [200], and two well-known fairness-aware classifiers, namely Agarwal's [4] and AdaFair [94]. Agarwal's method reduces the fair classification to a sequence of cost-sensitive classification problems with the lowest (empirical) error subject to the desired constraints, and AdaFair is based on AdaBoost that iteratively adjusts the weights of instances in each boosting round by considering a cumulative notion of fairness, which is influenced by the entire ensemble of current members.

In brief, in this evaluation, we use the following predictive models:

- Decision tree (DT)

[2]https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-post-training-bias-metric-te.html

- Naïve Bayes (NB)
- Multi-layer perceptron (MLP)
- Support vector machines (SVM)
- Agarwal's fairness-aware classification model (Agarwal's)
- AdaFair fairness-aware classification model (AdaFair)

Because our primary objective is to evaluate the group fairness measures, therefore, we simplify the splitting and hyperparameter turning processes. In detail, we utilize 70% of the data for training and 30% for testing using the single split strategy. Predicted models are executed with default parameters provided by Scikit-learn and Iosifidis and Ntoutsi [94]. Agarwal's fairness-aware model is implemented in the AI Fairness 360 (AIF360) toolkit[3].

4.4.3 Experimental results

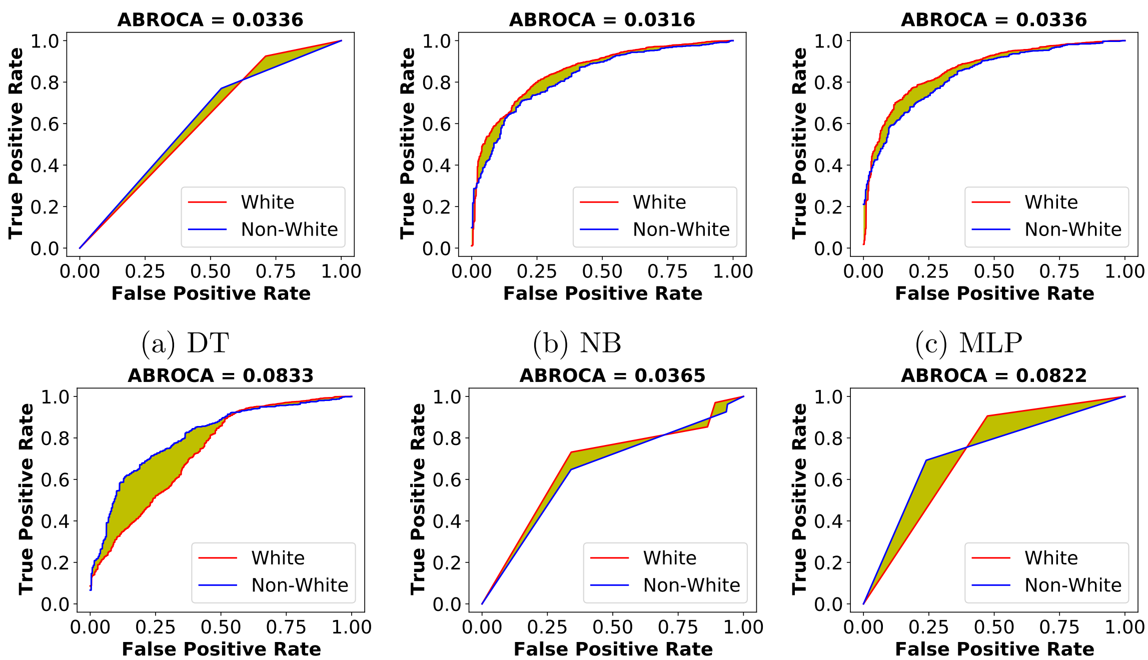
Law school dataset. Because the datasets are imbalanced, we report the performance of predictive models on both accuracy and balanced accuracy measures (Table 4.3). It is readily apparent that fairness is not achieved in the majority of traditional predictive models, as these models neglect to consider fairness in their objective function. In contrast, fairness-aware models show relatively good performances on most fairness measures. AdaFair is the best predictive model w.r.t. most fairness measures, although its balanced accuracy is significantly lower than that of other models. Moreover, the fairness measures show a quite large variation across the classification models, as demonstrated in Figure 4.7-a. *Equalized odds* is the measure with the highest variability, followed by *statistical parity* and *predictive equality* while *ABROCA* is relatively stable across all predictive models. Furthermore, the shape and position of the ROC curves, as visualized in Figure 4.2, have been changed across the predictive models, which indicates the change in the performance of models w.r.t. each group of the protected attribute. Interestingly, all fairness-aware models achieve equalized odds at some thresholds of TPR and FPR where the two ROC curves of protected and non-protected groups intersect (Figure 4.2-e, f). Because traditional predictive models do not consider fairness constraints, these ROC curves only intersect on some models, such as SVM and DT.

PISA dataset. The interesting point is SVM and DT show their superiority in terms of fairness measures, although AdaFair still has very good results on fairness metrics and accuracy (Figure 4.3 and Table 4.4). A possible explanation is that the intersection between ROC curves occurs several times in DT, MLP and SVM models, i.e., the predictive models achieve equalized odds [72], at some thresholds of FPR and

[3] https://github.com/Trusted-AI/AIF360

Table 4.3. Law school: performance of predictive models

Measures	DT	NB	MLP	SVM	Agarwal's	AdaFair
Accuracy	0.8458	0.8191	**0.9042**	0.8926	0.7952	0.8921
Balanced accuracy	0.6301	**0.7784**	0.6596	0.5029	0.5848	0.5
Statistical parity (SP)	0.1999	0.5250	0.2367	0.0052	0.0326	**0.0**
Equal opportunity (EO)	0.1557	0.4665	0.1237	0.0014	0.0202	**0.0**
Equalized odds (EOd)	0.3253	0.8105	0.5501	0.0169	0.0953	**0.0**
Predictive parity (PP)	0.1424	**0.0130**	0.0754	0.1857	0.1802	0.1885
Predictive equality (PE)	0.1696	0.3440	0.4265	0.0154	0.0751	**0.0**
Treatment equality (TE)	-0.0667	22.440	0.7770	0.0039	-1.9676	**0.0**
ABROCA	0.0336	**0.0316**	0.0336	0.0833	0.0365	0.0822



(a) DT (b) NB (c) MLP

(d) SVM (e) Agarwal's (f) AdaFair

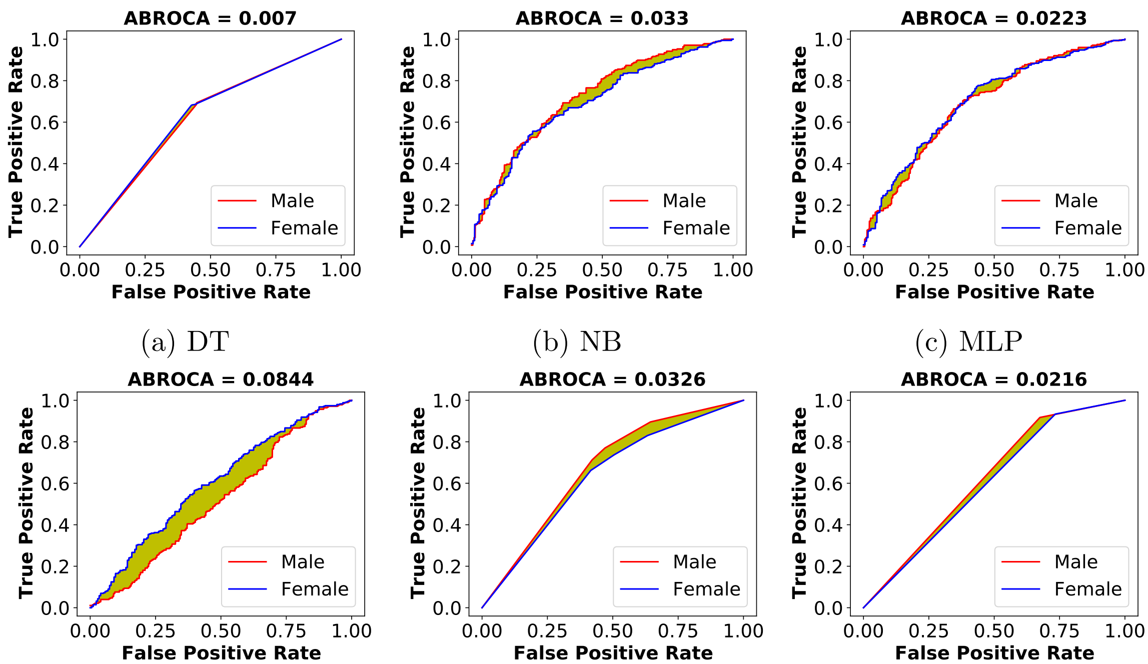
Figure 4.2. Law school: ABROCA slice plots

TPR. We create the class label for this dataset based on the *reading score* attribute (Section 3.3.3, Chapter 3). It might be the reason to explain why all prediction models give pretty low accuracy. Furthermore, fairness measures have the least variability in this dataset, as shown in Figure 4.7-b.

Student aca. dataset. The AdaFair outperforms other models w.r.t. fairness measures, however, the balanced accuracy is decreased considerably (Table 4.5). Besides, all fairness measures significantly vary across predictive models (Figure 4.7-c). *Equalized odds* and *predictive equality* are the two fairness measures with the highest variation. This can be explained by the ABROCA plots in Figure 4.4 when most

Table 4.4. PISA: performance of predictive models

Measures	DT	NB	MLP	SVM	Agarwal's	AdaFair
Accuracy	0.6360	0.6624	0.6526	0.6096	0.6614	**0.6810**
Balanced accuracy	0.6224	**0.6379**	0.5732	0.5026	0.6340	0.6130
Statistical parity	-0.0200	-0.0316	-0.0771	**-0.0022**	-0.0096	-0.0573
Equal opportunity	**0.0019**	0.0262	0.0330	0.0043	0.0414	0.0164
Equalized odds	0.0165	0.0709	0.1398	**0.0068**	0.0548	0.0752
Predictive parity	0.1012	**0.0683**	0.0826	0.1108	0.0785	0.0868
Predictive equality	0.0146	0.0446	0.1067	**0.0024**	0.0134	0.0588
Treatment equality	0.5642	0.3855	-0.0251	**-0.0033**	0.4609	0.0260
ABROCA	**0.0070**	0.0330	0.0223	0.0844	0.0326	0.0216



(a) DT (b) NB (c) MLP

(d) SVM (e) Agarwal's (f) AdaFair

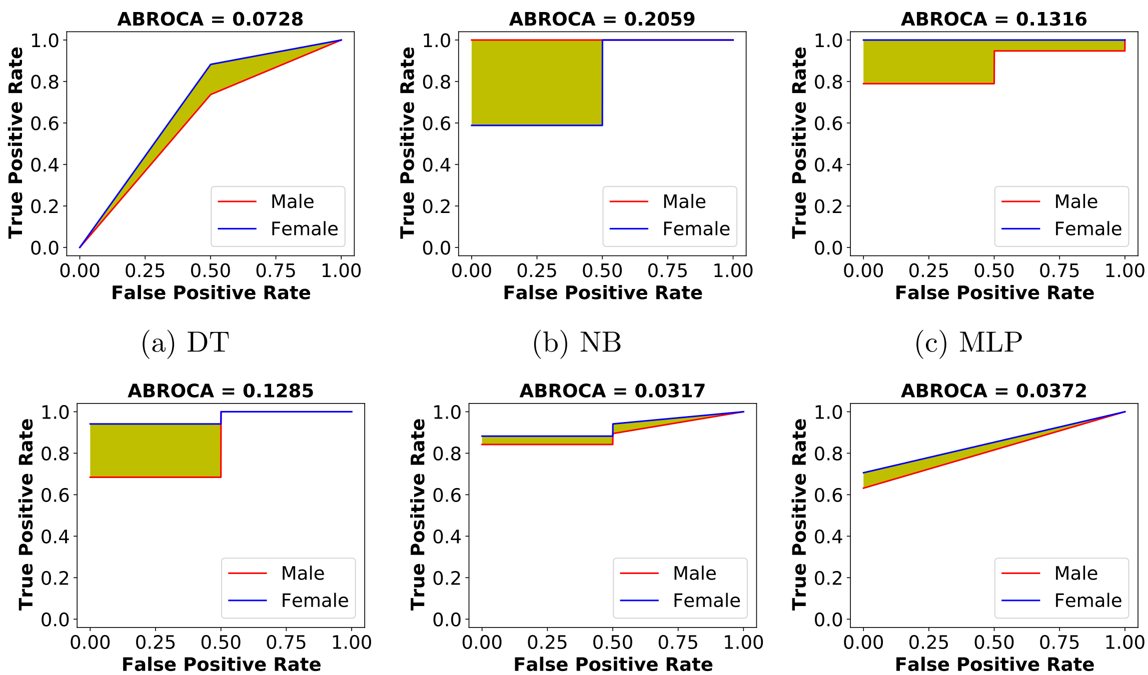
Figure 4.3. PISA: ABROCA slice plots

predictive models are unable to achieve *equalized odds* when the FPR is below 0.5.

Student-Por dataset. In general, all models show good accuracy (balanced accuracy) in predicting students' performance (Table 4.6). MLP and AdaFair models fairly guarantee the fairness of results on most measures because AdaFair takes into account the fairness constraint and an intersection between two ROC curves is observed in both these models (Figure 4.5-d, f). Equalized odds is also achieved by the MLP model when the ROC curves intersect (Figure 4.5-c). However, the values of fairness measures also do not vary significantly across predictive models (Figure 4.7-d), although the ABROCA slices are quite different in shape (Figure 4.5).

Table 4.5. Student aca.: performance of predictive models

Measures	DT	NB	MLP	SVM	Agarwal's	AdaFair
Accuracy	0.7750	0.8750	0.8750	**0.9250**	0.8750	0.9
Balanced accuracy	0.6528	**0.8194**	**0.8194**	0.6250	**0.8194**	0.5
Statistical parity	-0.1278	-0.1328	-0.1328	0.0526	0.0677	**0.0**
Equal opportunity	0.1455	0.0991	0.2105	**0.0**	0.0123	**0.0**
Equalized odds	0.1455	0.5991	0.7105	0.5	0.5124	**0.0**
Predictive parity	**0.0042**	0.0588	0.0552	0.0397	0.0556	0.01
Predictive equality	**0.0**	0.5	0.5	0.5	0.5	**0.0**
Treatment equality	-3.0	*N/A*	*N/A*	**0.0**	*N/A*	**0.0**
ABROCA	0.0728	0.2059	0.1316	0.1285	**0.0317**	0.0372



(a) DT (b) NB (c) MLP

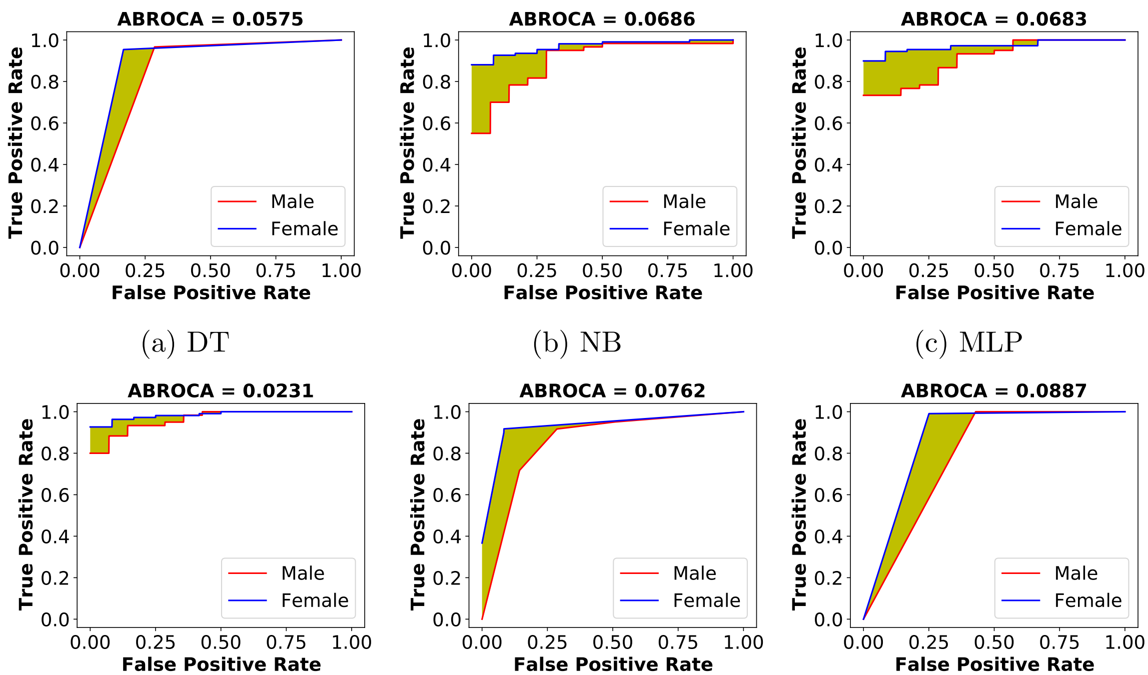
(d) SVM (e) Agarwal's (f) AdaFair

Figure 4.4. Student aca.: ABROCA slice plots

xAPI-Edu-Data dataset This is a surprising dataset because the traditional classification methods show a better performance not only in terms of accuracy or balanced accuracy measures but also fairness measures (Table 4.7). ROC curves that intersect (one or more times) in most models (both traditional and fairness-aware models) could be the explanation for the results. In addition, the variation in the values of fairness measures across the predictive models is not significant, as shown in Figure 4.7-e, except for the ABROCA measure with a noticeable change in the shape (Figure 4.6).

Table 4.6. Student-Por: performance of predictive models

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



(a) DT (b) NB (c) MLP

(d) SVM (e) Agarwal's (f) AdaFair

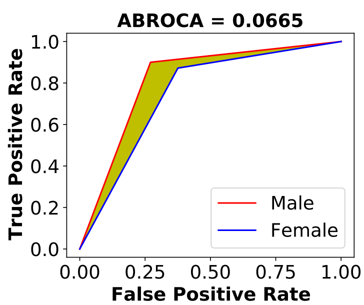
Figure 4.5. Student-Por: ABROCA slice plots

Regarding the *treatment equality* measure, this measure is entirely different from all other measures with an extensive range of values, which is visualized in Figure 4.7-f. In Figure 4.7, we use the abbreviations of the fairness measures and datasets. On the *PISA* datasets, this TE measure shows the best values across predicted models, followed by *Law school* and *Student aca.* datasets.

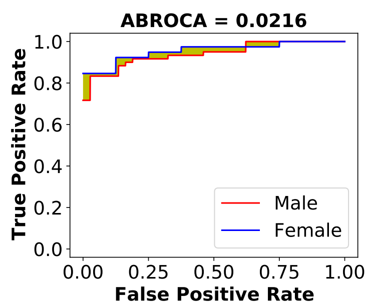
Summary of results: In general, *ABOCA* is the measure with the lowest variability across predictive methods and datasets. It also clearly presents the ML model's accuracy variation over each value of the protected attribute. *Equal opportunity* and

Table 4.7. xAPI-Edu-Data: performance of predictive models

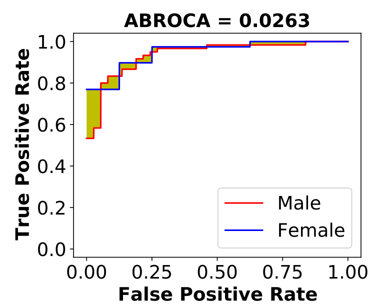
Measures	DT	NB	MLP	SVM	Agarwal's	AdaFair
Accuracy	0.8333	0.8750	0.8750	0.8611	**0.8681**	0.8056
Balanced accuracy	0.8	**0.8970**	0.8545	0.8505	0.8859	0.8162
Statistical parity	**-0.1274**	-0.2608	-0.2112	-0.2209	-0.2505	-0.2292
Equal opportunity	0.0282	0.0974	0.0654	**0.0308**	0.0974	0.0538
Equalized odds	0.1329	0.1954	**0.1262**	0.2706	0.1684	0.3207
Predictive parity	0.0752	0.0074	0.0654	0.0088	0.0122	**0.0057**
Predictive equality	0.1047	0.0980	**0.0608**	0.2399	0.0709	0.2669
Treatment equality	1.0667	-8.0	**0.0**	-0.2667	-2.0	-1.1667
ABROCA	0.0665	**0.0216**	0.0263	0.0796	0.0293	0.1065



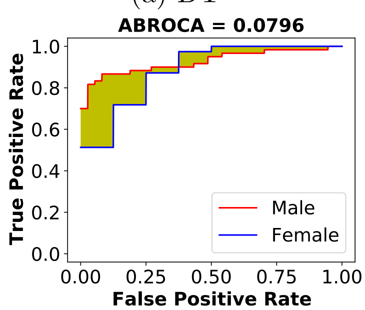
(a) DT



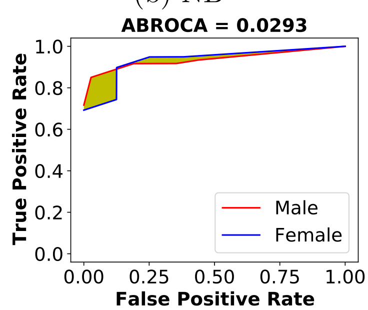
(b) NB



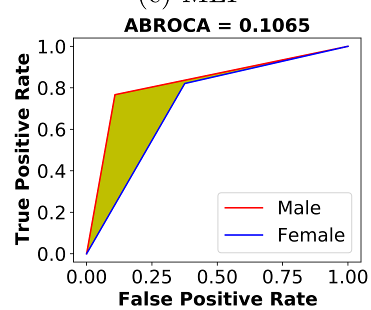
(c) MLP



(d) SVM



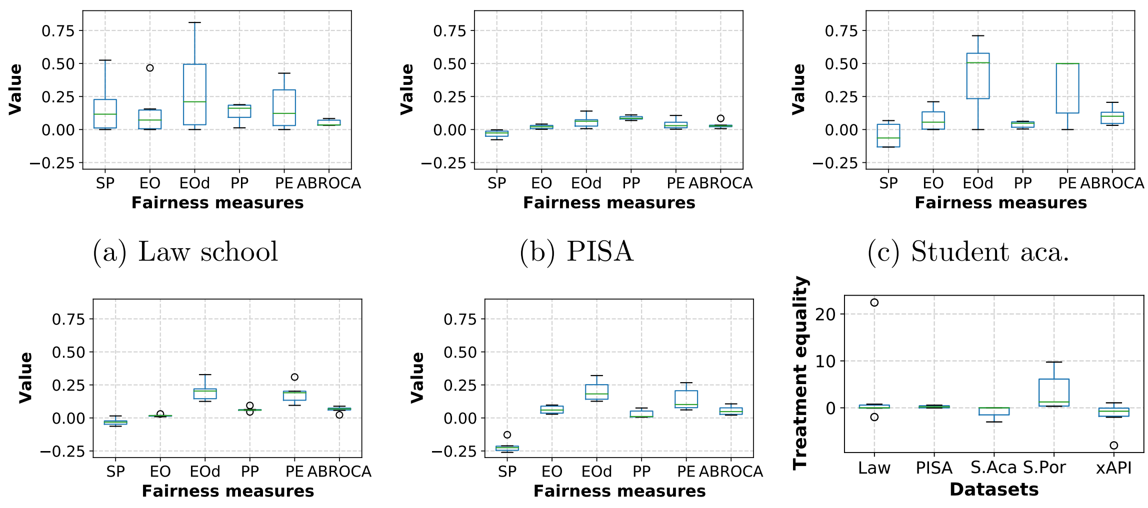
(e) Agarwal's



(f) AdaFair

Figure 4.6. xAPI-Edu-Data: ABROCA slice plots

predictive parity also have a slight variation across methods and datasets. *Equalized odds*, to some extent, can represent two measures *equal opportunity* and *predictive equality* as it is the sum of the other two metrics. Furthermore, *treatment equality* has a very wide range of values (sometimes the value may not be bounded), making it difficult to compare and evaluate.



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

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

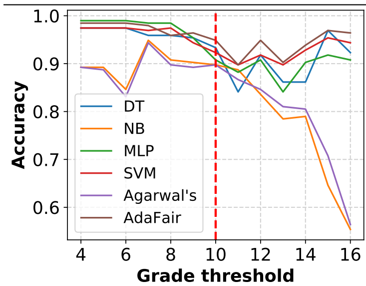
Figure 4.7. Variation of fairness measures

4.4.4 Effect of varying grade threshold on fairness

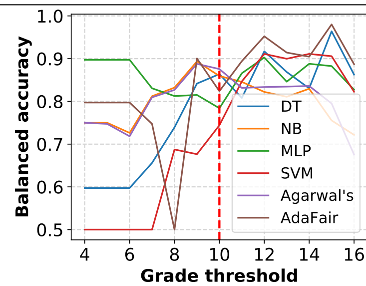
Grade thresholds are often chosen as a basis for determining whether a candidate passes or fails an exam, i.e., a passing grade. In the student-Por dataset, 10 (out of 20) is selected as the grade threshold [47, 119]. However, the selection of a threshold can affect the fairness of the predictive models, as shown in the *Integrated Public Use Microdata Series* (IPUMS) adult dataset [57]. Hence, we investigate the effect of grade threshold on fairness by varying the threshold in a range of [4, 16], corresponding to 25% to 75% of the maximum grade (20). The results in Figure 4.8 show that all fairness measures are affected by the grade threshold. When the grade threshold is gradually increased, the predictive models tend to be fairer (shown on the measures: equalized odds, predictive equality, and ABROCA). The opposite trend is observed in the remaining measures (except the treatment equality measure). The variation of the gender ratio (male/female) within the classes (Figure 4.9) could be the explanation for this result. Regarding balanced accuracy, two models (SVM and AdaFair) tend to predict more accurately. The NB model has a decreasing accuracy when the grade threshold is increased. Agarwal's model exhibits the same trend because it utilizes a Gaussian NB as an estimator (Figure 4.8-a, b).

4.5 Chapter summary

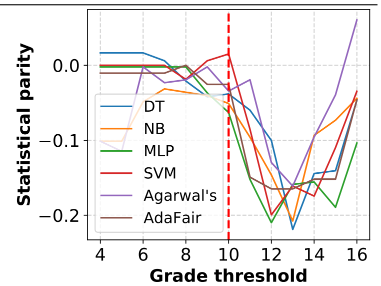
In this chapter, we evaluate seven popular group fairness measures for student performance prediction problems. We conduct experiments using four traditional ML models and two fairness-aware ML methods on five educational datasets. Our exper-



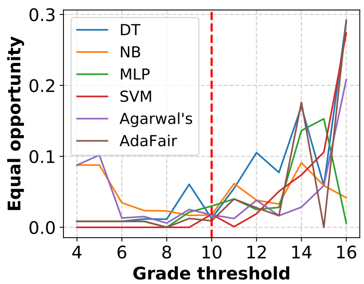
(a) Accuracy



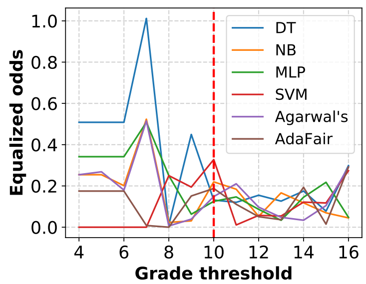
(b) Balanced accuracy



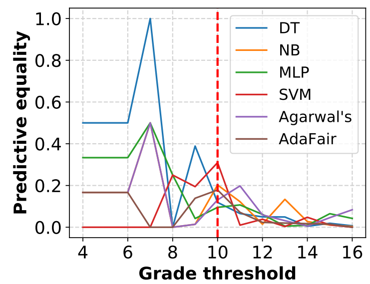
(c) Statistical parity



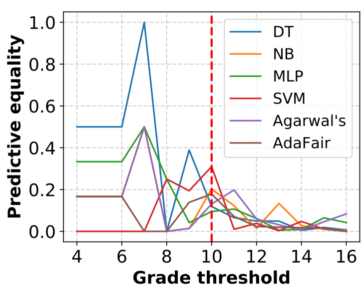
(d) Equal opportunity



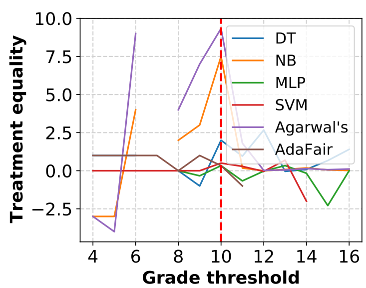
(e) Equalized odds



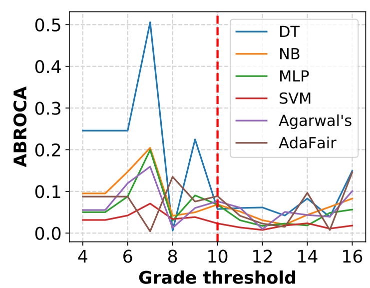
(f) Predictive parity



(g) Predictive equality



(h) Treatment equality



(i) ABROCA

Figure 4.8. Accuracy and fairness interventions with varying grade threshold on Student-Por dataset

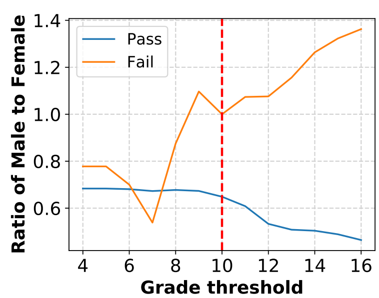


Figure 4.9. Gender ratio by classes with different grade thresholds

iments reflect varying behaviors of fairness measures across datasets and predictive models. The results provide an overview picture for the selection of fairness measures in a specific case. Besides, we investigate the effect of varying grade thresholds on the accuracy and fairness of ML models. The preliminary results suggest that choosing the grade threshold is an important factor contributing to ensuring fairness in the output of the ML models. In another aspect, users should also consider choosing efficient classification models, which can be traditional models, like SVM, MLP, or fairness-aware models, like AdaFair or Agarwal's. Moreover, we acknowledge that hyperparameter tuning was not conducted during the experiments; consequently, the impact of these hyperparameters will remain an area for further research. In addition, we plan to extend our evaluation of fairness w.r.t. more than one protected attribute, such as gender and race, and further explore the correlation between groups of fairness notions.

5 Fair-capacitated clustering

As mentioned, student performance prediction and student grouping are important EDM tasks. In Chapter 4, we studied group fairness measures in student performance prediction problems. In this chapter, we investigate the student grouping problem in collaborative learning. Traditionally, clustering algorithms focus on partitioning students into groups of similar ones. The similarity objective, however, is not sufficient in applications where a fair-representation of the groups in terms of protected attributes like gender or race, is required for each cluster. Studies also indicate that students might learn better in diverse student groups. Moreover, in many applications, to make the clusters useful for the end-user, a balanced cardinality among the clusters is required. To deal with this problem, we introduce the *fair-capacitated clustering* problem that partitions the data into clusters of similar instances while ensuring cluster fairness and balancing cluster cardinalities. Then, we propose a two-step solution to the problem: i) we rely on fairlets to generate minimal sets that satisfy the fair constraint and ii) we propose two approaches, namely hierarchical clustering and partitioning-based clustering, to obtain fair-capacitated clustering. The hierarchical approach embeds the additional cardinality requirements during the merging step while the partitioning-based approach alters the assignment step using a knapsack problem formulation to satisfy the additional requirements. We perform experiments on four educational datasets to evaluate our proposed methods. The experimental results show that our approaches deliver well-balanced clusters in terms of both fairness and cardinality while maintaining a good clustering quality.

5.1 Introduction

In educational activities, group assignments [68] and student team achievement divisions [186] are important tools in teaching and learning tasks to help students work together towards shared learning goals. In pedagogical contexts, the assignment to groups (clusters) can be left to the learners or can be done by the teacher or project manager. In the latter case, typical criteria are interests, prior knowledge, problem-solving ability, and also age and communication skills. Better communication, higher-order thinking, and conflict management are several examples of the advantages of group assignments [68]. Clustering algorithms are effective solutions for partitioning students into groups of *similar instances* [28, 149] in collaborative learning. Traditional algorithms, however, focus solely on the similarity objective and do not consider the fairness of the resulting clusters w.r.t. protected attributes like gender or race. However, studies indicate that students might learn better in diverse student groups, e.g., mixed-gender groups [77, 208]. Lately, fair-clustering solutions have been proposed, e.g., , which aim to discover clusters with a fair representation regarding some protected attributes. Therefore, this chapter focuses on fairness in clustering, i.e., the balance of members in cluster w.r.t. protected attributes.

In a teaching situation, one is often interested in certain group sizes which are usually between 2–4 students per group in primary, secondary and university education but might be much larger in adult education and MOOCs. It is obvious that the size of the groups should be comparable to allow for a fair allocation of work among students. Again, traditional clustering algorithms do not consider this requirement, and as a result, clusters of varying sizes might be extracted, reducing the usefulness and applicability of the partitioning for the end-user/teacher. This leads to the demand for clustering solutions that also take into account the size of the clusters. The problem is known as *the capacitated clustering problem (CCP)* [145], which aims to extract clusters with a limited capacity[1] while minimizing the total dissimilarity in the clusters. Capacitated clustering is useful in many applications, e.g., transferring goods/services from the service providers (post office, stores, etc.), garbage collection and sales-force territorial design [150] to various customer locations [73]. To the best of our knowledge, no solution exists that considers both the fairness and capacity of clusters on top of the similarity objective.

To this end, we propose a new problem, the so-called *fair-capacitated clustering* that ensures fairness and balanced cardinalities of the resulting clusters. We decompose the problem into two sub-problems: i) the fairness-requirement compliance step that preserves fairness at a minimum threshold of balance score and ii) the capacity-requirement compliance step that ensures clusters of comparable sizes. For the first step, we generate fairlets [44], which are minimal sets that satisfy fair representation w.r.t. a protected attribute while approximately preserving the clustering objective. For the second step, we propose two solutions for two different clustering types, namely hierarchical and partitioning-based clustering, that consider the capacity constraint during the merge step (for the hierarchical approach) and during the assignment step (for the partitioning approach). Experimental results, on four real datasets from the education domain, show that our methods result in fair and

[1] We use the terms cluster capacity, cluster size and cluster cardinality interchangeably to refer to the number of instances in a cluster.

capacitated clusters while preserving the clustering quality.

The rest of the chapter is structured as follows: Section 5.2 overviews the related work. The fair-capacitated clustering problem is introduced in Section 5.3. Section 5.4 describes our proposed approaches and section 5.5 presents the details of experimental evaluation on educational datasets. Finally, we summarize the chapter in Section 5.6.

5.2 Related work

Chierichetti et al. [44] introduced the fair clustering problem with the aim of ensuring equal representation for each protected attribute, such as gender, in every cluster. In their formulation, each instance is assigned one of two colors (red, blue). They proposed a two-phase approach: clustering all instances into fairlets - small clusters preserving the fairness measure and then applying vanilla clustering methods (k-center, k-median) on those fairlets. Experimental results show that their method can maintain the fairness of clusters; however, finding the fairlet decomposition may introduce a computational bottleneck.

Subsequent studies focus on generalization and scalability. Backurs et al. [21] presented an approximate fairlet decomposition algorithm that can formulate the fairlets in nearly linear time thus tackling the efficiency bottleneck of Chierichetti et al.'s approach. Rösner and Schmidt [171] generalized the fair clustering problem to more than two protected attributes. A more generalized and tunable notion of fairness for clustering was introduced by Bera et al. [25]. They did the experiments on five datasets from the UCI repository with three clustering methods: k-center, k-median, and k-mean. Chen et al. [40] proposed a new definition of fairness as proportionality. According to their theory, to cluster n points with k centers, any n/k points could form their own cluster. Anshuman and Prasant [43] introduced a fair hierarchical agglomerative clustering method for multiple protected attributes.

CCP - a combinatorial optimization problem (see Section 2.3.3, Chapter 2), was first introduced by Mulvey and Beck [145] who proposed solutions using heuristic and sub-gradient algorithms. Several approaches exist to improve the efficiency of solutions or CCP approaches for different cluster types. Khuller and Sussmann [105], for example, introduced an approximation algorithm (approximation factors of 5 and 6) for the capacitated k-center problem. Geetha et al. [73] improved the k-Means algorithm for CCP by using a priority measure to assign points to their centroid. Li et al. [122] showed a $(6 + 10\alpha)$-approximation algorithm for the hard uniform capacitated k-median problem. Lam and Mittenthal [111] proposed a heuristic hierarchical clustering method for CCP to solve the multi-depot location-routing problem.

In this chapter, we introduce the *fair-capacitated clustering* problem which builds upon notions from *fair* clustering and *capacitated* clustering. In particular, we build upon the notion of *fairlets* [44] to extract the minimal sets that preserve fairness. We follow the formulation of [145] to ensure balanced cluster cardinalities (CCP) . To the best of our knowledge, the combined problem has not been studied before and as already discussed, comprises a useful tool in many domains like education.

5.3 Problem definition

Let $X \in \mathbb{R}^d$ be a set of n instances to be clustered and let $dist() : X \times X \to \mathbb{R}$ be the distance function. For an integer k we use $[k]$ to denote the set $\{1, 2, ..., k\}$. A k-clustering $\mathcal{C}$ is a partition of X into k disjoint subsets, $\mathcal{C} = \{C_1, C_2, ..., C_k\}$, called *clusters* with $S = \{s_1, s_2, ..., s_k\}$ be the corresponding cluster centers.

The goal of clustering is to find an *assignment*[2] $\phi : X \to [k]$ that minimizes the objective function:

$$\mathcal{L}(X, \mathcal{C}) = \sum_{s_j \in S} \sum_{x \in C_j} dist(x, s_j) \tag{5.1}$$

As shown in Eq. 5.1, the goal is to find an assignment that minimizes the sum of distances between each point $x \in X$ and its corresponding cluster center $s_j \in S$. It is clear that such an assignment optimizes for similarity but does not consider the fairness or capacity of the resulting clusters.

Capacitated clustering: The goal of a vanilla capacitated clustering [145] is to discover clusters of given capacities while still minimizing the distance objective $\mathcal{L}(X, \mathcal{C})$. The capacity constraint is defined as an upper bound q_j on the cardinality of each cluster C_j:

$$|C_j| \leq q_j \tag{5.2}$$

Clustering fairness: We assume the existence of a binary protected attribute $\mathcal{P} \in \{p, \overline{p}\}$, e.g., $\mathcal{P}$ = "gender" = {female, male}, and p = "female", $\overline{p}$ = "male". Let $\psi : X \to \mathcal{P}$ denote the demographic group to which the point belongs, i.e., "male" or "female". Then, the fairness of a cluster is evaluated in terms of the *balance score* [44] as the minimum ratio between two groups (recall Eq. 2.13)

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

Fairness of a clustering $\mathcal{C}$ equals to the balance of the least balanced cluster $C_j \in \mathcal{C}$ (recall Eq. 2.14):

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

We now introduce the *fair-capacitated clustering* problem that combines all aforementioned objectives regarding distance, fairness, and cardinality.

[2]We focus on hard clustering where each instance is only assigned to one cluster, i.e., hard clustering.

Definition 5.1. ***Fair-capacitated clustering problem***
We define the problem of (t, k, q)*-fair-capacitated clustering as finding a clustering* $\mathcal{C} = \{C_1, \cdots C_k\}$ *that partitions the dataset* X *into* $|\mathcal{C}| = k$ *clusters such that:*

i) The cardinality of each cluster $C_j \in \mathcal{C}$ *does not exceed a threshold* q*, i.e.,* $|C_j| \leq q$ *(**the capacity constraint**);*

ii) The balance of each cluster is at least t*, i.e., balance*$(\mathcal{C}) \geq t$ *(**the fairness constraint**);*

*iii) **The objective function*** $\mathcal{L}(X, \mathcal{C})$ *is minimized.*

Parameters t, k, q *are user-defined referring to the number of clusters, minimum balance threshold, and maximum cluster capacity, respectively.*

5.4 Fair-capacitated clustering

In this section, we propose two (t, k, q) fair-capacitated clustering approaches, one for hierarchical clustering and the second for partitioning-based clustering. We adopt the heuristic approaches to solve the fair-capacitated clustering problem because clustering and fairness-aware clustering are NP-hard problems [11, 44, 109], and finding an optimal fairlet decomposition is NP-hard [44]. Therefore, heuristic methods are a good way to reach the solution in a polynomial time.For both cases, we decompose the complex problem into two simpler sub-problems: i) the *fairlet decomposition* step that divides the original points into a set of points, the so-called *fairlets*, each preserving a balance score subject to the balance threshold t (Section 5.4.1) and ii) the *final clustering* step that clusters these fairlets into k final clusters so that the cardinality constraint subject to the cardinality threshold q is met. Step (ii) depends on the clustering type: for hierarchical clustering, the merge step needs to be changed (Section 5.4.2), whereas for partitioning-based clustering the assignment step needs to change (Section 5.4.3).

5.4.1 Fairlet decomposition

Traditionally, the vanilla versions of clustering algorithms are not capable of ensuring fairness because they assign the data points to the closest center without fairness consideration. Hence, if we could divide the original data set into subsets such that each of them satisfies the balance threshold t then grouping these subsets to generate the final clustering would still preserve the fairness constraint. Each fair subset is defined as a fairlet. We follow the definition of fairlet decomposition by Chierichetti et al. [44].

Definition 5.2. ***Fairlet decomposition***
Suppose that balance$(X) \geq t$ *with* $t = f/m$ *for some integers* $1 \leq f \leq m$*, such that the greatest common divisor* $gcd(f, m) = 1$*. A decomposition* $\mathcal{F} = \{F_1, F_2, ..., F_l\}$ *of* X *is a fairlet decomposition if:*

i) Each point $x \in X$ *belongs to exactly one fairlet* $F_i \in \mathcal{F}$*;*

ii) $|F_i| \leq f + m$ *for each* $F_i \in \mathcal{F}$*, i.e., the size of each fairlet is small;*

iii) For each $F_i \in \mathcal{F}$*, balance*$(F_i) \geq t$*, i.e., the balance of each fairlet satisfies the threshold* t*.*

Each F_i *is called a fairlet.*

By applying fairlet decomposition on the original dataset X, we obtain a set of fairlets $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$. For each fairlet F_i we randomly select a point $r_i \in F_i$ as the *center*. For a point $x \in X$, we denote $\gamma : X \to [1, l]$ as the index of the mapped fairlet.

The second step is to cluster the set of fairlets $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$ into k final clusters, subject to the cardinality constraint. The clustering process is described below for the hierarchical clustering type (Section 5.4.2) and for the partitioning-based clustering type (Section 5.4.3). Clustering results in an assignment from fairlets to final clusters: $\delta : \mathcal{F} \to [k]$. The final fair-capacitated clustering $\mathcal{C}$ can be determined by the overall assignment function $\phi(x) = \delta(F_{\gamma(x)})$, where $\gamma(x)$ returns the index of the fairlet to which x is mapped.

5.4.2 Fair-capacitated hierarchical clustering

Given the set of fairlets: $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$, let $W = \{w_1, w_2, \ldots, w_l\}$ be their corresponding weights, where the weight w_i of a fairlet F_i is defined as its cardinality, i.e., number of data points in F_i.

Traditional agglomerative clustering approaches merge the two closest clusters, so rely solely on similarity. We extend the merge step by also ensuring that merging does not violate the cardinality constraint w.r.t. the cardinality threshold q.

Theorem 1. *The balance score of a cluster formed by the union of two or more fairlets is at least* t*.*

$$balance(\mathcal{Y}) \geq t, \text{ where } \mathcal{Y} = \cup_{i \leq l} F_i \text{ and } balance(F_i) \geq t$$

Proof. We use the method of induction to derive the proof. Assume we have a set of fairlets $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$, in which, $balance(F_i) \geq t$, $i = 1, \ldots, l$.

We first consider the case for any two fairlets $\{F_1, F_2\} \in \mathcal{F}$. We have $balance(F_1) = \frac{f_1}{m_1} \geq t$ and $balance(F_2) = \frac{f_2}{m_2} \geq t$. We denote by $\mathcal{Y}$ is the union of two fairlets F_1 and F_2, then

$$balance(\mathcal{Y}) = balance(F_1 \cup F_2) = \frac{f_1 + f_2}{m_1 + m_2} \tag{5.3}$$

It holds:

$$\begin{aligned}
\frac{f_1}{m_1} &\geq t \\
\text{or, } \frac{f_1}{m_1 + m_2} &\geq \frac{tm_1}{m_1 + m_2} \\
\text{Similarly, } \frac{f_2}{m_1 + m_2} &\geq \frac{tm_2}{m_1 + m_2} \\
\implies \frac{f_1}{m_1 + m_2} + \frac{f_2}{m_1 + m_2} &\geq \frac{tm_1}{m_1 + m_2} + \frac{tm_2}{m_1 + m_2} \\
\implies \frac{f_1 + f_2}{m_1 + m_2} &\geq \frac{t(m_1 + m_2)}{m_1 + m_2} = t
\end{aligned} \tag{5.4}$$

Therefore, from Eq. 5.3 and Eq. 5.4 we get,

$$balance(\mathcal{Y}) \geq t \tag{5.5}$$

Thus, the statement given in Theorem 1 is true for any cluster formed by the union of any two fairlets. Now we assume that the statement holds true for a cluster formed from j fairlets, i.e, $\mathcal{Y} = \cup_{i \leq j} F_i$, where $1 < j < l$. Then,

$$balance(\mathcal{Y}) = \frac{\sum_{i \leq j} f_i}{\sum_{i \leq j} m_i} \geq t \tag{5.6}$$

Consider another fairlet $F_{j+1} \in \mathcal{F}$ which is not in the formed cluster $\mathcal{Y}$, $balance(F_{j+1}) = \frac{f_{j+1}}{m_{j+1}} \geq t$. Then, by joining F_{j+1} with the cluster $\mathcal{Y}$ we get the new cluster $\mathcal{Y}'$ such that

$$balance(\mathcal{Y}') = \frac{f_{j+1} + \sum_{i \leq j} f_i}{m_{j+1} + \sum_{i \leq j} m_i} \tag{5.7}$$

Following the steps in Eq. 5.4, we can similarly show that

$$\begin{aligned}
\frac{f_{j+1} + \sum_{i \leq j} f_i}{m_{j+1} + \sum_{i \leq j} m_i} &\geq t \\
\implies balance(\mathcal{Y}') &\geq t
\end{aligned} \tag{5.8}$$

Hence, the theorem holds true for clusters formed with $j + 1$ fairlets if it is true for j fairlets. Since j is any arbitrary number of fairlets, the theorem holds true for all cases. □

The theorem 1 shows that for any cluster formed by a union of fairlets, the fairness constraint is always preserved. Henceforth, we don't need further interventions w.r.t. fairness.

The pseudocode of the fair-capacitated hierarchical clustering is demonstrated in Algorithm 4. In each step, the closest pair of clusters is identified (line 4) and a new cluster is created (line 6) only if its capacity does not exceed the capacity threshold q. Otherwise, the next closest pair is investigated. The procedure continues until k clusters remain. The remaining clusters are fair and capacitated according to the corresponding thresholds t and q. To compute the proximity matrix (line 1 and line 8), we use the distance between centroids of the corresponding clusters. The function $capacity(cluster)$ described in line 5 returns the size of a cluster.

Algorithm 4: Hierarchical fair-capacitated algorithm

Input: $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$: a set of fairlets
q: a given maximum capacity of final clusters
$W = \{w_1, w_2, \ldots, w_l\}$: weights of fairlets
k: number of clusters

Output: A fair-capacitated clustering

1 compute the proximity matrix ;
2 $clusters \leftarrow \mathcal{F}$ //each fairlet F_j is considered as cluster ;
3 **repeat**
4     $cluster_1, cluster_2 \leftarrow$ the closest pair of clusters ;
5   **if** $capacity(cluster_1) + capacity(cluster_2) \leq q$ **then**
6         $newcluster \leftarrow$ merge$(cluster_1, cluster_2)$;
7   update $clusters$ with $newcluster$;
8   update the proximity matrix ;
9   **else**
10   continue;
11   **end**
12 **until** k *clusters remain*;
13 **return** $clusters$;

Complexity: In the first phase, the complexity is $\mathcal{O}(n^2)$ for computing the fairlet decomposition [21, 44]. In the second phase with the hierarchical clustering model, the time complexity of the agglomerative algorithm is $\mathcal{O}(n^3)$. Therefore, the complexity is $\mathcal{O}(n^3)$, where n is the number of students.

5.4.3 Fair-capacitated partitioning-based clustering

Partitioning-based clustering algorithms, such as k-medoids, can be viewed as a distance minimization problem, in which, we try to minimize the objective function in Eq. 5.1, i.e., minimize the sum of the distance from every $x_i \in X$ to its medoid s_j.

The vanilla k-medoids method does not satisfy a cardinality constraint since the process of assigning points to clusters relies solely on the distances between them. Now, if we change the goal of this assignment step to find the "best" data points with a defined capacity for each medoid instead of searching for the most suitable medoid for each point, we can control the cardinality of clusters. We formulate the problem of *assigning points to clusters* subject to a capacity threshold q as a ***0-1 knapsack problem*** [137] (see Section 2.5.1, Chapter 2).

At a given k-medoids assignment step, let $S = \{s_1, s_2, ..., s_k\}$ be the cluster centers, i.e., medoids, $\mathcal{C} = \{C_1, C_2, ..., C_k\}$ be the resulting clusters. We change the assignments of points to clusters, using knapsack, in order to meet the capacity constraint q. In particular, we define a flag variable $y_i = 1$ if x_i is assigned to cluster C_j, otherwise $y_i = 0$. Now, if we assign a value v_i to data point x_i, which depends on the distance of x_i to C_j, with v_i being maximum if C_j is the best cluster for x_i, i.e, the distance between x_i and s_j is minimum. We define the value v_i of instance x_i based on an exponential decay distance function:

$$v_i = e^{-\frac{1}{\lambda} * dist(x_i, s_j)} \tag{5.9}$$

where $dist(x_i, s_j)$ is the Euclidean distance between the point x_i and the medoid s_j. The higher λ is the lower the effect of distance on the value of the points. The point which is closer to the medoid will have a higher value.

Then, the objective function for the assignment step is:

$$\text{maximize} \sum_{i=1}^{n} v_i y_i \tag{5.10}$$

Now, given $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$ and $W = \{w_1, w_2, \ldots, w_l\}$ are the set of fairlets and their corresponding weights respectively; q is the maximum capacity of the final clusters. Our target is to cluster the set of fairlets $\mathcal{F}$ into k clusters centered by k medoids. We apply the formulas in Eq. 5.9 and Eq. 5.10 on the set of fairlets $\mathcal{F}$, i.e, each fairlet F_i has the same role as x_i. Then, the problem of assigning the fairlets to each *medoid* in the cluster assignment step becomes finding a set of fairlets with total weights less than or equal to q and the total value is maximized. In other words, we can formulate the cluster assignment step in the partitioning-based clustering as a **0-1 knapsack problem**.

$$\begin{gathered}\text{maximize} \sum_{i=1}^{l} v_i y_i \\ \text{subject to} \quad \sum_{i=1}^{l} w_i y_i \leq q \quad \text{and} \quad y_i \in \{0, 1\}\end{gathered} \tag{5.11}$$

In which, y_i is the flag variable for F_i, $y_i = 1$ if F_i is assigned to a cluster, otherwise $y_i = 0$; v_i is the value of F_i which is computed by the Eq. 5.9; q is the desired maximum capacity.

Algorithm 5: k-medoids fair-capacitated algorithm

Input: $\mathcal{F} = \{F_1, F_2, \ldots, F_l\}$: a set of fairlets
$W = \{w_1, w_2, \ldots, w_l\}$: weights of fairlets
q: a given maximum capacity of final clusters
k: number of clusters

Output: A fair-capacitated clustering

Function `ClusterAssignment`($medoids$):
  $clusters \leftarrow \emptyset$;
  for *each medoid s in medoids* **do**
    candidates $\leftarrow$ all fairlets which are not assigned to any cluster ;
    $len \leftarrow$ length(candidates) ;
    $w \leftarrow$ weights(candidates) ;
    for *each $fairlet_i$ in candidates* **do**
      $values[i] \leftarrow v(fairlet_i)$ //Computed by Eq. 5.9 ;
    end
    $clusters[s]$←knapsack(len, $values$, w, q) ;
  end
  return $clusters$;

Function `main()`:
  $medoids \leftarrow$ select k of the l fairlets arbitrarily ;
  ClusterAssignment(medoids) ;
  $cost_{best} \leftarrow$ current clustering cost;
  $s_{best} \leftarrow null$;
  $o_{best} \leftarrow null$;
  repeat
    for *each medoid s in medoids* **do**
      for *each non-medoid o in $\mathcal{F}$* **do**
        consider the swap of s and o, compute the current clustering cost;
        if *current clustering cost $< cost_{best}$* **then**
          $s_{best} \leftarrow s$;
          $o_{best} \leftarrow o$;
          $cost_{best} \leftarrow$ *current clustering cost*;
        end
      end
    end
    update $medoids$ by the swap of s_{best} and o_{best} ;
    ClusterAssignment(medoids)
  until *no improvements can be achieved by any replacement*;

return $clusters$;

Algorithm 5 depicts the pseudocode of our k-medoids fair-capacitated approach. In the algorithm, for each *medoid* we would search for the adequate points (line 3) by using the function $knapsack(len, values, w, q)$ (line 10) implemented using *dynamic programming* (presented in Algorithm 3, Chapter 2). The 0-1 knapsack returns a list of items with a maximum total value and the total weight not exceeding q. In the main function, we optimize the clustering cost by replacing *medoids* with *non-medoid* instances when the clustering cost is decreased (line 12). This optimization procedure will stop when there is no improvement in the clustering cost (lines 19 to 32).

Complexity: In the first phase, the complexity is $\mathcal{O}(n^2)$ for computing the fairlet decomposition [21, 44]. In the second phase with the k-medoids and 0-1 knapsack problem, the complexity of the k-medoids algorithm is $\mathcal{O}(k(n-k)^2)$ and it costs $\mathcal{O}(n \times q)$ to solve the 0-1 knapsack problem. Therefore, the complexity is $\mathcal{O}(n^2)$, where n is the number of students.

5.5 Experiments

In this section, we describe our experiments and the performance of our proposed algorithms on four real educational datasets.

5.5.1 Experimental setup

Datasets

We evaluate our proposed methods on four public educational datasets[3]. An overview of datasets is presented in Table 5.1. The detailed description of these four datasets is presented in Chapter 3. We select randomly 4,000 instances in the OULAD dataset and 4,000 instances in the course 6.002x of the MOOC dataset to investigate and perform the experiments. The *balance scores* are computed on the cleaned datasets.

Table 5.1. An overview of four educational datasets

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

[3]Student performance dataset consists of two subsets: Portugues and Mathematics subjects which are indicated as "Student-Math" and "Student-Por" in Table 5.1

Baselines

We compare our approaches against well-known clustering methods, including fairness-aware clustering algorithms and traditional clustering.

- ***k*-medoids.** k-medoids clustering [101] is a traditional partitioning technique of clustering that divides the dataset into k clusters and minimizes the clustering cost. k-medoids uses the actual instances as centers.
- **Vanilla fairlet.** This is the approach proposed by Chierichetti et al. [44]. The first phase computes a vanilla fairlet decomposition that ensures fair clusters, but it might not give the optimal cost value. A vanilla k-center clustering algorithm [78] is employed to cluster those fairlets into k clusters in the second step.
- **MCF fairlet.** In this version [44], the fairlet decomposition is transformed into a *minimum cost flow* (MCF) problem, by which an optimized version of fairlet decomposition in terms of cost value is computed. Like the vanilla version, a k-center method is used to cluster fairlets in the second phase.

In our experiments, both resulting fairlets generated by vanilla fairlet and MCF fairlet methods are used for our proposed fair-capacitated clustering algorithms. Therefore, we have two versions of each proposed method, namely *Vanilla fairlet hierarchical fair-capaciated* and *MCF fairlet hierarchical fair-capacitated* (for the hierarchical approach), *Vanilla fairlet k-Medoids fair-capacitated* and *MCF fairlet k-Medoids fair-capacitated* (for the partitioning approach). Section 5.5.2 presents the experimental results of these clustering methods.

Evaluation measures

We report our experimental results on clustering cost, balance score, and capacity. The *clustering cost* is used for evaluating the quality of clustering, which is measured by the formula given in Eq. 5.1. The fairness of clustering is measured by the *balance score* in Eq. 2.14.

Parameter selection

Regarding fairness, a minimum threshold of balance t is set to 0.5 for all datasets in our experiments. It means that the proportion of the minority group (e.g., female) is at least 50% in the resulting clusters. Regarding the λ factor in Eq. 5.9, a value $\lambda = 0.3$ is chosen for our experiments from a range of [0.1, 1.0] via grid-search. We evaluate the clustering cost and balance score on a small dataset, i.e., Student-Math w.r.t. λ.

Theoretically, the **ideal capacity** of clusters is $\left\lceil \frac{n}{k} \right\rceil$ where n is the population of dataset X, k is the number of desired clusters. However, in many cases, the clustering models cannot satisfy this constraint, especially the hierarchical clustering model. Hence, the **maximum capacity** q of clusters is computed by Eq. 5.12. In which, ε is a fine-tuning parameter to ensure that each cluster has an (integer) number of members; parameter ε is chosen for each fair-capacitated clustering approach.

$$q = \left\lceil \frac{n * \varepsilon}{k} \right\rceil \tag{5.12}$$

In our experiments, to find the appropriate value of ε, we set a range of [1.0, 1.3] to ensure all the generated clusters have members. We evaluate the cardinality of resulting clusters on the Student-Math dataset. Based on this, ε is set to 1.01 and 1.2, for k-medoids fair-capacitated and hierarchical fair-capacitated methods, respectively.

5.5.2 Experimental results

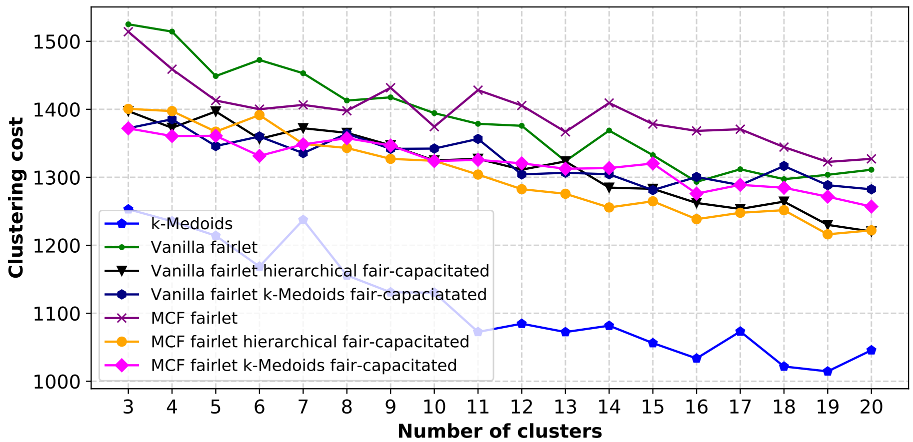
Student performance dataset

Student-Math dataset

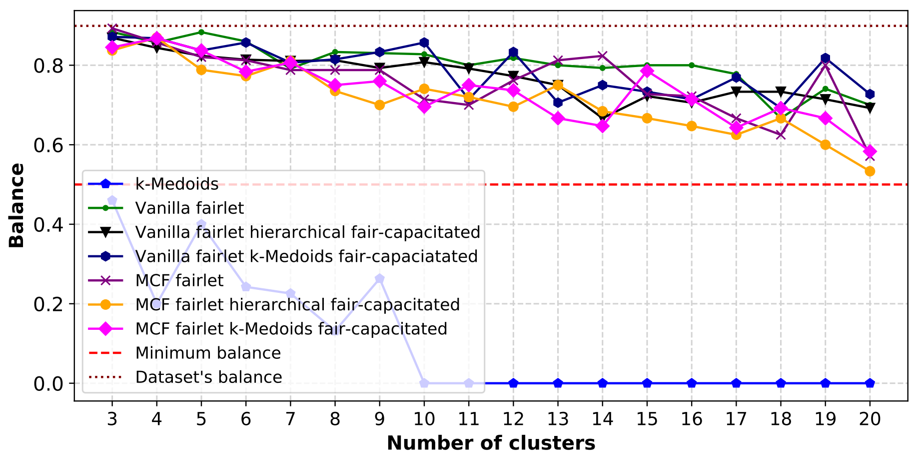
In Figure 5.1-a, the clustering cost of all methods is worse compared to that of the vanilla k-medoids clustering. This is expected as these methods have to satisfy constraints on fairness or/and cardinality. However, both of our approaches outperform the vanilla fairlet and MCF fairlet methods. In which, *MCF fairlet hierarchical fair-capacitated* shows the best performance due to the optimization in the merging step. Regarding fairness, as shown in Figure 5.1-b, the minimum threshold of balance t is visualized as a dashed line while the actual balance from the dataset is plotted as a dotted line. All of our methods are comparative to the competitors in most cases. Interestingly, our *vanilla fairlet k-medoids fair-capacitated* method outperforms the competitive methods when k is less than 10. In terms of cardinality, as presented in Figure 5.1-c, the maximum capacity thresholds q are indicated by the figure's dashed and dotted lines. Our capacitated variants are superior (lower dispersion as shown by the interquartile ranges). We have to thicken the boxplots of our proposed methods since in many cases the dispersion in the size of the resulting clusters is too small. MCF fairlet shows the worst performance in terms of cardinality, followed by Vanilla fairlet and vanilla k-medoids.

Student-Por dataset

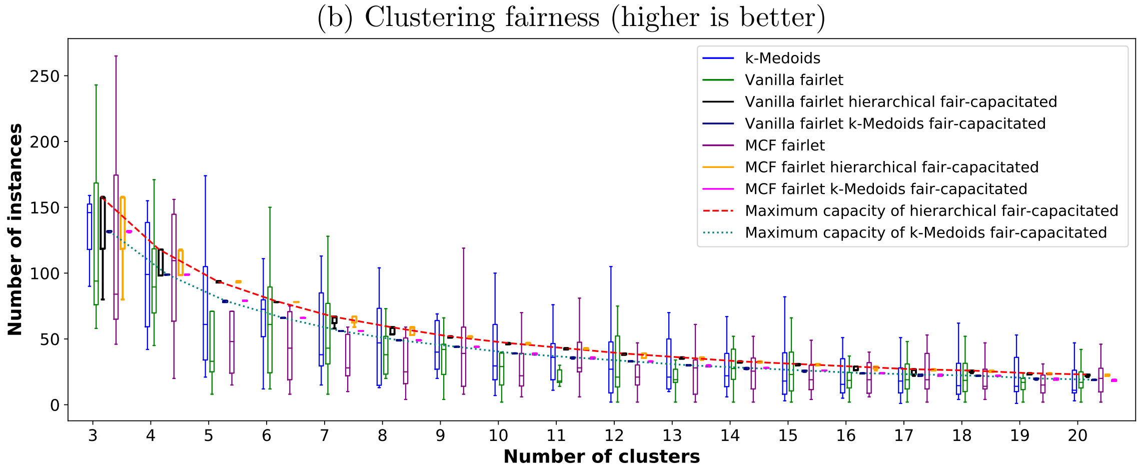
When k is less than 4, as shown in Figure 5.2-a, the clustering quality of our models can be close to that of the vanilla k-medoids method. However, the clustering cost fluctuates thereafter due to the effort to maintain the fairness and cardinality of methods. Our *vanilla fairlet hierarchical fair-capacitated* outperforms other competitors in most cases. Vanilla fairlet and MCF fairlet show the worst clustering cost as an effect of the k-center method. Figure 5.2-b depicts the clustering fairness. As we



(a) Clustering quality (lower is better)



(b) Clustering fairness (higher is better)



(c) Clustering cardinality

Figure 5.1. Student-Math: Performance of different methods w.r.t. clustering quality (a), fairness (b) and cardinality (c)

can observe, in terms of fairness, *vanilla fairlet hierarchical fair-capacitated* has the best performance when k is less than 10. Contrary to that, by selecting each point for each cluster in the cluster assignment step, the *k-medoids fair-capacitated* method can maintain well the fairness in many cases. Regarding cardinality, as illustrated in Figure 5.2-c, our approaches outperform the competitors when they can keep the number of instances for each cluster under the specified thresholds.

PISA dataset

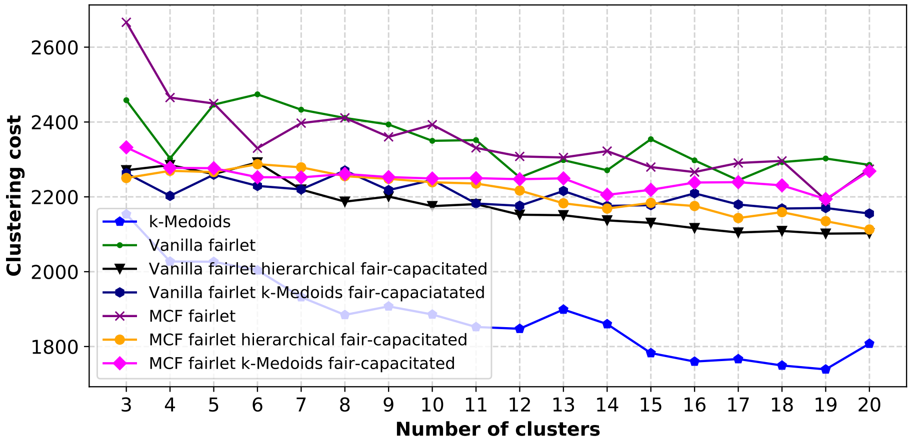
Although the clustering cost increases in most methods, as presented in Figure 5.3-a, our approaches outperform the competitors, i.e., vanilla fairlet and MCF fairlet. The hierarchical approach shows the best performance compared to other methods which are concerned with equity and capacity. Interestingly, our proposed methods outperform the competitors when they can preserve very well the balance score for all number of clusters in terms of fairness (Figure 5.3-b). This is explained by fairness in the fairlets that are used as the input for our clustering method. It is easy to observe in Figure 5.3-c that our proposed methods strictly follow the maximum capacities of clusters regarding cardinality. MCF fairlet is the worst model, followed by k-medoids and vanilla fairlet, and MCF fairlet.

OULAD dataset

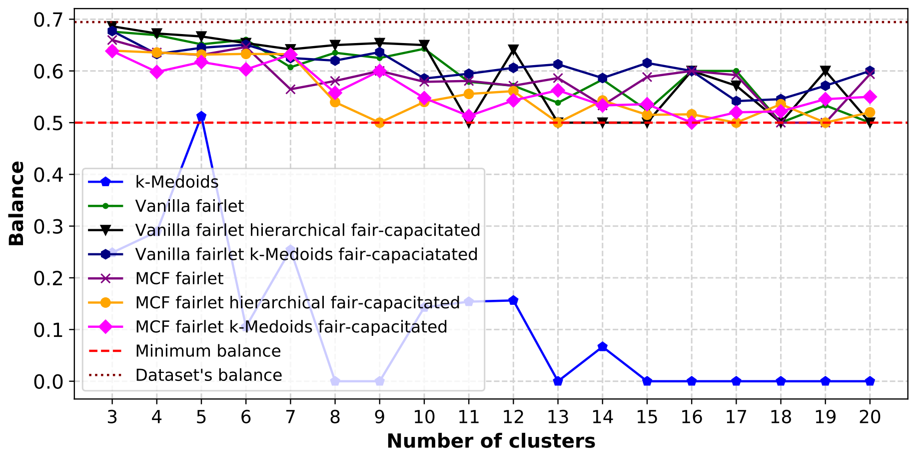
Our *MCF fairlet k-medoids fair-capacitated* approach outperforms other methods in terms of clustering cost, although there is an increase compared to the vanilla k-medoids algorithm, as we can see in Figure 5.4-a. Concerning fairness, in Figure 5.4-b, k-medoids is the weakest method while others can achieve the highest balance. The balance of *Gender* feature in the dataset is the main reason for this result. All fairlets are fully fair; this is a prerequisite for our methods of being able to maintain the perfect balance. Regarding cardinality, our approaches demonstrate their strength in ensuring the capacity of clusters (Figure 5.4-c). The difference in the size of the clusters generated by our methods is tiny. This is in stark contrast to the trend of competitors.

MOOC dataset

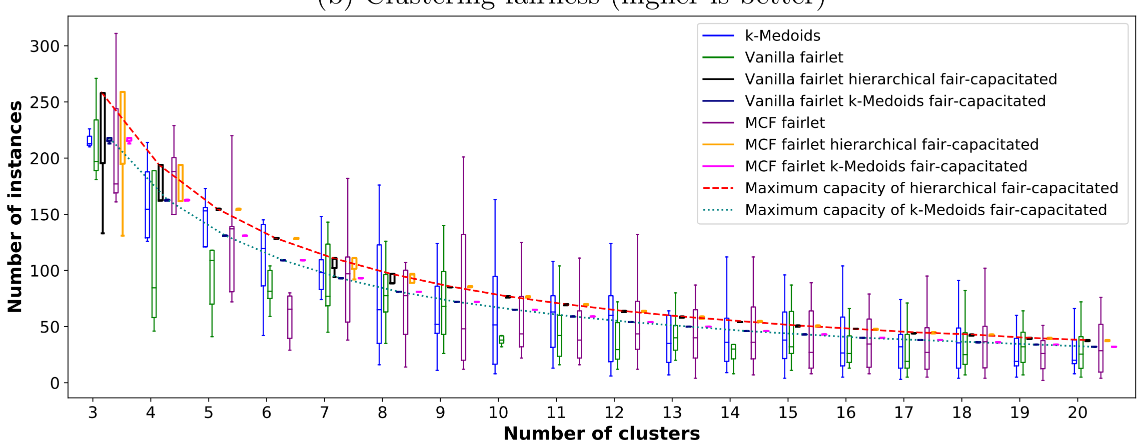
The results of clustering quality are described in Figure 5.5-a. Although an increase in the clustering cost is the main trend, our methods outperform the vanilla fairlet and MCF fairlets methods. Regarding clustering fairness, as depicted in Figure 5.5-b, our approaches can maintain the perfect balance for all experiments. This is the result of the actual balance in the dataset and the fairlets. The emphasis is our methods can divide all the experimented instances into capacitated clusters, as presented in Figure 5.5-c, which proves their superiority in presenting the results over the competitors regarding the cardinality of clusters.



(a) Clustering quality (lower is better)

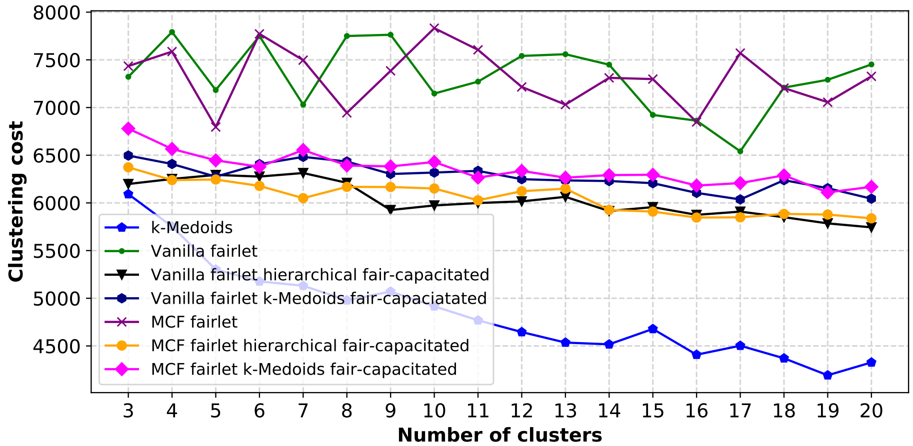


(b) Clustering fairness (higher is better)

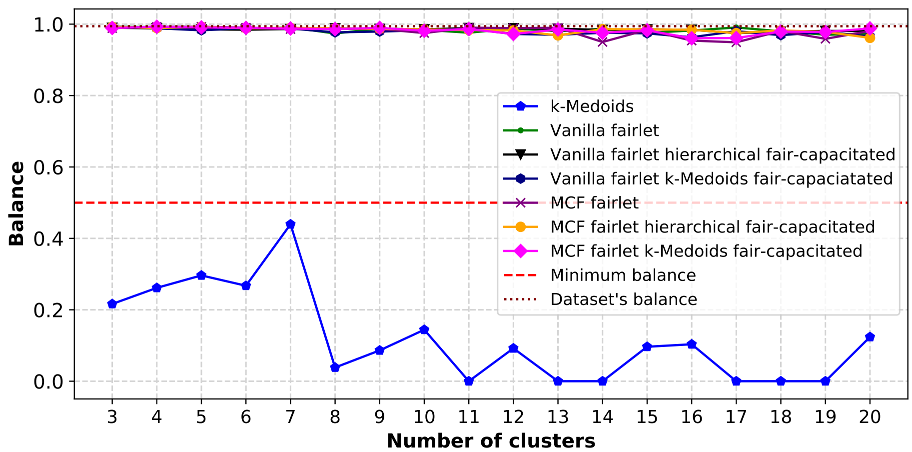


(c) Clustering cardinality

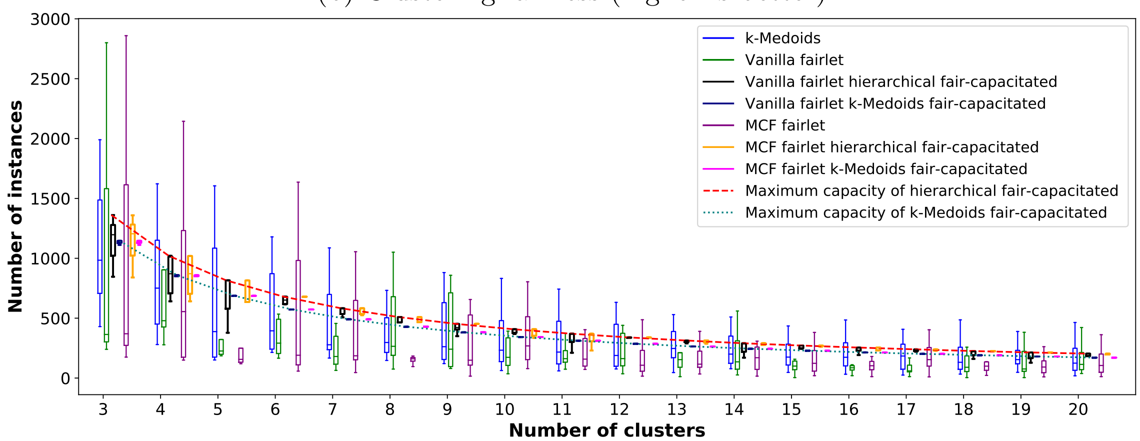
Figure 5.2. Student-Por: Performance of different methods w.r.t. clustering quality (a), fairness (b) and cardinality (c)



(a) Clustering quality (lower is better)

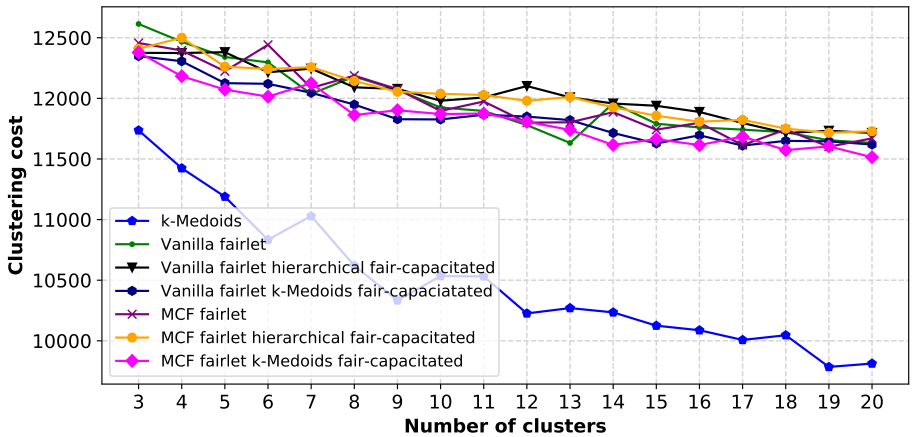


(b) Clustering fairness (higher is better)

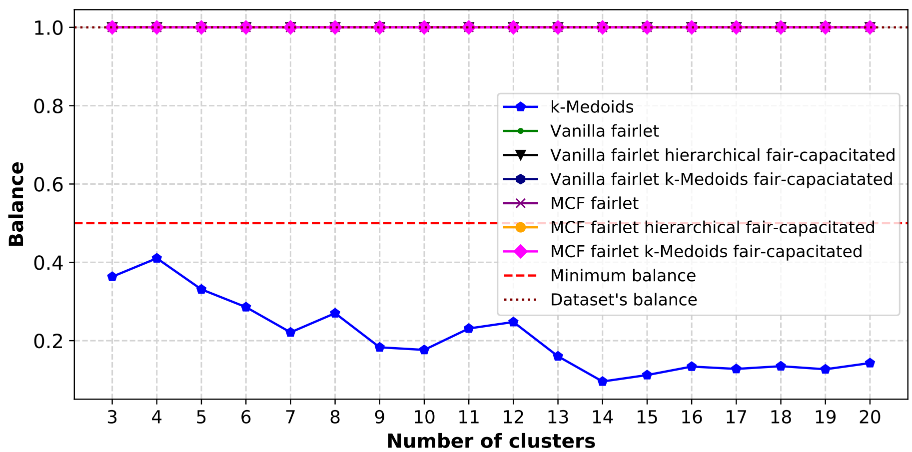


(c) Clustering cardinality

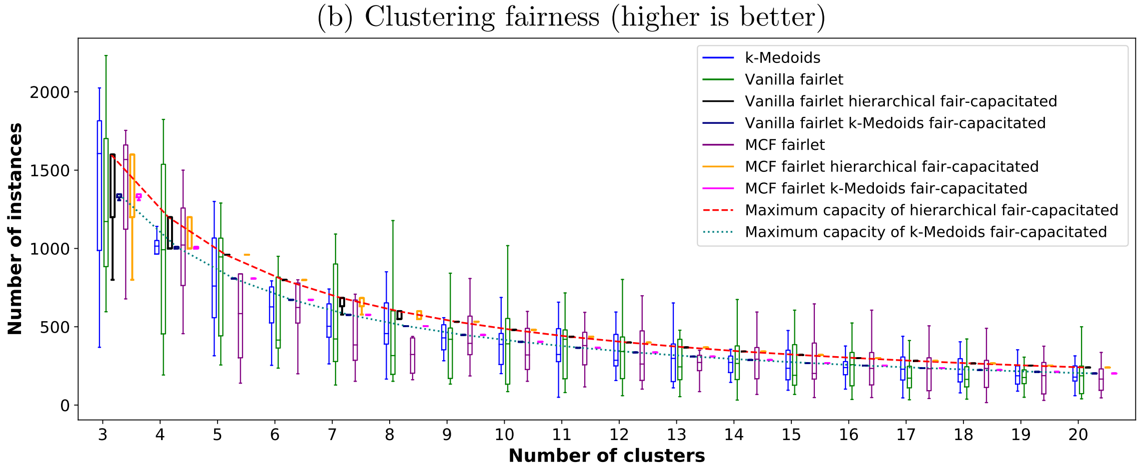
Figure 5.3. PISA: Performance of different methods w.r.t. clustering quality (a), fairness (b) and cardinality (c)



(a) Clustering quality (lower is better)

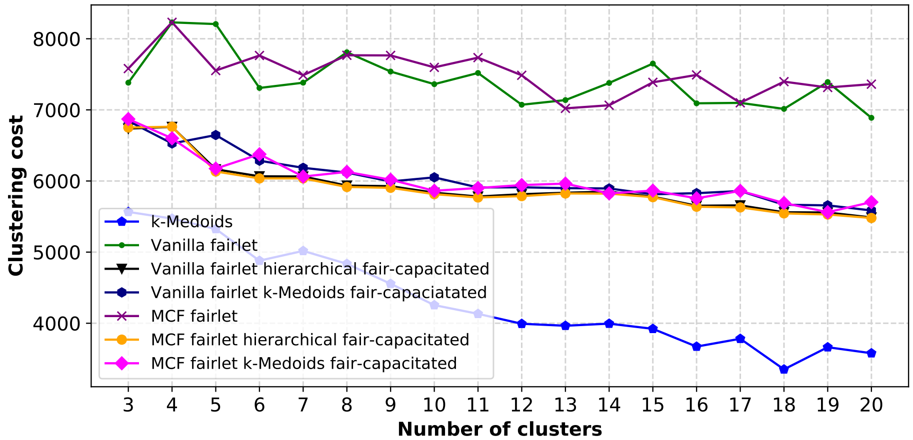


(b) Clustering fairness (higher is better)

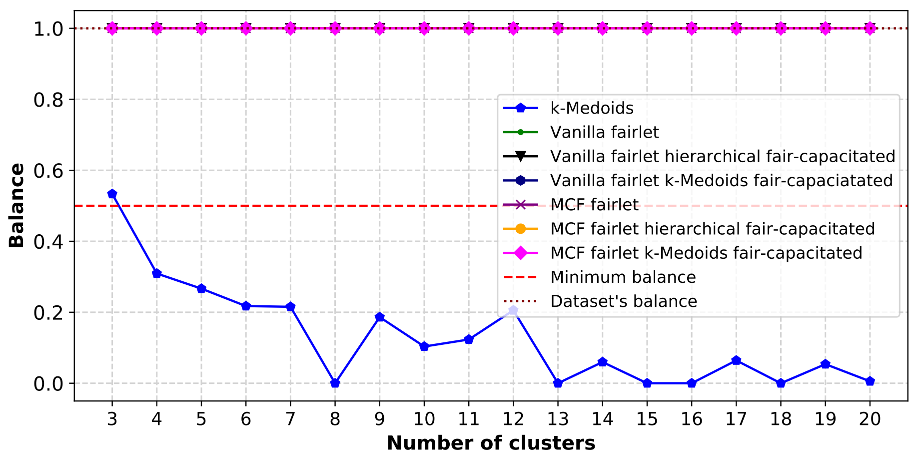


(c) Clustering cardinality

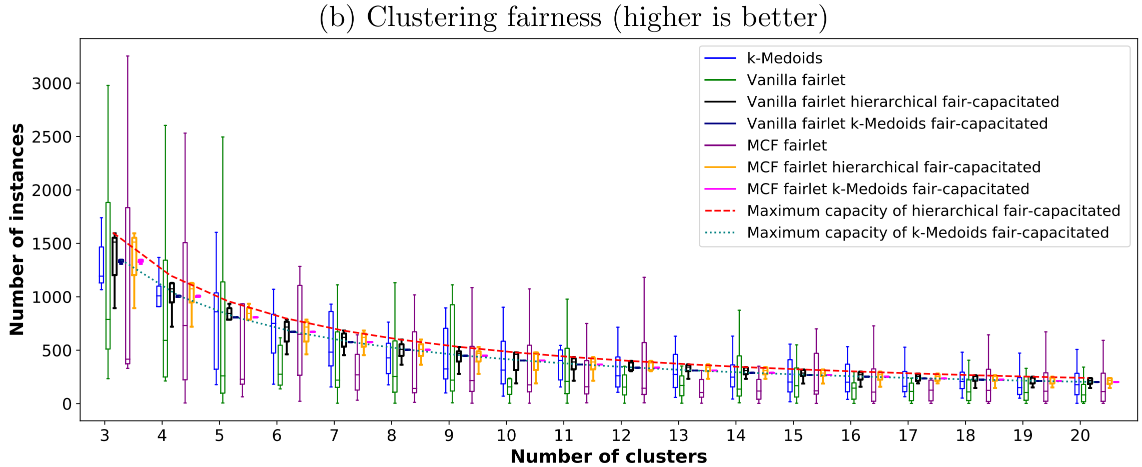
Figure 5.4. OULAD: Performance of different methods w.r.t. clustering quality (a), fairness (b) and cardinality (c)



(a) Clustering quality (lower is better)



(b) Clustering fairness (higher is better)



(c) Clustering cardinality

Figure 5.5. MOOC: Performance of different methods w.r.t. clustering quality (a), fairness (b) and cardinality (c)

Summary of the results

In general, fairness is well maintained in all of our experiments. When the data is fair, in the case of OULAD and MOOC datasets, our methods achieve perfect fairness. In terms of cardinality, our methods are able to maintain the cardinality of resulting clusters within the maximum capacity threshold, which is significantly superior to competitive methods. The fair-capacitated partitioning-based method is better than the hierarchical approach since it can determine the capacity threshold closest to the *ideal cardinality* mentioned in Section 5.5.1. Regarding the clustering cost, the hierarchical approach has an advantage over other methods by outperforming its competitors in most experiments.

5.6 Chapter summary

In this chapter, we introduced the fair-capacitated clustering problem that extends traditional clustering, solely focusing on similarity, by also aiming at a balanced cardinality among the clusters and a fair representation of instances in each cluster according to a protected attribute, like gender or race. Our solutions work on the fairlets derived from the original instances: the hierarchical-based approach takes into account the cardinality requirement during the merging step, whereas the partitioning-based approach takes into account the cardinality of the final clusters during the assignment step which is formulated as a 0-1 knapsack problem. Our experiments show that our methods are effective in terms of fairness and cardinality while maintaining clustering quality. Apart from the educational field, the fair-capacitated clustering problem can contribute to other applications such as the clustering of customers in marketing studies, vehicle routing, and communication network design. An immediate future direction is to improve the clustering quality by optimizing the cluster assignment phase of the partitioning-based approaches. Moreover, we plan to extend our work for multiple protected attributes.

6

Multi-fair capacitated students-topics grouping problem

In Chapter 5, we cluster students w.r.t. group's cardinality and fairness in terms of the protected attribute without considering students' preferences. In this chapter, we focus on the problem of grouping students based on students' preferences, which should reflect students' aspirations as much as possible. In addition, the resulting groups should also be balanced in terms of protected attributes like gender. Moreover, to allow a fair workload across the groups, the cardinalities of the different groups should be balanced. We introduce a *multi-fair capacitated* (MFC) grouping problem that fairly partitions students into non-overlapping groups while ensuring balanced group cardinalities (with a lower and an upper bound) and maximizing the diversity of members regarding the protected attribute. We propose three approaches to solve the MCF grouping problem: a greedy heuristic approach, a knapsack-based approach using vanilla maximal 0-1 knapsack formulation, and an MFC knapsack approach based on group fairness knapsack formulation. Experimental results on a real dataset and a semi-synthetic dataset show that our proposed methods can satisfy students' preferences and deliver balanced and diverse groups regarding cardinality and the protected attribute, respectively.

6.1 Introduction

Teamwork plays a vital role in educational activities, as students can work together to achieve shared learning goals. By working in groups, students have better communication skills and become more social and creative. Moreover, they can learn about leadership, higher-order thinking, conflict management [68, 80] etc. A common approach to group students into teams is as follows: The instructor provides a list of topics, projects, or tasks, etc.[1], according to which the different non-overlapping

[1] We use "topic" to refer to all the terms: topic, project, task, etc.

groups of students should be formed. The grouping procedure can be performed randomly or based on students' preferences [143] typically expressed as a ranking over the provided topics. Or, the teacher just says: "Find yourself into groups"; in this case, a grouping is not random and does not consider students' preferences w.r.t. topics but it is triggered by social connections. The common case in educational settings is the grouping w.r.t. students' preferences. The important case and often in educational settings is the grouping with regard to students' preferences. In the classroom, this costs time, and a suitable algorithm could help. Therefore, in this work, we consider the case of grouping w.r.t. students' preferences.

The grouping process should consider various requirements. First, students' preferences should be taken into account (i.e., *student satisfaction*). A grouping is considered satisfactory if it can satisfy the students' preferences as much as possible. Second, the groups should be balanced in terms of their cardinalities, so all students share a similar workload (i.e., *group cardinality*) because when groups have unequal sizes, and the minority group is smaller than a critical size, the minority cohesion widens inequality [152]. Third, the instructor might be interested in fair-represented groups w.r.t. some protected attributes like gender or race [108] (i.e., *group fairness*), as studies suggest that mixed-gender grouping may have a positive effect on groups' performance [65].

These requirements have been discussed in the related work but are typically treated independently. For example, fairness w.r.t. workload distribution and students' preferences has been discussed in group assignments [68], assignment of group members to tasks [143] or students to projects [162]. Student satisfaction is typically assessed as the number of topics staffed [130] or the sum of the utilities of the topics assigned to students based on the ranking of preferences chosen by students [133]. The group cardinality can be satisfied by the heuristic method [145], or the hierarchical clustering approach [118]. However, providing a grouping solution that simultaneously satisfies all these constraints is difficult. And, "in general, it is not possible to assign all students to their most preferred project" [162].

To this end, we introduce *multi-fair capacitated (MFC) grouping* problem that aims to ensure fairness of the resulting groups in multiple aspects. In particular, we target fairness in terms of i) maximizing students' satisfaction, ii) ensuring fairness in group representation w.r.t. the protected attribute, and iii) balancing group cardinalities. For the satisfaction aspect, we employ the *Nash social welfare* notation [148]; for the fairness w.r.t. protected attribute we use the balance score notion [44]. To solve the MFC problem, we propose three approaches: i) a greedy heuristic algorithm; ii) a knapsack-based approach that reformulates the assignment step as a maximal 0-1 knapsack problem; iii) an MFC knapsack model based on the group fairness knapsack formulation [153].

The rest of this chapter is structured as follows: we overview the related work in Section 6.2. The MFC grouping problem is introduced in Section 6.3. Section 6.4 presents the solutions to the MFC problem. The experimental evaluation on several educational datasets is described in Section 6.5. Finally, section 6.6 provides the summary of the chapter.

6.2 Related work

In the education domain, Agrawal et al. [6] proposed the problem of grouping students in a large class w.r.t. the overall gain of students. Miles et al. [143] investigated the problem of assignment of group members to tasks w.r.t. the workload distribution. They examined the viability of four methods to assign students into groups: random, ability, personal influence, and personal influence with justification. Concerning a diversity of features such as skills, genders, and academic backgrounds, Krass et al. [108] studied the problem of assigning students to multiple non-overlapping groups. The problem was solved by an integer programming model to minimize the number of overlaps. In a similar study, Cutshall et al. [49] assigned students into groups based on their academic background and gender. However, students' preferences were not considered.

To consider both efficiency and fairness, Magnanti et al. [133] solved a CPLEX integer programming formulation with two objectives: maximizing the total utility computed by the rank of student's preferences (efficiency) and minimizing the number of students assigned to the projects which they do not prefer (fairness). They used the optimization solver CPLEX with the SolverStudio Excel interface[2]. Besides, Rezaeinia et al. [162] introduced a lexicographic approach to prioritize the goals. The efficiency objective is computed based on the utility, similar to the work of Magnanti et al. [133]; however, the authors adapted Jain's index [95] to measure the fairness of the assignment.

A related problem is the problem of assigning reviewers to papers [83, 98, 129, 183]. Each reviewer can be assigned to several papers, and each paper can be assigned to several reviewers. However, in the students grouping problem, we attempt to generate non-overlapping groups [108], where each student can be assigned to only one group [162].

The knapsack problem formulation has been used for finding good clustering assignments [118]. However, the minimum capacity of a group (cluster) is not ensured. Recently, Stahl et al. [182] introduced a fair knapsack model to balance the price given by the data provider and the suggested price of the customer. The data vendors propose the data for an *ask price*, and customers can negotiate a *bid price*. The data quality is adjusted to satisfy the price bargained by the customer and ensure the final selling price is fair. Next, Fluschnik et al. [67] proposed three concepts of fair knapsack (individually best, diverse, and fair knapsack) to solve the problem of choosing a subset of items with the total cost is not greater than a given *budget* while taking into account the preferences of the voters.

[2]https://solverstudio.org/

The Nash social welfare (or Nash equilibrium) [148] was used as the solution concept for fairness [67], i.e., fairness is ensured by the objective function. The group fairness definition for the knapsack problem was investigated recently by Patel et al. [153]. In their study, each item is characterized by a *category*, their goal is to select a subset of items such that the total value of the selected items is maximized, and the total weight does not surpass a given weight while each category is *fairly* represented. The notion of *group fairness* is defined based on three criteria (the number of items, the total value of items, and the total weight of items in each category).

In this work, we introduce the MFC grouping problem that ensures fairness in multiple aspects. In particular, we guarantee fairness in terms of maximizing students' satisfaction (by objective function) in parallel with fairness w.r.t. protected attributes and balancing group cardinalities (lower bound and upper bound on the group cardinality). To the best of our knowledge, the proposed problem has not been studied before and, as already discussed, comprises a useful tool to ensure fairness in educational activities.

6.3 Problem definition

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of n students, $T = \{t_1, t_2, \cdots, t_m\}$ be a set of m topics. For an integer n we use [n] to denote the set $\{1, 2, \cdots, n\}$. Each student can choose h topics as their preference ($h \ll m$). The students' preferences are stored in matrix $wishes$. Row $wishes_i$ contains the list of h topics preferred by student i. We use a matrix V to record the student's level of interest in the topics. The preference of topic t_j chosen by student x_i is represented by a number v_{ij}. The more preferred topic will have a higher value of v_{ij}. Matrix V is computed as: $V_{i,wishes_{io}} = h/o$ with $o \in [h]$, where o indicates the order of preferences. Likewise, each topic t_j can be chosen by several students. A priority matrix W consists of values computed based on the registration time, where w_{ij} represents the priority of student x_i on topic t_j. Students who register earlier will have a higher value of w_{ij}. If the topic t_j is not preferred by student x_i then $v_{ij} = 0$ and $w_{ij} = 0$.

Let $\varphi : V \times W \to \mathbb{R}$ be the aggregate function of matrices V and W. For each student x_i, we define a *welfare* value w.r.t. topic t_j: $welfare_{ij} = \varphi(v_{ij}, w_{ij})$. In detail:

$$\varphi(v_{ij}, w_{ij}) = \alpha v_{ij} + \beta w_{ij} \tag{6.1}$$

where α and β are the parameters indicating the weight of each component.

Figure 6.1 illustrates a dataset with 5 students and 4 topics. The matrix $welfare$ is computed with $\alpha = 1$ and $\beta = 1$ (preferences and registration time are equally considered).

The goal of a grouping problem is to distribute n students into k disjoint groups $\mathcal{G} = \{G_1, G_2, \cdots, G_k\}$, $(k \leq m)$, that maximizes the students' preferences w.r.t. the

$h = 3$ 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*$_{n\times h}$

$m = 4$ 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*$_{n\times m}$

$m = 4$ 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*$_{n\times m}$

$m = 4$ 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*$_{n\times m}$

Figure 6.1. A dataset with matrices $wishes$, V, W and $welfare$

registration time, formulated by the objective function:

$$\mathcal{L}(X, \mathcal{G}) = \prod_{r=1}^{k} \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r} \tag{6.2}$$

In other words, the goal is to maximize the product of the total *welfare* obtained from each group G_r. In Eq. 6.3, a set of indexes $J = \{j_1, j_2, \cdots, j_k\}$ of k selected topics is defined as $J = \{j | x_i \in G_r, welfare_{ij} > 0\}$, $\forall r \in [k]$. Variable y_{ij_r} is the flag of x_i; $y_{ij_r} = 1$ if x_i is assigned to the group of topic t_{j_r}, otherwise $y_{ij_r} = 0$.

Similar to formula in the study of Fluschnik et. al. [148], Eq. 6.2 is the representation of the Nash social welfare function[3]. Therefore, we can call a grouping satisfactory if it maximizes the product in the objective function $\mathcal{L}(X, \mathcal{G})$.

Furthermore, we add one to the sum $\sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r}$ to avoid the phenomenon that the sum of *welfare* in a certain group might be zero. The objective function $\mathcal{L}(X, \mathcal{G})$ is rewritten as follows:

$$\mathcal{L}(X, \mathcal{G}) = \prod_{r=1}^{k} (1 + \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r}) \tag{6.3}$$

[3]The Nash social welfare was defined as $\prod_{v_i \in V}(1 + \sum_{a \in S} u_i(a))$ [67] (the typical formula is $\prod_{v_i \in V} \sum_{a \in S} u_i(a)$, where v_i is a voter in a set of voters V, a is an item of the knapsack S, and $u_i(a)$ represents the extent to which v_i enjoys a. The knapsack S is fair if that product is maximized.

Fairness of grouping w.r.t. a protected attribute: Assume that each student is characterized by a binary protected attribute $\mathcal{P} = \{p, \overline{p}\}$, where p is the protected group (e.g., $\mathcal{P}$ = "gender" = {female, male}; $p = female$) and $\overline{p}$ is the non-protected group (e.g., $\overline{p} = male$). Let $\psi : X \rightarrow \mathcal{P}$ be the demographic category to which the student belongs. Fairness of a grouping $\mathcal{G}$ w.r.t. protected attribute [44] is computed as:

$$balance(\mathcal{G}) = \min_{\forall G_r \in \mathcal{G}} balance(G_r) \tag{6.4}$$

where fairness of a group G_r is the minimum ratio between two categories:

$$balance(G_r)_{\forall G_r \in \mathcal{G}} = \min \left(\frac{|\{x \in G_r \mid \psi(x) = p\}|}{|\{x \in G_r \mid \psi(x) = \overline{p}\}|}, \frac{|\{x \in G_r \mid \psi(x) = \overline{p}\}|}{|\{x \in G_r \mid \psi(x) = p\}|} \right) \tag{6.5}$$

Capacitated grouping: Inspired by the capacitated clustering problem [145], we call a grouping *capacitated* if the cardinality of each group G_r, i.e., $|G_r|$, is between a given lower bound $C^l \geq 0$ and an upper bound $C^u \geq C^l$.

Definition 6.1. ***MFC grouping problem***
We describe the MFC grouping problem as finding a grouping $\mathcal{G} = \{G_1, G_2, \cdots, G_k\}$ that distributes a set of students X into k groups corresponding to k topics, and satisfies the following requirements:

i) The assignment is satisfactory, i.e., maximizing students' satisfaction (Eq. 6.3);

ii) The balance of each group G_r is maximized, i.e., the fairness constraint w.r.t. the protected attribute (Eq. 6.4);

iii) The cardinality of each group $G_r \in \mathcal{G}$ is bounded within $[C^l, C^u]$.

6.4 Solving the MFC grouping problem

To solve the MFC grouping problem, we first propose a greedy heuristic algorithm (Section 6.4.1); then we formulate the assignment phase as a vanilla maximal 0-1 knapsack (Section 6.4.2) or a group fairness knapsack problem (Section 6.4.3). In general, our MFC grouping problem is an NP-hard problem since it is a case of assignment problems [33, 38]. Hence, we employ heuristic methods to efficiently find solutions within a polynomial time.

6.4.1 A greedy heuristic approach

The main idea of our greedy heuristic approach is to assign a student to the topic that is the highest favorite topic in the student's preferences. This approach is divided into two main phases, as presented in the Algorithm 6.

In the first step, we maximize the students' preferences by assigning them to their most preferred topic. If a topic is preferred by many students we select the student who has the highest *welfare* value according to Eq. 6.1 (lines 4, 5). In the second step, we adjust the assignment to satisfy the fairness w.r.t. the protected attribute and cardinality requirements by *GroupAdjustment* function (Algorithm 7). The number of students of each group (e.g., male, female) w.r.t. protected attribute $(p_0^l, p_0^u, p_1^l, p_1^u)$ are computed based on the resulting groups' cardinalities (C^l, C^u) and the balance score θ (line 2). If there are ungrouped students, we try to assign them to the existing groups (lines 3 - 6). If all groups are full, we choose the topic that is most preferred by the remaining ungrouped students and assign them to such a topic (lines 10 - 14). We disband groups containing too few students and assign those ungrouped students to other groups until all groups have the desired capacity (lines 19 - 25).

Complexity: The first step consumes $\mathcal{O}(n \times h)$ and the second step costs $\mathcal{O}(C^l \times n \times m)$ as the algorithm has to deal with every group having cardinality less than C^l. As $C^l \ll n$ and $C^u \ll n$, the complexity of the greedy heuristic model is $\mathcal{O}(n \times m)$, where n is the number of students, m is the number of topics and h is the number of wishes.

Algorithm 6: Greedy heuristic algorithm

Input: X: a set of students; n: #students; h: #preferences; m: #topics; C^l, C^u: capacities ; matrices $wishes_{n \times h}$, $V_{n \times m}$, $W_{n \times m}$; θ: balance score

Output: A grouping with k groups

1 *groups* $\leftarrow \emptyset$; $welfare \leftarrow \varphi(V, W)$;//Step 1: Assign students to groups;
2 **for** $i \leftarrow 1$ **to** n **do**
3   **for** $j \leftarrow 1$ **to** h **do**
4   **if** *(topic* $wishes_{ij}$ *is the most preferred topic of student* i*) and (*$welfare_{i,wish_{ij}}$ *is the highest value among students choosing topic* $wishes_{ij}$*) and (*$len(groups[wishes_{ij}] < C^l)$*)* **then**
5             $groups[wishes_{ij}]$.append(i);
6   **end**
7   **end**
8   GroupAdjustment(*groups*) //Step 2: Adjustment;
9 **end**
10 **return** *groups*;

6.4.2 A knapsack-based approach

In the greedy heuristic approach, we tend to assign students to their highest favorite topics. This assignment can be detrimental to students' satisfaction because there may be some students who have no more topics to be assigned, even though they also have a high degree of interest in that topic. Therefore, assigning students to their second or third favorite topic could improve student satisfaction overall, for example.

Algorithm 7: Group adjustment algorithm

Input: *groups*: a set of groups; n: #students; h: #preferences; m: #topics; C^l, C^u: capacities; θ: balance score

Output: An adjusted grouping

Function `GroupAdjustment`(*groups*):
  $p_0^l \leftarrow \left\lceil \frac{C^l}{\frac{1+\theta}{\theta}} \right\rceil ; p_0^u \leftarrow \left\lceil \frac{C^u}{\frac{1+\theta}{\theta}} \right\rceil ; p_1^l \leftarrow C^l - p_0^l ; p_1^u \leftarrow C^u - p_0^u + 1$;
  for $i \leftarrow 1$ **to** n **do**
    for $q \leftarrow 1$ **to** m **do**
      if $(i \notin groups[q])$ *and* $len(groups[q] < C^l)$ *and* $((n\_students\_0 < p_0^l)$ *or* $(n\_students\_1 < p_1^l))$ **then**
        $groups[q].\text{append}(i)$;
      end
    end
  end
  while $len(unassigned\_students) > 0$ **do**
    $id \leftarrow$ the most prevalent topic preferred by remaining students;
    for $i \in unassigned\_students$ **do**
      if $len(groups[id]) < C^u$ *and* $((n\_students\_0 < p_0^u)$ *or* $(n\_students\_1 < p_1^u))$ **then**
        $groups[id].\text{append}(i)$;
      end
    end
  end
  $n\_items \leftarrow 1$;
  while *(cardinalities of all groups* $\notin [C^l, C^u]$*)* **do**
    if $n\_items < C^l$ **then**
      Resolve the groups with cardinality n_items;
      if $(n\_students\_0 < p_0^u)$ *or* $(n\_students\_1 < p_1^u)$ **then**
        Assign ungrouped students to the remaining groups having cardinality $< C^u$;
      end
      $n\_items++$;
    end
  end
return *groups*;

Therefore, we propose a new approach whereby we will search for the most suitable students for each topic. We will formulate the task of selecting the "best" students for a group of the MFC grouping problem as a *maximal 0-1 knapsack* problem [137].

Let *capacity* be a cardinality array with $capacity_i = 1, \forall i \in [n]$; $welfare_{ij} = \varphi(v_{ij}, w_{ij})$ and the indexes of k topics $J = \{j_1, j_2, \cdots, j_k\}$ will be chosen for the resulting groups. For each topic $t_{j_r} \in T$, $\forall r \in [k]$, i.e., r is the index of the selected knapsack, the goal is to select a subset of students (G_r), such that:

$$\begin{aligned} &\text{maximize} \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r} \\ &\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^l \end{cases} \end{aligned} \tag{6.6}$$

where

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

In other words, for each selected topic, we find a set of students that maximizes the total $welfare$, while the total *capacity*, is within the given bounds. The pseudo-code is described in Algorithm 8 with two steps. In the first step, we find the most suitable candidates among the unassigned students by the solution of a maximal 0-1 knapsack problem [137] for each topic. We use dynamic programming to solve the maximal 0-1 knapsack problem (Eq. 6.6), where the dynamic programming approach is described in Algorithm 3 in Chapter 2. The second step is demonstrated in Algorithm 7 which performs a fine-tuning of the assignment.

Complexity: In the first step, the complexity is $\mathcal{O}(m \times n \times C^u)$ since it costs $\mathcal{O}(n \times C^u)$ for each topic to solve the knapsack problem. The running time of the second step is $\mathcal{O}(C^l \times n \times m)$. Therefore, the complexity is $\mathcal{O}(n \times m)$.

6.4.3 An MFC knapsack approach

In the knapsack-based approach, we select the most suitable students for each topic by solving a *maximal knapsack* problem. However, the fairness constraint w.r.t. the protected attribute is not directly considered in the knapsack formulation. Inspired by the knapsack problem with *group fairness* constraints of Patel et al. [153], we propose an *MFC knapsack* algorithm to find the group of suitable students, which satisfies the MFC problem' requirements.

The goal of the MFC knapsack is to select a subset of student (G_r), such that:

Algorithm 8: Knapsack-based algorithm

Input: X: a set of students; n: #students; h: #preferences; m: #topics; C^l, C^u: capacities; matrices $wishes_{n \times h}$; $V_{n \times m}$; $W_{n \times m}$.

Output: A grouping with k groups

$groups \leftarrow \emptyset$ //Step 1: Assign students to groups ;
$welfare \leftarrow \varphi(V, W)$;
for $id \leftarrow 1$ **to** m **do**
  $capacity \leftarrow$ get_capacity($unassigned\_students$);
  $values \leftarrow$ get_welfare($unassigned\_students, welfare$);
  $n\_items \leftarrow$ len($unassigned\_students$);
  if $n\_items > 0$ **then**
    if $n \bmod C^l = 0$ **then**
      $selected\_students \leftarrow$ knapsack($values, capacity, n, C^l$);
    end
    else
      $selected\_students \leftarrow$ knapsack($values, capacity, n, C^u$);
    end
    $groups[id] \leftarrow selected\_students$;
  end
  GroupAdjustment($groups$) //Step 2: Adjustment;
end
return $groups$;

$$
\begin{aligned}
&\text{maximize} \sum_{i=1}^{n} welfare_{ij_r} \times y_{ij_r} \\
&\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i \times y_{ij_r} \leq C^l \\ balance(G_r) \text{ is maximized} \end{cases}
\end{aligned}
\tag{6.7}
$$

where

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

We use dynamic programming to solve the MFC knapsack problem (presented in Algorithm 9). The input parameters include a set of unassigned students $\mathcal{S} \subseteq X$. A dynamic programming table $\mathcal{A}(g, s, w)$ is used to record the total welfare of the first s students in the set $\mathcal{S}$ with capacity w on group g, $\forall g \in \{p, \overline{p}\}$, e.g., $\{female, male\}$ w.r.t. protected attribute $\mathcal{P}$. The computation of table $\mathcal{A}$ is described in line 3 and line 4 of the algorithm. Then, we construct table $\mathcal{B}(g, w)$ to find the total welfare with capacity w w.r.t. the protected attribute. The number of students in the protected group and the non-protected group is computed based on a given balance score θ (line 6). Table $\mathcal{B}$ is calculated in line 9 and line 10 of the algorithm.

After that, we apply a two-phase approach to solve the MFC grouping problem. In the first step, we assign students to groups based on the MFC knapsack's solution. We replace the *knapsack* function in Algorithm 8 with the new *MFC knapsack* function (Algorithm 9). In the second step, we use the group adjustment algorithm (Algorithm 7) to fine-tune the assignment.

Complexity: The MFC knapsack problem takes $\mathcal{O}(n \times C^u)$ for each topic. To solve the MFC problem, the first step consumes $\mathcal{O}(m \times n \times C^u)$, and the second step costs $\mathcal{O}(C^l \times n \times m)$. Therefore, the complexity of the MFC knapsack approach is $\mathcal{O}(n \times m)$.

Algorithm 9: MFC knapsack algorithm

Input: $\mathcal{S} = \{x_1, x_2, \ldots, x_z\}$: a set of unassigned students; C^l, C^u: capacities; $welfare_{n\times m}$: a welfare matrix; θ: balance score

Output: An optimal total welfare value

1 $avg = \dfrac{\sum_{i=1}^{n} welfare_{ij_r}}{(C^l + C^u)/2}$;

2 Let $\mathcal{A}(g, s, w), \forall g \in \{p, \overline{p}\}$, be the total welfare of the first s students in the set $\mathcal{S}$ with capacity w on group g ;

3 Initialize $\mathcal{A}(g, 0, w) \leftarrow 0$; $\mathcal{A}(g, s, 0) \leftarrow 0$;

4 $\mathcal{A}(g, s, w) \leftarrow max\{\mathcal{A}(g, s-1, w), \mathcal{A}(g, s-1, w-1) + \sum_{i=1}^{s} welfare_{ij_r}\}$;

5 Let $\mathcal{B}(g, w)$ be the total welfare of group g with capacity w;

6 $p_0^l \leftarrow \left\lceil \dfrac{C^l}{\frac{1+\theta}{\theta}} \right\rceil$; $p_0^u \leftarrow \left\lceil \dfrac{C^u}{\frac{1+\theta}{\theta}} \right\rceil$; $S_0 \leftarrow \{x \in \mathcal{S} | \psi(x) = p\}$; $S_1 \leftarrow \{x \in \mathcal{S} | \psi(x) = \overline{p}\}$;

7 $\mathcal{B}(p, w) \leftarrow max\{\mathcal{A}(p, |S_0|, w) | p_0^l \leq w \leq p_0^u\}$;

8 $\mathcal{B}(\overline{p}, w) \leftarrow max\{\mathcal{B}(p, w') + \mathcal{A}(\overline{p}, |S_1|, w - w') | C^l - p_0^l \leq w - w' \leq C^u - p_0^u, p_0^l \leq w' \leq p_0^u$, and $\frac{w'}{w-w'} \geq \theta\}$;

9 **return** $argmax\{\mathcal{B}(\overline{p}, w) | min\{\mathcal{B}(\overline{p}, w) - avg\}\}$;

6.5 Experiments

In this section, we present our experiments and the performance of our proposed approaches on two educational datasets.

6.5.1 Datasets

We evaluate our proposed methods on two variations of a real dataset often used in EDM and a real data science dataset collected at our institute. An overview of datasets is summarized in Table 6.1.

Table 6.1. An overview of the datasets for the MFC grouping problem

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

Real data science dataset[4]. This dataset was collected in a seminar on data science at our institute. Students have to register 3 desired topics out of 16 topics. The advisor will assign students into groups based on their preferences and the registration time. The data contain demographic information of students (attributes: *ID, Name, Gender*) with their preferences (attributes: *wish1, wish2, wish3*), registration time (attribute: *Time*) and priority matrix W which is represented by 16 attributes (*T1, . . . , T16*).

Student performance dataset. This dataset contains information of students in two Portuguese schools [47] with two subsets: Student-Math and Student-Por (presented in Section 3.3.1, Chapter 3). Because there is no given information about the topics and preferences of students in the original dataset, we create a *semi-synthetic* dataset by generating the preferences and the topics. The number of preferences h and the number of topics m are the main parameters of the data generator. For each student, we randomly generate h different preferred topics. Then, for each topic, we list the students who select the topic and randomly generate (different) priorities and store them in m attributes (matrix W). Hence, the semi-synthetic version has $(h+m)$ new attributes apart from the original attributes.

6.5.2 Experimental setup

Parameter selection

Naturally, a group should contain at least 2 students; therefore, the number of topics is chosen to satisfy each group of 2 members assigned to a topic. Hence, we set $m = 200$ and $m = 325$ as the number of topics for the student-Math and student-Por datasets, respectively. We set the number of wishes $h = 3$ for the real data science dataset and $h = 5$ for the student performance dataset. In addition, we set the parameters $\alpha = 1$ and $\beta = 1$ (Eq. 6.1), i.e., each component has the same weight. The balance scores θ are computed based on the datasets (Table 6.1).

Furthermore, since the real data science dataset is very small, our methods are evaluated with the lower bound C^l in the range of $(2, \dots, 8)$. For the student performance dataset, we set $C^l = (2, \dots, 18)$, as the average number of students per group should not exceed 20 [187]. The upper bound C^u is set as $C^u = C^l + 1$ for all datasets.

[4] https://tailequy.github.io/fair-grouping/

Baseline

In the experiments, we compare our proposed methods with the CPLEX integer programming model which considers both efficiency and fairness [133].

Evaluation measures

We report our experimental results w.r.t. fairness in terms of students' satisfaction, protected attribute, and cardinality with the following measures:

Nash social welfare. The Nash social welfare is computed by the Eq. 6.3. However, the number of groups (k) is determined during the group assignment process, i.e., k is different for the same set C^l, C^u, for each method. Hence, we normalize the Nash social welfare of the final group assignment by the following logarithmic function:

$$Nash = log_k \mathcal{L}(X, \mathcal{G}) \tag{6.8}$$

Balance. The fairness in terms of the protected attribute (Eq. 6.4).

Satisfaction level. It is computed by the ratio of the number of satisfied students, i.e., the students are assigned to their preferred topic, out of the total number of students:

$$Satisfaction = \frac{\mid \{i | wishes_{io} = k, i \in groups_k, o \in [h]\} \mid}{n} \tag{6.9}$$

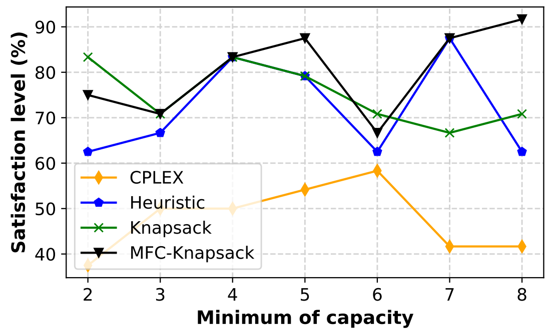
6.5.3 Experimental results

Real data science dataset

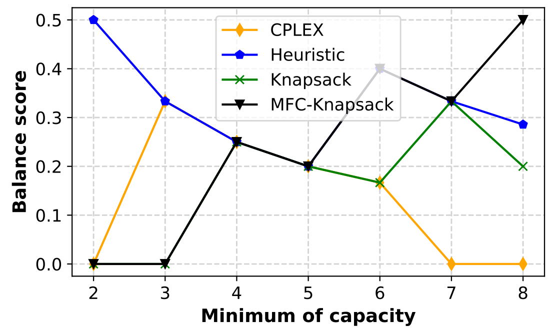
In Figure 6.2, we present the performance of proposed methods on various evaluation measures. The MFC knapsack method is better in terms of the Nash social welfare and satisfaction level (Figure 6.2-a, c). In terms of fairness w.r.t. protected attribute, the MFC knapsack method outperforms others when a group has at least 4 people (Figure 6.2-b). Besides, the CPLEX model fails to assign students while maintaining only a constant number of groups (Figure 6.2-d).

Student-Math dataset

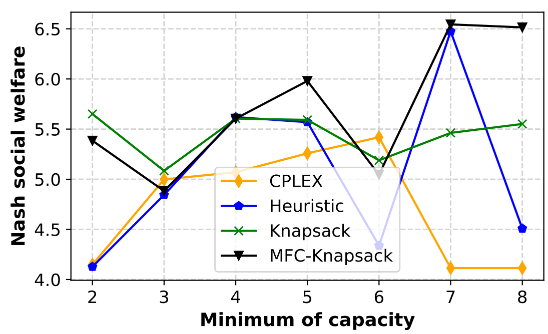
The knapsack-based approach outperforms others regarding Nash social welfare and satisfaction level in most experiments (Figure 6.3-a, c). The satisfaction level tends to decrease because students have only a limited number of preferences. When the group's cardinality increases, the desired topics become more diverse, and it is challenging to satisfy most students. In terms of fairness w.r.t. protected attribute (*gender*), the knapsack-based and MFC knapsack methods tend to achieve a higher balance score in comparison to the heuristic method (Figure 6.3-b). When groups'



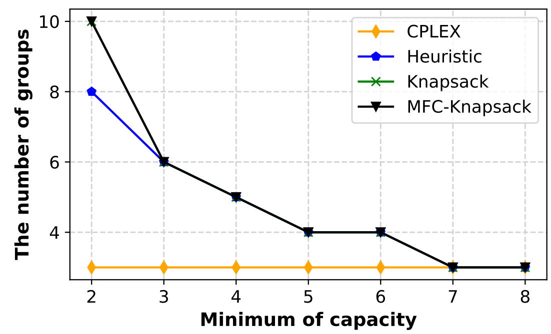
(a) Satisfaction level of students' preferences (higher is better)



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



(c) Nash social welfare (higher is better)



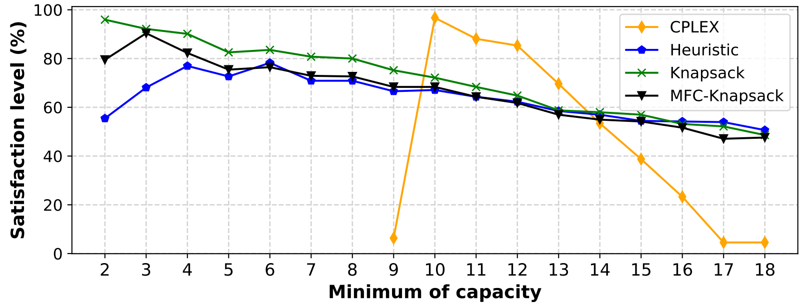
(d) Number of groups

Figure 6.2. Real data science: Performance of different methods

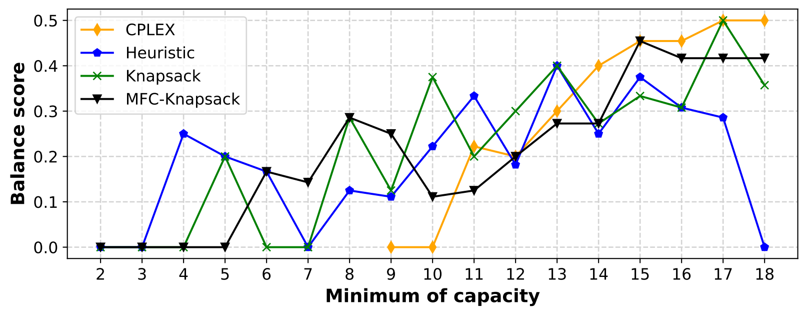
cardinality is less than 4, the greedy heuristic and MFC knapsack methods tend to create more groups than the knapsack-based method (Figure 6.3-d). The CPLEX model cannot return a solution when the groups' cardinality is less than 9 and it also fails since it is not possible to assign all students to groups.

Student-Por dataset

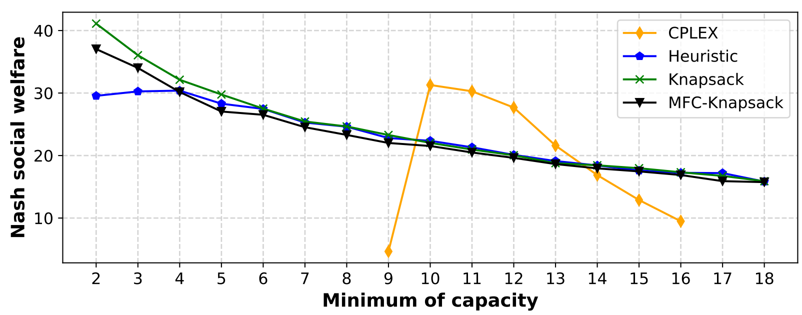
The knapsack-based method once again demonstrates the ability to create groups with higher Nash social welfare and satisfaction level than others in many cases (Figure 6.4-a and Figure 6.4-c). Regarding fairness w.r.t. gender, a higher and more stable balance score is observed in the grouping generated by the MFC knapsack model (Figure 6.4-b). The main reason for this phenomenon can be attributed to the model's emphasis on maximizing the balance constraint w.r.t. protected attribute. Besides, the MFC knapsack and greedy heuristic models divide students into more groups (Figure 6.4-d) while the CPLEX model also cannot assign all students to groups.



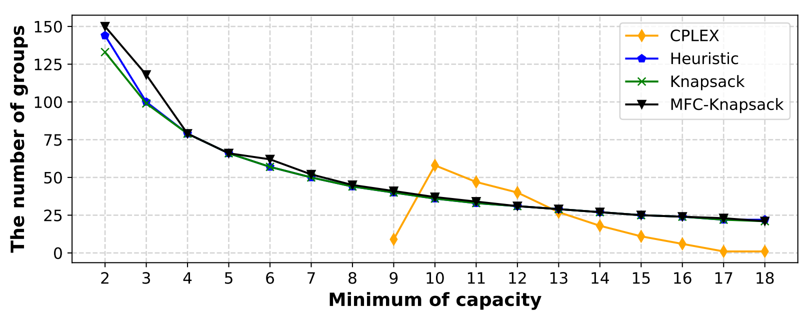
(a) Satisfaction level of students' preferences (higher is better)



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



(c) Nash social welfare (higher is better)

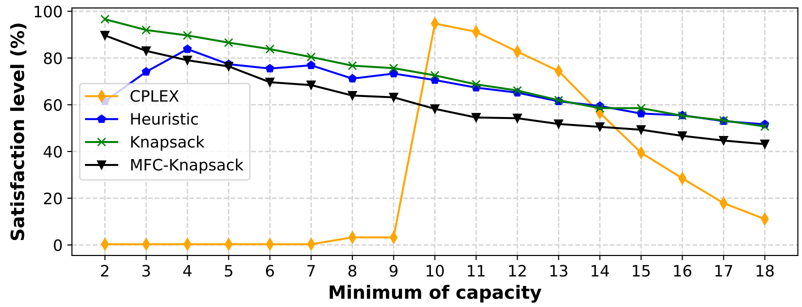


(d) Number of groups

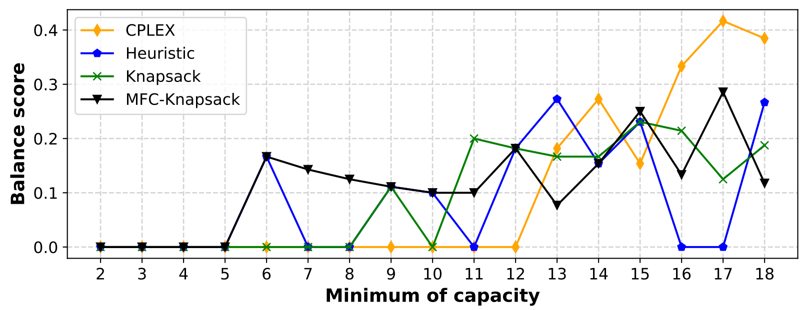
Figure 6.3. Student-Math: Performance of different methods

Impact of parameters

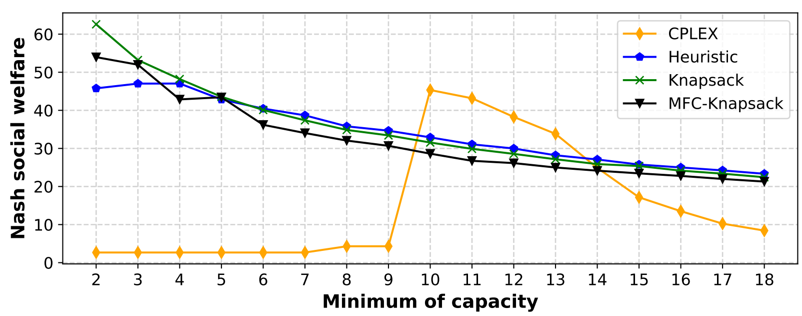
We investigate the influence of α, β parameters on the knapsack-based model. The results are illustrated in Figure 6.5 (real data science dataset), Figure 6.6 (student-



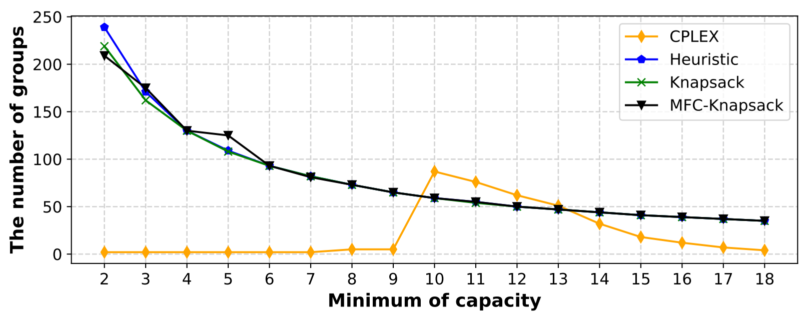
(a) Satisfaction level of students' preferences (higher is better)



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

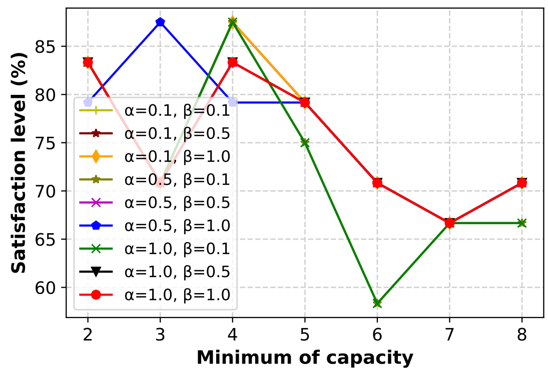


(c) Nash social welfare (higher is better)

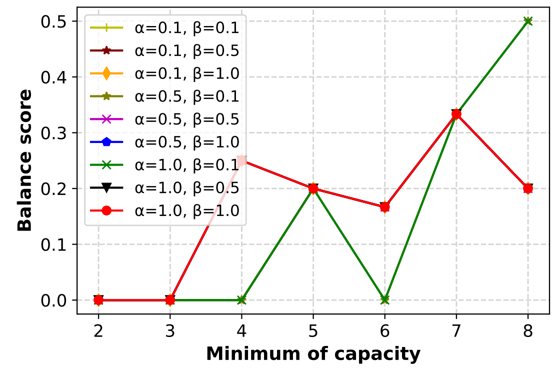


(d) Number of groups

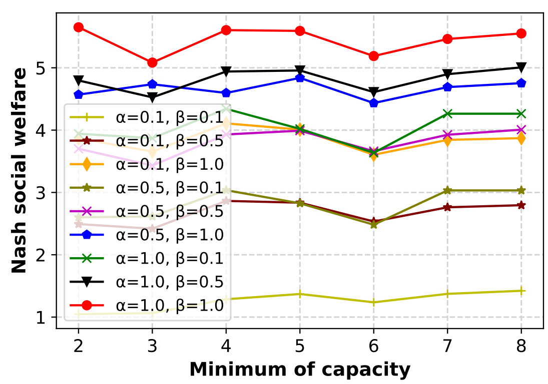
Figure 6.4. Student-Por: Performance of different methods



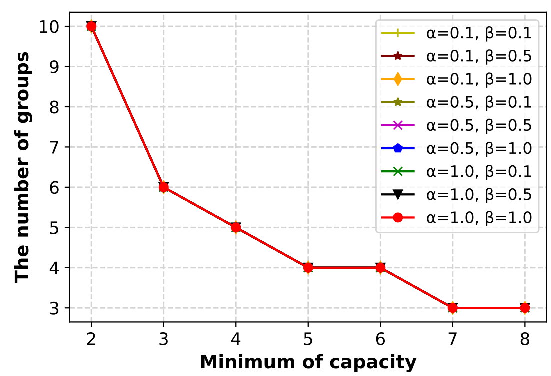
(a) Satisfaction level of students' preferences (higher is better)



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



(c) Nash social welfare (higher is better)



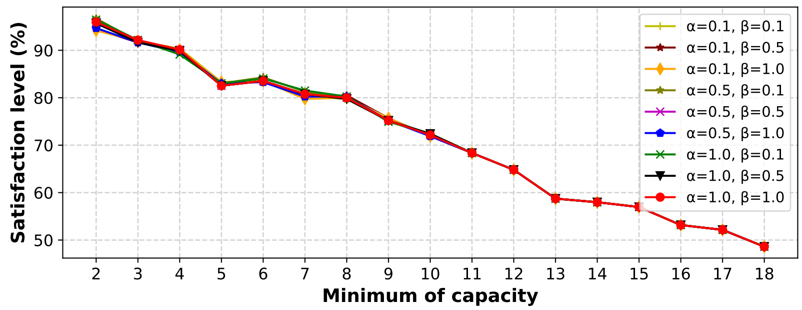
(d) Number of groups

Figure 6.5. Real data science: Impact of α, β parameters on the knapsack-based model

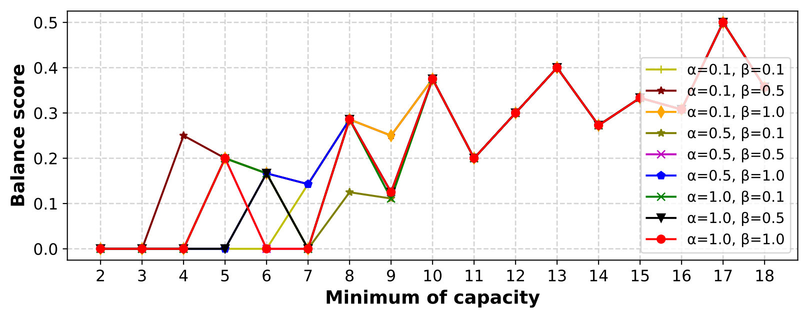
Math dataset), and Figure 6.7 (student-Por dataset). In all datasets, the knapsack-based model shows the best performance with the combination of $\alpha = 1.0$ and $\beta = 1.0$.

Summary of results

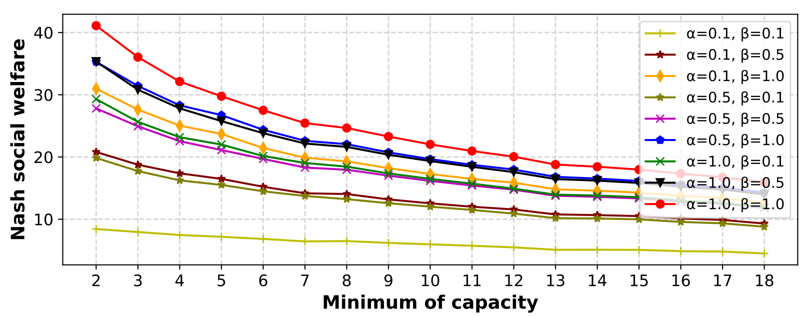
In general, the knapsack-based approach outperforms other models regarding Nash social welfare and satisfaction level. The MFC knapsack method shows its preeminence in terms of fairness w.r.t. gender in many cases, especially when the resulting groups have more members. However, in some cases, the knapsack-based approach tends to create fewer groups than the greedy heuristic method, i.e., the groups' cardinality is higher, which has both pros and cons. On the one hand, the larger groups can produce more ideas in brainstorming and discussions [31]. On the other hand, the group's performance may decline with the increase in the group's size [204].



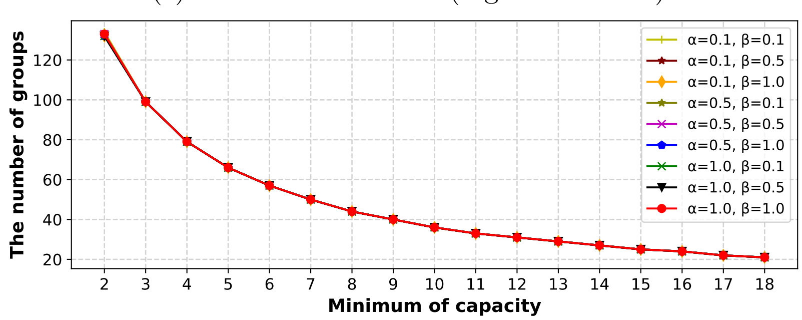
(a) Satisfaction level of students' preferences (higher is better)



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

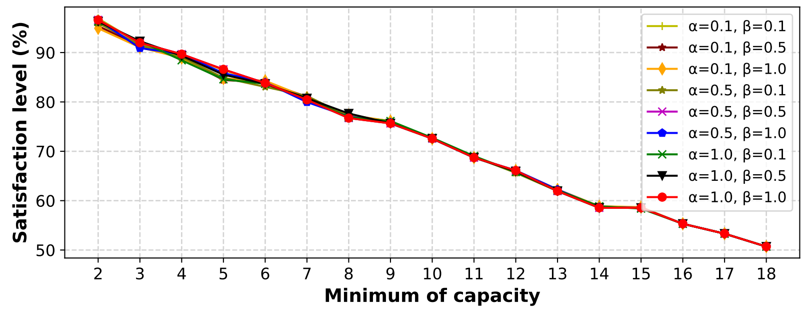


(c) Nash social welfare (higher is better)

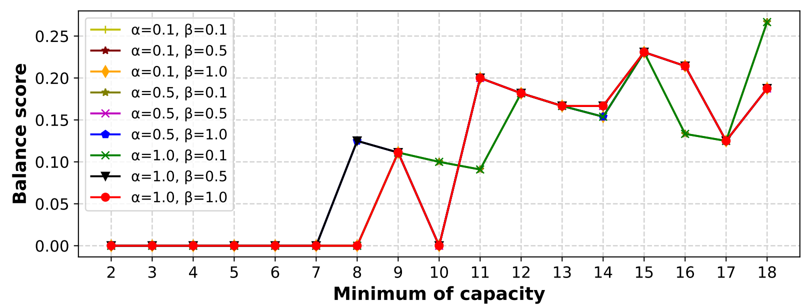


(d) Number of groups

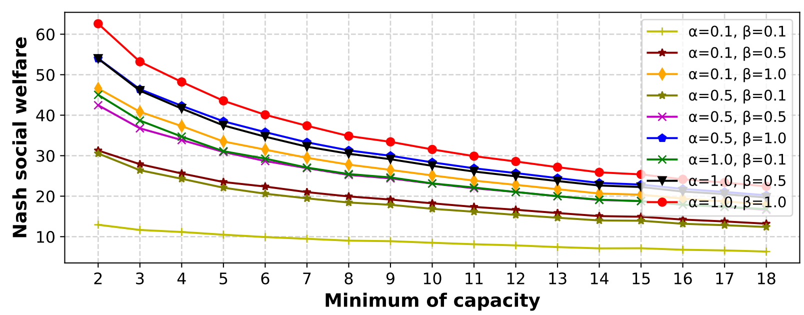
Figure 6.6. Student-Math: Impact of α, β parameters on the knapsack-based model



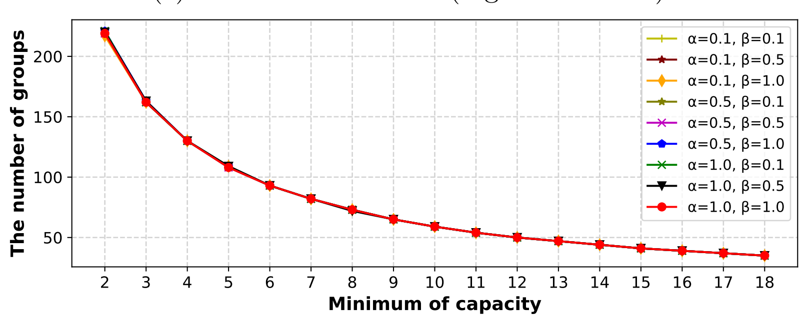
(a) Satisfaction level of students' preferences (higher is better)



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



(c) Nash social welfare (higher is better)



(d) Number of groups

Figure 6.7. Student-Por: Impact of α, β parameters on the knapsack-based model

6.6 Chapter summary

In this chapter, we introduced the MFC grouping problem that ensures fairness in multiple aspects: i) in terms of student satisfaction w.r.t. students' preferences and ii) regarding the protected attribute and maintaining the groups' cardinality within the given bounds. We proposed three methods: the greedy heuristic approach that prioritizes the students' preferences in the assignment; the knapsack-based approach with the assignment step is formulated as a maximal 0-1 knapsack problem; the MFC knapsack method considers fairness, cardinality, and students' preferences in the MFC knapsack formulation. The experiments show that our methods are effective regarding student satisfaction and fairness w.r.t. the protected attribute while maintaining cardinality within the given bounds. In the future, we aim to discuss the hardness and approximability of the problem as well as provide theoretical guarantees for proposed algorithms. In addition, we plan to extend our approach to more than one protected attribute, as well as to further investigate the groups' characteristics w.r.t. students' abilities, and other fairness notions.

7 Conclusions and outlook

In this chapter, we first summarize the main conclusions from the findings of the research described in Chapters 3, 4, 5 and 6 before discussing open issues and research directions.

7.1 Conclusions

In this thesis, we focused on answering the research questions formalized in Chapter 1 and provided contributions to fairness-aware ML in EDM with three aspects: i) bias-aware data analysis, ii) evaluation of fairness measures in EDM, and iii) development of fair-capacitated clustering models for student grouping problem in collaboration learning and multi-fair capacitated students-topics grouping models w.r.t. students' preferences.

In Chapter 3, we provided a bias-aware analysis of seven well-known educational datasets. These datasets continue to be investigated and studied in Chapters 4, 5, and 6. Through BNs generated from the datasets and exploratory analysis, we have discovered that bias appears in most datasets w.r.t. protected attributes, such as gender and race. Therefore, we suggest that the ML algorithms should consider these protected attributes to mitigate bias and achieve fairness in education.

In Chapter 4, we evaluated seven prevalent group fairness measures in student performance prediction problems. Our experimental results showed the varying behavior of fairness measures across datasets and predictive models. Based on the utility of investigated fairness measures, we remark that equal opportunity, equalized odds, and ABROCA fairness measures could be good candidates for fairness-aware ML in EDM. We also discovered that choosing the suitable passing grade threshold can have a strong effect on ensuring fairness in the output of the ML models.

In Chapter 5, we explored the student grouping problem in collaborative learning. We first introduced the fair-capacitated clustering problem, which aims to distribute students into clusters with a balanced cardinality and a fair representation w.r.t. the protected attribute. To solve the fair-capacitated clustering problem, we proposed two approaches: the hierarchical-based approach considers the cardinality constraint during the merging step, while the k-medoids fair-capacitated approach formulates the assignment step as a 0-1 knapsack problem in order to satisfy the cardinality requirement of the final clusters. The experimental results showed that our methods are effective in terms of fairness and cardinality while maintaining clustering quality. Moreover, beyond education, the fair-capacitated clustering problem and proposed approaches are applicable to other fields such as marketing studies, vehicle routing, and communication network design.

In Chapter 6, we focused on the student grouping problem w.r.t. students' preferences. We introduced the MFC students-topics grouping problem that fairly partitions students into non-overlapping groups while ensuring balanced groups with bounded cardinalities, satisfying students' preferences, and maximizing the fairness w.r.t. the protected attribute. We developed three approaches to deal with the MFC grouping problems: a greedy heuristic approach, a knapsack-based approach using vanilla maximal 0-1 knapsack formulation, and an MFC knapsack approach based on group fairness knapsack formulation. The experimental result confirmed that our methods are effective regarding student satisfaction and fairness w.r.t. the protected attribute while maintaining cardinality within the given bounds.

7.2 Outlook

First, in this thesis, we provided an analysis of a limited number of educational datasets, and it is crucial to collect and develop a benchmark educational dataset for EDM [142, 168]. However, due to privacy issues, a majority of educational datasets are non-public and hard to acquire. Besides, datasets are often collected and used for specific purposes or problems. Therefore, generating synthetic datasets is a potential solution to overcome these difficulties.

Second, we only consider a single protected attribute in the formulation and experiments. However, in practice, the roots of discrimination can be recognized with multiple protected attributes, such as the combination of race and gender. Therefore, an interesting and useful research direction is to address the issue of fairness on multiple protected attributes, aiming to tackle the more complex discriminatory issues in education systems.

Third, because fairness notions differ across disciplines, it isn't easy to evaluate the effectiveness of fairness-aware clustering algorithms. Therefore, selecting or defining the appropriate fairness measures for the educational domain is crucial and necessary, and choosing the appropriate fairness measures for fair clustering models in EDM is still a significant challenge for researchers. We will evaluate the prevalent fairness measures for fairness-aware clustering models in EDM and develop a new fairness measure for fair clustering.

Fourth, we introduce the fair-capacitated clustering and MFC students-topic grouping problem with experiments on the real and semi-synthetic datasets in Chapter 5 and Chapter 6. However, it is needed to discuss the hardness and approximability of the problem as well as provide theoretical guarantees for proposed algorithms. We will consider the theoretical aspect of the proposed algorithm as a research direction.

Fifth, explaining decisions is becoming increasingly important in education, especially in ML-based learning systems. However, the problem with many state-of-the-art classification and clustering models is a lack of transparency and interpretability [16, 59], they produce cluster assignments and predicted outcomes that are difficult to explain. Hence, a fairly obvious requirement is that clustering models and predictive models should provide explanations for the model's results in a way that humans can understand. In the future, we plan to deploy the implementation of an explainable fair clustering algorithm to achieve the clarification of the resulting clustering.

To conclude, in this thesis, we investigate the problems of fairness-aware ML in EDM with different aspects from the datasets, and fairness measures to ML models. We believe that the contributions of this thesis will be valuable to the ML and EDM communities because they have an impact in terms of technical methods and applications in educational problems. Finally, ensuring the fairness of ML algorithms in EDM will greatly contribute to improving and preserving fairness in education in general.

Bibliography

[1] Savitha Sam Abraham, Deepak Padmanabhan, and Sowmya S Sundaram. Fairness in clustering with multiple sensitive attributes. In *Proceedings of the 23rd International Conference on Extending Database Technology (EDBT)*, pages 287–298, 2020.

[2] Amjed Abu Saa, Mostafa Al-Emran, and Khaled Shaalan. Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24:567–598, 2019.

[3] Ryan P Adams. Hierarchical agglomerative clustering. In *COS 324: Introduction to Machine Learning course*. Princeton University, 2015.

[4] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna Wallach. A reductions approach to fair classification. In *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, pages 60–69. PMLR, 2018.

[5] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi. Study navigator: An algorithmically generated aid for learning from electronic textbooks. *Journal of Educational Data Mining*, 6(1):53–75, 2014.

[6] Rakesh Agrawal, Behzad Golshan, and Evimaria Terzi. Grouping students in educational settings. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1017–1026, 2014.

[7] Sara Ahmadian, Alessandro Epasto, Marina Knittel, Ravi Kumar, Mohammad Mahdian, Benjamin Moseley, Philip Pham, Sergei Vassilvitskii, and Yuyan Wang. Fair hierarchical clustering. *Advances in Neural Information Processing Systems*, 33:21050–21060, 2020.

[8] Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. Clustering without over-representation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 267–275, 2019.

[9] Shoaib Akbar, Edward Gehringer, and Zhewei Hu. Poster: Improving formation of student teams: A clustering approach. In *Proceedings of the IEEE/ACM 40th International Conference on Software Engineering: Companion (ICSE-Companion)*, pages 147–148. IEEE, 2018.

[10] Marwah Alaofi and Grace Rumantir. Personalisation of generic library search results using student enrolment information. *Journal of Educational Data Mining*, 7(3):68–88, 2015.

[11] Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat. Np-hardness of euclidean sum-of-squares clustering. *Machine learning*, 75:245–248, 2009.

[12] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. Preprocessing and analyzing educational data set using X-API for improving student's performance. In *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pages 1–5. IEEE, 2015.

[13] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. Mining educational data to predict student's academic performance using ensemble methods. *International journal of database theory and application*, 9(8):119–136, 2016.

[14] Jake Anders, Catherine Dilnot, Lindsey Macmillan, and Gill Wyness. Grade expectations: How well can we predict future grades based on past performance? *CEPEO Working Paper No. 20-14*, 2020.

[15] Henry Anderson, Afshan Boodhwani, and Ryan S Baker. Assessing the fairness of graduation predictions. In *Proceedings of the 12th International Conference on Educational Data Mining*. ERIC, 2019.

[16] Plamen P Angelov, Eduardo A Soares, Richard Jiang, Nicholas I Arnold, and Peter M Atkinson. Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(5):e1424, 2021.

[17] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. *ProPublica, May*, 23, 2016.

[18] Neal M Ashkanasy. Emotion and performance. *Human performance*, 17(2):137–144, 2004.

[19] 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.

[20] Nida Aslam, Irfan Khan, Leena Alamri, and Ranim Almuslim. An improved early student's academic performance prediction using deep learning. *International Journal of Emerging Technologies in Learning (iJET)*, 16(12):108–122, 2021.

[21] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, pages 405–413. PMLR, 2019.

[22] Ryan S Baker and Aaron Hawn. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, pages 1–41, 2021.

[23] Behdad Bakhshinategh, Osmar R Zaiane, Samira ElAtia, and Donald Ipperciel. Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23:537–553, 2018.

[24] Vaclav Bayer, Martin Hlosta, and Miriam Fernandez. Learning analytics and fairness: Do existing algorithms serve everyone equally? In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, pages 71–75. Springer, 2021.

[25] Suman Bera, Deeparnab Chakrabarty, Nicolas Flores, and Maryam Negahbani. Fair algorithms for clustering. *Neural Information Processing Systems (NIPS 2019)*, 32, 2019.

[26] Fiseha Berhanu and Addisalem Abera. Students' performance prediction based on their academic record. *International Journal of Computer Applications*, 131(5):27–35, 2015.

[27] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1):3–44, 2021.

[28] Sanyam Bharara, Sai Sabitha, and Abhay Bansal. Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies*, 23(2):957–984, 2018.

[29] Kalwant Bhopal and Martin Myers. The impact of COVID-19 on A level exams in England: Students as consumers. *British Educational Research Journal*, 49(1):142–157, 2023.

[30] Peter Blatchford, Harvey Goldstein, Clare Martin, and William Browne. A study of class size effects in english school reception year classes. *British Educational Research Journal*, 28(2):169–185, 2002.

[31] Thomas J Bouchard Jr and Melana Hare. Size, performance, and potential in brainstorming groups. *Journal of Applied Psychology*, 54(1p1):51, 1970.

[32] Javier Bravo and Alvaro Ortigosa. Detecting symptoms of low performance using production rules. In *Proceedings of the 2nd International Conference on Educational Data Mining (EDM)*. ERIC, 2009.

[33] Rainer E Burkard. Selected topics on assignment problems. *Discrete applied mathematics*, 123(1-3):257–302, 2002.

[34] Leslie Burr and DH Spennemann. Patterns of user behaviour in university online forums. *International Journal of Instructional Technology and Distance Learning*, 1(10):11–28, 2004.

[35] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. Building classifiers with independency constraints. In *2009 IEEE International Conference on Data Mining Workshops*, pages 13–18. IEEE, 2009.

[36] Toon Calders and Sicco Verwer. Three naive bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, 21:277–292, 2010.

[37] Flavio Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R Varshney. Optimized pre-processing for discrimination prevention. *Advances in Neural Information Processing Systems*, 30, 2017.

[38] Michael W Carter and Craig A Tovey. When is the classroom assignment problem hard? *Operations Research*, 40(1-supplement-1):S28–S39, 1992.

[39] Darshan Chakrabarti, John P Dickerson, Seyed A Esmaeili, Aravind Srinivasan, and Leonidas Tsepenekas. A new notion of individually fair clustering: α-equitable k-center. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 6387–6408. PMLR, 2022.

[40] Xingyu Chen, Brandon Fain, Liang Lyu, and Kamesh Munagala. Proportionally fair clustering. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, pages 1032–1041. PMLR, 2019.

[41] Yi-Chun Chen, Tim A Wheeler, and Mykel J Kochenderfer. Learning discrete bayesian networks from continuous data. *Journal of Artificial Intelligence Research*, 59:103–132, 2017.

[42] Anshuman Chhabra, Karina Masalkovaitė, and Prasant Mohapatra. An overview of fairness in clustering. *IEEE Access*, 2021.

[43] Anshuman Chhabra and Prasant Mohapatra. Fair algorithms for hierarchical agglomerative clustering. In *Proceedings of the 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 206–211. IEEE, 2022.

[44] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Neural Information Processing Systems*, pages 5036–5044, 2017.

[45] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2):153–163, 2017.

[46] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806, 2017.

[47] Paulo Cortez and Alice Maria Gonçalves Silva. Using data mining to predict secondary school student performance. *Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008)*, pages 5–12, 2008.

[48] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232, 1958.

[49] Rex Cutshall, Srinagesh Gavirneni, and Kenneth Schultz. Indiana university's kelley school of business uses integer programming to form equitable, cohesive student teams. *Interfaces*, 37(3):265–276, 2007.

[50] Kahneman Daniel. *Thinking, fast and slow.* Farrar Straus Giroux, 2017.

[51] Tobias Dantzig. *Number, the Language of Science; a Critical Survey Written for the Cultured Non-Mathematician.* Macmillan, first edition, 1930.

[52] Anupam Datta, Matthew Fredrikson, Gihyuk Ko, Piotr Mardziel, and Shayak Sen. Use privacy in data-driven systems: Theory and experiments with machine learnt programs. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1193–1210, 2017.

[53] P Deepak and Savitha Sam Abraham. Fair outlier detection. In *Proceedings of the 21th International Conference on Web Information Systems Engineering (WISE 2020)*, pages 447–462, 2020.

[54] Francesca Del Bonifro, Maurizio Gabbrielli, Giuseppe Lisanti, and Stefano Pio Zingaro. Student dropout prediction. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, pages 129–140. Springer, 2020.

[55] Soledad Delgado, Federico Morán, José Carlos San José, and Daniel Burgos. Analysis of students' behavior through user clustering in online learning settings, based on self organizing maps neural networks. *IEEE Access*, 9:132592–132608, 2021.

[56] Dong Ding, Junhuai Li, Huaijun Wang, and Zhu Liang. Student behavior clustering method based on campus big data. In *Proceedings of the 13th International Conference on Computational Intelligence and Security (CIS)*, pages 500–503. IEEE, 2017.

[57] Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring adult: New datasets for fair machine learning. *Advances in Neural Information Processing Systems*, 34:6478–6490, 2021.

[58] Neil J Dorans and Linda L Cook. *Fairness in educational assessment and measurement*. Routledge New York, NY, 2016.

[59] Filip Karlo Došilović, Mario Brčić, and Nikica Hlupić. Explainable artificial intelligence: A survey. In *Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 0210–0215. IEEE, 2018.

[60] Erik Dovgan, Bojan Leskošek, Gregor Jurak, Gregor Starc, Maroje Sorić, and Mitja Luštrek. Enhancing BMI-based student clustering by considering fitness as key attribute. In *Proceedings of the International Conference on Discovery Science*, pages 155–165. Springer, 2019.

[61] Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan. A systematic review on educational data mining. *IEEE Access*, 5:15991–16005, 2017.

[62] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226, 2012.

[63] Shadi Esnashari, Lesley Gardner, and Paul Watters. Clustering student participation: implications for education. In *Proceedings of the 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pages 313–318. IEEE, 2018.

[64] Bankole Kolawole Fasanya and Masoud Fathizadeh. Clustering from grouping: A key to enhance students' classroom active engagement. In *Proceedings of the 2019 ASEE Annual Conference & Exposition*, 2019.

[65] Graham D Fenwick and Derrick J Neal. Effect of gender composition on group performance. *Gender, Work & Organization*, 8(2):205–225, 2001.

[66] Howard L Fleischman, Paul J Hopstock, Marisa P Pelczar, and Brooke E Shelley. Highlights from pisa 2009: Performance of us 15-year-old students in reading, mathematics, and science literacy in an international context. nces 2011-004. *National Center for Education Statistics*, 2010.

[67] Till Fluschnik, Piotr Skowron, Mervin Triphaus, and Kai Wilker. Fair knapsack. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1941–1948, 2019.

[68] Marilyn Ford and Jenny Morice. How fair are group assignments? A survey of students and faculty and a modest proposal. *Journal of Information Technology Education: Research*, 2(1):367–378, 2003.

[69] Ian Foster, Rayid Ghani, Ron S Jarmin, Frauke Kreuter, and Julia Lane. *Big data and social science: A practical guide to methods and tools*. CRC Press, 2016.

[70] Bindhia K Francis and Suvanam Sasidhar Babu. Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems*, 43:1–15, 2019.

[71] Enrique García, Cristóbal Romero, Sebastián Ventura, and Carlos de Castro. An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction*, 19:99–132, 2009.

[72] Josh Gardner, Christopher Brooks, and Ryan Baker. Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19)*, pages 225–234, 2019.

[73] S Geetha, G Poonthalir, and PT Vanathi. Improved k-means algorithm for capacitated clustering problem. *INFOCOMP Journal of Computer Science*, 8(4):52–59, 2009.

[74] Mehrdad Ghadiri, Samira Samadi, and Santosh Vempala. Socially fair k-means clustering. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*, pages 438–448, 2021.

[75] 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.

[76] 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.

[77] Dana Gnesdilow, Amanda Evenstone, Julia Rutledge, Sarah Sullivan, and Sadhana Puntambekar. *Group work in the science classroom: How gender composition may affect individual performance*. International Society of the Learning Sciences, 2013.

[78] Teofilo F Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[79] Hong Guo and Miqi Wang. Analysis on the penetration of emotional education in college physical education based on emotional feature clustering. *Scientific Programming*, 2022, 2022.

[80] Randall S Hansen. Benefits and problems with student teams: Suggestions for improving team projects. *Journal of Education for Business*, 82(1):11–19, 2006.

[81] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 3323–3331, 2016.

[82] Maggie Hartnett. The importance of motivation in online learning. In *Motivation in Online Education*, pages 5–32. Springer, 2016.

[83] David Hartvigsen, Jerry C Wei, and Richard Czuchlewski. The conference paper-reviewer assignment problem. *Decision Sciences*, 30(3):865–876, 1999.

[84] HarvardX. HarvardX Person-Course Academic Year 2013 De-Identified dataset, version 3.0. 2014.

[85] Yomna MI Hassan, Abeer Elkorany, and Khaled Wassif. Utilizing social clustering-based regression model for predicting student's gpa. *IEEE Access*, 10:48948–48963, 2022.

[86] Dawn E Holmes and Lakhmi C Jain. *Introduction to Bayesian networks*. 2008.

[87] Danial Hooshyar, Margus Pedaste, and Yeongwook Yang. Mining educational data to predict students' performance through procrastination behavior. *Entropy*, 22(1):12, 2019.

[88] Chiao-Tzu Huang, Wen-Tsann Lin, Shen-Tsu Wang, and Wen-Shan Wang. Planning of educational training courses by data mining: Using china motor corporation as an example. *Expert Systems with Applications*, 36(3):7199–7209, 2009.

[89] Sue Hubble and Paul Bolton. A level results in england and the impact on university admissions in 2020-21. *House of Commons Library*, 2020.

[90] Dirk Husmeier, Richard Dybowski, and Stephen Roberts. *Probabilistic Modeling in Bioinformatics and Medical Informatics*. Springer London, 2006.

[91] Sadiq Hussain, Neama Abdulaziz Dahan, Fadl Mutaher Ba-Alwib, and Najoua Ribata. Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2):447–459, 2018.

[92] Stephen Hutt, Margo Gardner, Angela L Duckworth, and Sidney K D'Mello. Evaluating fairness and generalizability in models predicting on-time graduation from college applications. *Proceeding of the 12th International Conference on Educational Data Mining (EDM)*, pages 79–88, 2019.

[93] Vasileios Iosifidis and Eirini Ntoutsi. Dealing with bias via data augmentation in supervised learning scenarios. *Jo Bates Paul D. Clough Robert Jäschke*, 24:11, 2018.

[94] Vasileios Iosifidis and Eirini Ntoutsi. AdaFair: Cumulative fairness adaptive boosting. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 781–790, 2019.

[95] Rajendra K Jain, Dah-Ming W Chiu, William R Hawe, et al. A quantitative measure of fairness and discrimination. *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 21, 1984.

[96] Weijie Jiang and Zachary A Pardos. Towards equity and algorithmic fairness in student grade prediction. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 608–617, 2021.

[97] Matthew Jones, Huy Nguyen, and Thy Nguyen. Fair k-centers via maximum matching. In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, pages 4940–4949. PMLR, 2020.

[98] Yordan Kalmukov. An algorithm for automatic assignment of reviewers to papers. *Scientometrics*, 124(3):1811–1850, 2020.

[99] Faisal Kamiran and Toon Calders. Classifying without discriminating. In *Proceedings of the 2009 2nd International Conference on Computer, Control and Communication*, pages 1–6. IEEE, 2009.

[100] Faisal Kamiran, Sameen Mansha, Asim Karim, and Xiangliang Zhang. Exploiting reject option in classification for social discrimination control. *Information Sciences*, 425:18–33, 2018.

[101] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 1990.

[102] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. An empirical study of rich subgroup fairness for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, pages 100–109, 2019.

[103] Lorenz Kemper, Gerrit Vorhoff, and Berthold U Wigger. Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, 10(1):28–47, 2020.

[104] Anupam Khan and Soumya K Ghosh. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, 26(1):205–240, 2021.

[105] Samir Khuller and Yoram J Sussmann. The capacitated k-center problem. *SIAM Journal on Discrete Mathematics*, 13(3):403–418, 2000.

[106] René F Kizilcec and Hansol Lee. Algorithmic fairness in education. *Ethics in Artificial Intelligence in Education*, 2022.

[107] Jared E Knowles. Of needles and haystacks: Building an accurate statewide dropout early warning system in wisconsin. *Journal of Educational Data Mining*, 7(3):18–67, 2015.

[108] Dmitry Krass and Anton Ovchinnikov. The university of Toronto's rotman school of management uses management science to create MBA study groups. *Interfaces*, 36(2):126–137, 2006.

[109] M Křivánek and J Morávek. On np-hardness in hierarchical clustering. In *Compstat 1984: Proceedings in computational statistics*, pages 189–194. Springer, 1984.

[110] Jakub Kuzilek, Martin Hlosta, and Zdenek Zdrahal. Open university learning analytics dataset. *Scientific Data*, 4:170171, 2017.

[111] Marco Lam and John Mittenthal. Capacitated hierarchical clustering heuristic for multi depot location-routing problems. *International Journal of Logistics Research and Applications*, 16(5):433–444, 2013.

[112] Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin. How we analyzed the compas recidivism algorithm. *ProPublica (5 2016)*, 9(1), 2016.

[113] 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)*, pages 507–519. Springer, 2023.

[114] 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.

[115] 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.

[116] 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.

[117] 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. ERIC, 2021.

[118] 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. ERIC, 2021.

[119] 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.

[120] Tai Le Quy, Sergej Zerr, Eirini Ntoutsi, and Wolfgang Nejdl. Data augmentation for dealing with low sampling rates in NILM. In *NILM*, 2018.

[121] Kenneth Leithwood and Doris Jantzi. A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research*, 79(1):464–490, 2009.

[122] Shanfei Li. An improved approximation algorithm for the hard uniform capacitated k-median problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, volume 28 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 325–338. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014.

[123] Yuwei Li and Xiaoya Sun. Data analysis and feedback system construction of university students' psychological fitness based on fuzzy clustering. *Wireless Communications and Mobile Computing*, 2022, 2022.

[124] Oriana Licchelli, Teresa MA Basile, Nicola Di Mauro, Floriana Esposito, Giovanni Semeraro, and Stefano Ferilli. Machine learning approaches for inducing student models. In *Proceedongs of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2004)*, pages 935–944. Springer, 2004.

[125] Mihai Lintean, Vasile Rus, and Roger Azevedo. Automatic detection of student mental models during prior knowledge activation in metatutor. *International Journal of Artificial Intelligence in Education*, 21(3):169–190, 2012.

[126] Huazhong Liu, Jihong Ding, Laurence T Yang, Yimu Guo, Xiaokang Wang, and Anyuan Deng. Multi-dimensional correlative recommendation and adaptive clustering via incremental tensor decomposition for sustainable smart education. *IEEE Transactions on Sustainable Computing*, 5(3):389–402, 2019.

[127] Shuangyan Liu and Mathieu d'Aquin. Unsupervised learning for understanding student achievement in a distance learning setting. In *Proceedings of the IEEE Global Engineering Education Conference (EDUCON)*, pages 1373–1377. IEEE, 2017.

[128] Ioannis E Livieris, Vassilis Tampakas, N Karacapilidis, and P Pintelas. A semi-supervised self-trained two-level algorithm for forecasting students' graduation time. *Intelligent Decision Technologies*, 13(3):367–378, 2019.

[129] Cheng Long, Raymond Chi-Wing Wong, Yu Peng, and Liangliang Ye. On good and fair paper-reviewer assignment. In *Proceedings of the 2013 IEEE 13th International Conference on Data Mining*, pages 1145–1150. IEEE, 2013.

[130] Leo Lopes, Meredith Aronson, Gary Carstensen, and Cole Smith. Optimization support for senior design project assignments. *Interfaces*, 38(6):448–464, 2008.

[131] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 53(3):950–965, 2009.

[132] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Fairness through causal awareness: Learning causal latent-variable models for biased data. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, pages 349–358, 2019.

[133] Thomas L Magnanti and Karthik Natarajan. Allocating students to multidisciplinary capstone projects using discrete optimization. *Interfaces*, 48(3):204–216, 2018.

[134] Silvano Martello and Paolo Toth. Algorithms for knapsack problems. *North-Holland Mathematics Studies*, 132:213–257, 1987.

[135] Mihai Masala, Stefan Ruseti, Mihai Dascalu, and Ciprian Dobre. Extracting and clustering main ideas from student feedback using language models. In *Proceedings of the 22nd International Conference on Artificial Intelligence in Education (AIED)*, pages 282–292. Springer, 2021.

[136] Thomas Masterson. An empirical analysis of gender bias in education spending in paraguay. *World Development*, 40(3):583–593, 2012.

[137] George B Mathews. On the partition of numbers. *Proceedings of the London Mathematical Society*, 1(1):486–490, 1896.

[138] Daniel A McFarland, Saurabh Khanna, Benjamin W Domingue, and Zachary A Pardos. Education data science: Past, present, future. *AERA Open*, 7, 2021.

[139] M Meaney and Tom Fikes. Early-adopter iteration bias and research-praxis bias in the learning analytics ecosystem. In *LAK'19, Fairness and Equity in Learning Analytics Systems Workshop*, pages 14–20, 2019.

[140] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6):1–35, 2021.

[141] Kirsten Meyer. *Education, justice and the human good: Fairness and equality in the education system*. Routledge, 2014.

[142] Marian Cristian Mihaescu and Paul Stefan Popescu. Review on publicly available datasets for educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(3):e1403, 2021.

[143] Jeffrey A Miles and Howard J Klein. The fairness of assigning group members to tasks. *Group & Organization Management*, 23(1):71–96, 1998.

[144] James S Moore. An expert system approach to graduate school admission decisions and academic performance prediction. *Omega*, 26(5):659–670, 1998.

[145] John M Mulvey and Michael P Beck. Solving capacitated clustering problems. *European Journal of Operational Research*, 18(3):339–348, 1984.

[146] Pedro J Muñoz-Merino, Manuel Fernández Molina, Mario Muñoz-Organero, and Carlos Delgado Kloos. Motivation and emotions in competition systems for education: An empirical study. *IEEE Transactions on Education*, 57(3):182–187, 2014.

[147] Abdallah Namoun and Abdullah Alshanqiti. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1):237, 2020.

[148] John Forbes Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[149] Álvaro Alexander Martínez Navarro and Pablo Moreno Ger. Comparison of clustering algorithms for learning analytics with educational datasets. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(2):9–16, 2018.

[150] Marcos Negreiros and Augusto Palhano. The capacitated centred clustering problem. *Computers & operations research*, 33(6):1639–1663, 2006.

[151] Eirini Ntoutsi, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis, et al. Bias in data-driven artificial intelligence systems—an introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1356, 2020.

[152] Marcos Oliveira, Fariba Karimi, Maria Zens, Johann Schaible, Mathieu Génois, and Markus Strohmaier. Group mixing drives inequality in face-to-face gatherings. *Communications Physics*, 5(1):1–9, 2022.

[153] Deval Patel, Arindam Khan, and Anand Louis. Group fairness for knapsack problems. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1001–1009, 2021.

[154] Sakshi Patel, Shivani Sihmar, and Aman Jatain. A study of hierarchical clustering algorithms. In *Proceedings of the 2nd International Conference on Computing for Sustainable global Development (INDIACom)*, pages 537–541. IEEE, 2015.

[155] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11):100336, 2021.

[156] Alejandro Peña-Ayala. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4):1432–1462, 2014.

[157] David Pisinger and Paolo Toth. Knapsack problems. *Handbook of Combinatorial Optimization: Volume 1–3*, pages 299–428, 1998.

[158] Oktariani Nurul Pratiwi, Budi Rahardjo, and Suhono Harso Supangkat. Clustering multiple mix data type for automatic grouping of student system. In *Proceedings of the International Conference on Information Technology Systems and Innovation (ICITSI)*, pages 172–176. IEEE, 2017.

[159] L Ramanathan, G Parthasarathy, K Vijayakumar, L Lakshmanan, and S Ramani. Cluster-based distributed architecture for prediction of student's performance in higher education. *Cluster Computing*, 22(1):1329–1344, 2019.

[160] Christophe Reffay and Thierry Chanier. How social network analysis can help to measure cohesion in collaborative distance-learning. In *Proceedings of the International Conference on Computer Support for Collaborative Learning 2003*, pages 343–352. Springer, 2003.

[161] Pablo Reyes and Pierre Tchounikine. Mining learning groups' activities in forum-type tools. In *Proceedings of the 2005 International Conference on Computer Supported Collaborative Learning (CSCL)*. International Society of the Learning Sciences (ISLS), 2005.

[162] Nahid Rezaeinia, Julio César Góez, and Mario Guajardo. Efficiency and fairness criteria in the assignment of students to projects. *Annals of Operations Research*, pages 1–19, 2021.

[163] Shirin Riazy, Katharina Simbeck, and Vanessa Schreck. Systematic literature review of fairness in learning analytics and application of insights in a case study. In *Proceedings of the International Conference on Computer Supported Education*, pages 430–449. Springer, 2020.

[164] Nikita Romashkin, Dmitry I Ignatov, and Elena Kolotova. How university entrants are choosing their department? Mining of university admission process with FCA taxonomies. In *Proceedings of the 4th International Conference on Educational Data Mining (EDM)*, pages 229–234. ERIC, 2011.

[165] Cristóbal Romero and Sebastián Ventura. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6):601–618, 2010.

[166] Cristobal Romero and Sebastian Ventura. Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, 3(1):12–27, 2013.

[167] Cristóbal Romero and Sebastián Ventura. Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1):e1187, 2017.

[168] Cristobal Romero and Sebastian Ventura. Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1355, 2020.

[169] Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan SJd Baker. *Handbook of educational data mining.* CRC press, 2010.

[170] Cristóbal Romero, Amelia Zafra, Jose María Luna, and Sebastián Ventura. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. *Expert Systems*, 30(2):162–172, 2013.

[171] Clemens Rösner and Melanie Schmidt. Privacy preserving clustering with constraints. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107, pages 96:1–96:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.

[172] Farrukh Saleem, Zahid Ullah, Bahjat Fakieh, and Faris Kateb. Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning. *Mathematics*, 9(17):2078, 2021.

[173] Yu Sang, Heng Qi, Keqiu Li, Yingwei Jin, Deqin Yan, and Shusheng Gao. An effective discretization method for disposing high-dimensional data. *Information Sciences*, 270:73–91, 2014.

[174] Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. Fair coresets and streaming algorithms for fair k-means. In *Proceedings of the International Workshop on Approximation and Online Algorithms*, pages 232–251. Springer, 2019.

[175] Erich Schubert and Peter J Rousseeuw. Fast and eager k-medoids clustering: O (k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 101:101804, 2021.

[176] Amirah Mohamed Shahiri, Wahidah Husain, et al. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72:414–422, 2015.

[177] Ruimin Shen, Peng Han, Fan Yang, Qiang Yang, and Joshua Z Huang. Data mining and case-based reasoning for distance learning. *International Journal of Distance Education Technologies (IJDET)*, 1(3):46–58, 2003.

[178] Camelia Simoiu, Sam Corbett-Davies, and Sharad Goel. The problem of inframarginality in outcome tests for discrimination. *The Annals of Applied Statistics*, 11(3):1193–1216, 2017.

[179] Sónia Rolland Sobral and Catarina Félix de Oliveira. Clustering algorithm to measure student assessment accuracy: a double study. *Big Data and Cognitive Computing*, 5(4):81, 2021.

[180] International Educational Data Mining Society. *Online: https://educationaldatamining.org*. ERIC, Jul 2011.

[181] Xiaomei Song. The fairness of a graduate school admission test in china: Voices from administrators, teachers, and test-takers. *The Asia-Pacific Education Researcher*, 27(2):79–89, 2018.

[182] Florian Stahl and Gottfried Vossen. Fair knapsack pricing for data marketplaces. In *Proceedings of the East European Conference on Advances in Databases and Information Systems (ADBIS)*, pages 46–59. Springer, 2016.

[183] Ivan Stelmakh, Nihar Shah, and Aarti Singh. Peerreview4all: Fair and accurate reviewer assignment in peer review. *Journal of Machine Learning Research*, 22(163):1–66, 2021.

[184] Juan-Francisco Superby, JP Vandamme, and N Meskens. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Proceedings of 8th International Conference on Intelligent Tutoring systems, Educational Data Mining Workshop*, volume 32, page 234, 2006.

[185] T Thilagaraj and N Sengottaiyan. Implementation of fuzzy clustering algorithms to analyze students performance using r-tool. In *Intelligent Computing and Innovation on Data Science*, pages 287–294. Springer, 2020.

[186] Monchai Tiantong and Sanit Teemuangsai. Student team achievement divisions (stad) technique through the moodle to enhance learning achievement. *International Education Studies*, 6(4):85–92, 2013.

[187] Argelia B Urbina Nájera, Jorge De La Calleja, and Ma Auxilio Medina. Associating students and teachers for tutoring in higher education using clustering and data mining. *Computer Applications in Engineering Education*, 25(5):823–832, 2017.

[188] Sahil Verma and Julia Rubin. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness (FairWare'18)*, pages 1–7, 2018.

[189] Chau Thi Ngoc Vo and Phung Hua Nguyen. A weighted object-cluster association-based ensemble method for clustering undergraduate students. In *Proceedings of the Asian Conference on Intelligent Information and Database Systems (ACIIDS)*, pages 587–598. Springer, 2018.

[190] Yao-Te Wang, Yu-Hsin Cheng, Ting-Cheng Chang, and SM Jen. On the application of data mining technique and genetic algorithm to an automatic course scheduling system. In *Proceedings of the 2008 IEEE Conference on Cybernetics and Intelligent Systems*, pages 400–405. IEEE, 2008.

[191] Yuan Wang and Qi Wang. A student grouping method for massive online collaborative learning. *International Journal of Emerging Technologies in Learning (iJET)*, 17(3):18–33, 2022.

[192] Zhengfa Wang and Junheng Wang. Analysis of emotional education infiltration in college physical education based on emotional feature clustering. *Wireless Communications and Mobile Computing*, 2022, 2022.

[193] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

[194] Natasha Warikoo, Stacey Sinclair, Jessica Fei, and Drew Jacoby-Senghor. Examining racial bias in education: A new approach. *Educational Researcher*, 45(9):508–514, 2016.

[195] Richard Warner and Robert H Sloan. Making artificial intelligence transparent: Fairness and the problem of proxy variables. *Criminal Justice Ethics*, 40(1):23–39, 2021.

[196] Jacob Whitehill, Joseph Williams, Glenn Lopez, Cody Coleman, and Justin Reich. Beyond prediction: First steps toward automatic intervention in mooc student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM)*. ERIC, 2015.

[197] Linda F Wightman. LSAC national longitudinal bar passage study. LSAC research report series. 1998.

[198] Albert KW Wu and Chi Hong Leung. Evaluating learning behavior of web-based training (WBT) using web log. In *Proceedings of the International Conference on Computers in Education*, pages 736–737. IEEE, 2002.

[199] Yongchao Wu, Jalal Nouri, Xiu Li, Rebecka Weegar, Muhammad Afzaal, and Aayesha Zia. A word embeddings based clustering approach for collaborative learning group formation. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, pages 395–400. Springer, 2021.

[200] Wen Xiao, Ping Ji, and Juan Hu. A survey on educational data mining methods used for predicting students' performance. *Engineering Reports*, 2021.

[201] Yi Yang. Evaluation model and application of college students' physical fitness based on clustering extraction algorithm. In *Proceedings of the 4th International Conference on Information Systems and Computer Aided Education*, pages 547–552, 2021.

[202] Ying Yang, Geoffrey I Webb, and Xindong Wu. Discretization methods. *Data mining and knowledge discovery handbook*, pages 101–116, 2010.

[203] Samuel Yeom, Anupam Datta, and Matt Fredrikson. Hunting for discriminatory proxies in linear regression models. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 4573–4583, 2018.

[204] Philip Yetton and Preston Bottger. The relationships among group size, member ability, social decision schemes, and performance. *Organizational Behavior and Human Performance*, 32(2):145–159, 1983.

[205] Renzhe Yu, Qiujie Li, Christian Fischer, Shayan Doroudi, and Di Xu. Towards accurate and fair prediction of college success: Evaluating different sources of student data. In *Proceedings of the 13th International Conference on Educational Data Mining (EDM)*. ERIC, 2020.

[206] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180, 2017.

[207] Osmar R Zaiane. Building a recommender agent for e-learning systems. In *Proceedings of the International Conference on Computers in Education*, pages 55–59. IEEE, 2002.

[208] Zehui Zhan, Patrick SW Fong, Hu Mei, and Ting Liang. Effects of gender grouping on students' group performance, individual achievements and attitudes in computer-supported collaborative learning. *Computers in Human Behavior*, 48:587–596, 2015.

[209] Ningyu Zhang, Gautam Biswas, and Yi Dong. Characterizing students' learning behaviors using unsupervised learning methods. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, pages 430–441. Springer, 2017.

[210] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Record*, 25(2):103–114, 1996.

[211] Yupei Zhang, Yue Yun, Rui An, Jiaqi Cui, Huan Dai, and Xuequn Shang. Educational data mining techniques for student performance prediction: method review and comparison analysis. *Frontiers in Psychology*, 12:698490, 2021.

[212] O Zlatkin-Troitschanskaia, J Schlax, J Jitomirski, R Happ, C Kühling-Thees, S Brückner, and HA Pant. Ethics and fairness in assessing learning outcomes in higher education. *Higher Education Policy*, 32(4):537–556, 2019.

[213] Indrė Žliobaitė. On the relation between accuracy and fairness in binary classification. In *Proceedings of the 2nd workshop on Fairness, Accountability, and Transparency in Machine Learning (FATML) at ICML'15*, volume 15, 2015.

[214] Indrė Žliobaitė. Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 31(4):1060–1089, 2017.

[215] Abu Zohair and Lubna Mahmoud. Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(1):1–18, 2019.

[216] Lukas Zoubek and Michal Burda. Visualization of differences in data measuring mathematical skills. In *Proceedings of the 2nd International Conference on Educational Data Mining (EDM)*. ERIC, 2009.

Resources

Publication	Resources (code, slide, etc.)
A survey on datasets for fairness-aware machine learning	https://tailequy.github.io/fairdata-survey
Evaluation of group fairness measures in student performance prediction problems	https://tailequy.github.io/fairness-measures
Fair-capacitated slustering	https://tailequy.github.io/fair-capacitated
Multi-fair capacitated students-topics grouping problem	https://tailequy.github.io/fair-grouping
A neighborhood-augmented LSTM model for taxi-passenger demand prediction	https://neighborlstm.github.io
Taxi demand prediction using an LSTM-based deep sequence model and points of interest	https://poilstm.github.io
Data augmentation for dealing with low sampling rates in NILM	https://github.com/tailequy/Data-Augmentation-for-NILM

Curriculum Vitae

Tai Le Quy

tailequy | tailequy | tailequy.github.io | quytai3985@gmail.com | +84.933030985
Gender Male |Date of birth 03 September 1985 | Nationality Vietnamese

RESEARCH INTEREST

Data mining, fairness-aware machine learning, clustering, educational data mining, responsible AI, deep learning, time series forecasting.

EDUCATION

- **PhD in Computer Science** (July 2016 – October 2023)
  L3S Research Center – Leibniz University Hannover, Germany
  Thesis: *Fairness-aware machine learning in educational data mining*
  Supervisors: Prof. Dr. Eirini Ntoutsi, Prof. Gunnar Friege

- **Master of Information Technology** (October 2007 – June 2011)
  College of Technology, Vietnam National University, Hanoi, Vietnam
  Thesis: *Natural language processing in Vietnamese text summarization* (in Vietnamese)
  Supervisor: Prof. Dr. Pham Bao Son
  GPA: 7.53/10 (2.95/4). Thesis mark: 8.5/10 (letter A)

- **Bachelor of Education in Information Technology** (September 2003 – June 2007)
  Thai Nguyen University of Education, Vietnam
  Thesis: *Data compression with error detection and correction* (in Vietnamese)
  Supervisor: Dr. Nguyen Van Truong
  GPA: 7.56/10. Thesis mark: 9.0/10

WORK EXPERIENCE

- **Lecturer at IU International University of Applied Sciences** (March 2023 – Present)
  Courses: Artificial Intelligence (Q2, 2023), Introduction to Computer Science (Q2, 2023), Introduction to Programming with Python (Q3, 2023), Artificial Intelligence (Q3, 2023), Machine Learning - Supervised learning (Q3, 2023), Machine Learning - Unsupervised learning and feature engineering (Q3, 2023), Programming with Python (Q4, 2023) Object Oriented and Functional

Programming with Python (Q4, 2023)

- **Teaching Assistant at Free University Berlin** (April 2021 – September 2021)
  Courses: Seminar Advance Topic in Data Mining (SS 2021).

- **Teaching Assistant at Leibniz University Hannover** (October 2016 – March 2020)
  Courses: Data Mining 1 (SS 2017, SS 2018, SS 2019), Seminar Advance Topic in Data Mining (WS 2017/2018), Seminar Data Mining (SS 2019), Data Mining Lab (WS 2019/2020).

- **Student supervision at Leibniz University Hannover** (October 2017 – September 2019)
  Co-supervision students: BSc. Thi Tu Nguyen (WS 2017/2018), MSc. Bahman Askari (SS 2019).

- **Lecturer at Banking Academy of Vietnam** (August 2010 – June 2016)
  Courses: Programming fundamentals with C (SS 2012, SS 2012, SS 2014, SS 2015); Windows Programming with C# (WS 2014/2015, WS 2015/2016); Data Structure and Algorithms (SS 2013, SS 2014, SS 2015); Introduction to Informatics (WS 2010/2011, WS 2011/2012, WS 2012/2013, WS 2013/2014, WS 2014/2015, WS 2015/2016).
  Student supervision: BSc. Huong Nguyen Thu (2012), BSc. Tinh Nguyen Thi (2013), BSc. Huong Nguyen Thi (2014), BSc. Trang Ngo Thi (2015), BSc. Van Bui Quynh (2015), BSc. Huyen Le Thi Thanh (2015).

- **Lecturer at Thai Nguyen University of Education, Vietnam** (November 2007 – December 2010)
  Courses: Introduction to Informatics (WS 2009/2010)

PUBLICATIONS

1. Tai Le Quy, Gunnar Friege, and Eirini Ntoutsi. Multi-fair capacitated students-topics grouping problem. *Proceedings of the 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2023)*, 2023. https://doi.org/10.1007/978-3-031-33374-3_40

2. 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. https://doi.org/10.1007/978-981-99-0026-8_2

3. 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. https://doi.org/10.1007/978-3-031-23618-1_8

4. 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. https://doi.org/10.1007/s42979-022-01158-1

5. 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. https://doi.org/10.1002/widm.1452

6. 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.

7. 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.

8. 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. https://doi.org/10.1109/BigComp51126.2021.00022

9. 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. https://doi.org/10.1109/COMPSAC48688.2020.000-7

10. 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. https://doi.org/10.1007/978-3-030-38081-6_8

11. Tai Le Quy, Sergej Zerr, Eirini Ntoutsi, and Wolfgang Nejdl. Data augmentation for dealing with low sampling rates in NILM. In *NILM*, 2018 (Accepted).

PERSONAL SKILLS

- Languages: English: IETLS 6.0 (L: 6.0, S: 6.0, W: 6.5, R: 6.0)
- Communication skills: Good communication skills, team-work
- Organizational skills: Leadership, Classroom management
- Other skills: Curriculum design, Lesson planning, Educational assessments, MOS 2010 Master certificate

AWARDS AND FUNDING

- PAKDD 2023 conference Student registration award (May 2023)
- KDD 2022 conference Student volunteer award (August 2022)
- IEEE Big Data 2021 conference Travel award (December 2021)
- The PhD program "LernMINT: Data-assisted teaching in the MINT subjects", supported by the Ministry of Science and Culture of Lower Saxony, Germany (August 2020 – April 2023)
- PhD scholarship from the project on training PhD lecturers for universities and colleges in the period from 2010 to 2020 (Project 911), supported by the Ministry of Education and Training of Vietnam (May 2016 – April 2020)
- Part-time Toshiba Scholarship, Vietnam National University, Hanoi, Vietnam (2008 – 2009)
- Certificate of the Third Prize of the Vietnamese Informatics Olympiad contest, Vietnam (2006)
- University scholarship, Thai Nguyen University of Education, Vietnam (2003 – 2007)