Korea Advanced Institute of Technology

Department of Mathematical Sciences

Contemplation on *Fair Clustering Through Fairlets*

Author:
Junghyun Lee

Instructor:
Prof. Chang Dong Yoo

Final Project - Graph

EE531: Statistical Learning Theory, Fall 2019

December 26, 2019

Abstract

This writing is an extensive outline of the paper *Fair Clustering Through Fairlets* by Chierichetti *et al.*, including summaries of the theorems/lemmas and additional explanation on some of the literatures/points that the paper omitted.

(This is to be accompanied by the pdf file used in the final presentation. Also, the theorem/lemma/corollary numbering is arbitrary, but the statements are exact.)

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Machine learning is becoming ubiquitous in our life; from self-driving car to drug prediction in chemistry, it is becoming part of our life. It is possible that the algorithms, though they aren't inherently biased, may pick up and amplify biases already present in the training data. Thus a recent line of work has emerged on the *fairness* of some of these machine learning algorithms.

Many different notions of fairness have been studied. (Refer to [8][13] for a survey of fairness) Here we'll be focused on the notion of **disparate impact**. Since **Griggs v. Duke Power Co.**[5], disparate impact has earned its place as one of the fundamental rules for fair employment in the U.S. laws. Most notably, the 80%-rule[2] is the direct result of this. According to [10], disparate impact is *substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group.*

This work answers the question of how to formalize this notion of disparate impact in the context of clustering problem, and how to actually solve it in that framework.

1.2 Previous / Related Works

Currently, there are two "big" tracks in fairness research:

- Codifying the meaning of fairness in algorithms
- Modifying algorithms to make it achieve fair outcomes under a specific notion of fairness

In the case of disparate impact, Feldman *et al.*[3] did some work in the first track. This work, on the other hand, is closer to the second track, and is one of the first in the unsupervised learning tasks. Unlike other works, *strong guarantees* on the quality of any fair clustering solution.

The general framework of this work follows that of Zemel *et al.*[12]. Instead of trying to modify an existing algorithm to be fair, our goal here is try to **learn a set of intermediate representations to satisfy two competing goals:**

- The intermediate representation should encode the data as well as possible.
- The encoded representation is sanitized in the sense that it should be **blind to whether or not the individual is from the protected group**.

Under this framework, any classification algorithm can be transformed into a fair classifier, by simply *applying the classifer to the sanitized representation of the data.*

This work is also closely related to that of Zafar *et al.*[11]. Part of their work was focused on designing a convex margin-based classifier that maximizes accuracy subject to fairness constraints, and helps ensure compliance with a non-discrimination policy or law (e.g., a given p%-rule) This work addresses an open question in that work, which asked for a general framework to solve an unsupervised learning task respecting the p%-rule.

Chapter 2

Preliminaries

2.1 (Classical) Clustering

First, let us define what a "clustering" is:

Definition 1
Let (M, d) be a metric space, equipped with the metric function d.
Given a set of points $X \subset M$, a k-clustering of X is a partition of X into k disjoint subsets, $C_1, \ldots, C_k$, called clusters.

An alternate formulation of definition would be to use something called an *assignment function*:

Definition 2

- *A k-clustering of X is an assignment function, $\alpha : X \to [k]$.*
- *Each cluster C_i is the preimage of i under α i.e. $C_i = \alpha^{-1}(i)$*

There are many ways to quantify "how good a given clustering is". Depending on the objective, different variants of clustering problems are possible. Here, we consider two specific types of k-clustering:

Problem (k-center problem)
Given a set of points $X \subset M$, find a k-clustering of X, denoted as $\mathcal{C}$, that minimizes

$$\phi(X, \mathcal{C}) = \max_{C \in \mathcal{C}} \left[\min_{c \in C} \max_{x \in C} d(x, c) \right]$$

Problem (k-median problem)
Given a set of points $X \subset M$, find a k-clustering of X, denoted as $\mathcal{C}$, that minimizes

$$\psi(X, \mathcal{C}) = \sum_{C \in \mathcal{C}} \left[\min_{c \in C} \sum_{x \in C} d(x, c) \right]$$

2.2 Incorporating Fairness

Now we ask ourselves how to incorporate fairness into above problems. To consider a "fair" version of clustering, we first have to identify the *unprotected attribute* and *protected attribute*. We shall consider the *coordinate* as the unprotected attribute. For simplicity, let us represent the protected attribute as the *coloring* of the points. To simplify things further (as in the paper), let us only consider the case of binary coloring.

For $Y \subset X$, let us denote:

- $\chi : X \to \{\text{RED}, \text{BLUE}\}$ is the given binary coloring.
- $R(Y) = \{x \in X : \chi(x) = \text{RED}\}$, $r(Y) = |R(Y)|$
- $B(Y) = \{x \in X : \chi(x) = \text{BLUE}\}$, $b(Y) = |B(Y)|$

Definition 3

- *For $\emptyset \neq Y \subset X$, the balance of Y is defined as:*

$$\text{balance}(Y) = \min\left(\frac{r(Y)}{b(Y)}, \frac{b(Y)}{r(Y)}\right) \in [0, 1]$$

- *The balance of a clustering $\mathcal{C}$ is defined as:*

$$\text{balance}(\mathcal{C}) = \min_{C \in \mathcal{C}} \text{balance}(C)$$

- *If* $\text{balance}(Y)$ *is* 0 *(resp.* 1*),* Y *is fully unbalanced(resp. perfectly balanced)*

Let's call a clustering algorithm *colorblind* if it doesn't take the protected attribute (coloring) into its decision making. It is easy to make an instance (and it is actually common) where colorblind algorithm results in a very unfair clustering. (Unfair in the sense that the resulting clustering is very unbalanced)

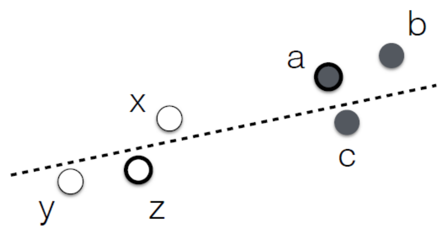


Figure 2.1: An instance where a colorblind clustering produces an unfair result

Therefore a "fair" clustering must take into account not just the position of the centers, but also the assignment function!

2.3 Fairlets and Fair Clustering

Definition 4
Let b, r be some integers such that $1 \leq b \leq r$ and $\gcd(b, r) = 1$.

- *A clustering $\mathcal{Y}$ of X is called a $(b, r)-$fairlet decomposition of X if (i) $\forall Y \in \mathcal{Y}$ $|Y| \leq b + r$ and (ii) $\text{balance}(\mathcal{Y}) = b/r = \text{balance}(X)$*
- *Each $Y \in \mathcal{Y}$ is called a $(b, r)-$fairlet, or simply fairlet.*

Intuitively, fairlet can be thought of as **a group of points that are fair and cannot be split further into true subsets that are also fair**. Observe that the balance of the original set of points is preserved while keeping each cluster "small".

Lemma 2.3.1
Let $\text{balance}(X) = b/r$ *for some integers* $1 \leq b \leq r$ *such that* $\gcd(b, r) = 1$*.*
Then there exists a $(b, r)-$fairlet decomposition of X.

This lemma tells us that every fair solution to the clustering problem induces a set of minimal fairlets, as shown in below figure:

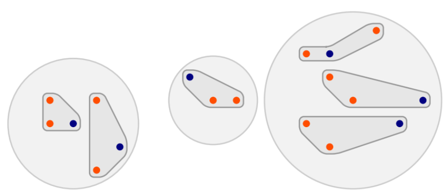


Figure 2.2: A decomposition of a fair clustering into fairlets with two red points and one blue point

Below is the final formulation of a "fair clustering":

Problem ($(t,k)-$fair center (resp. median) problem)
Partition X into $\mathcal{C}$ such that

- $|\mathcal{C}| = k$
- $\text{balance}(\mathcal{C}) \geq t$
- $\phi(X, \mathcal{C})$ *(resp. $\psi(X, \mathcal{C})$) is minimized.*

Note that if fairness is not taken into account ($t = 0$), the assignment function is implicit through a set $\{c_1, \ldots, c_k\}$ of centers i.e.

$$\alpha(x) = \text{argmin}_{i \in [k]} d(x, c_i)$$

and thus equivalent with the classical clustering algorithm. However with fairness taken into account, an explicit assignment function is required.

Chapter 3

Main results

3.1 Fairlet decomposition and fair clustering

Denote:

- $\mathcal{Y} = \{Y_1, \ldots, Y_m\}$ is a fairlet decomposition of X
- $y_j \in Y_j$ is the *center* of Y_j. (Its choice is arbitrary)
- $\beta : X \to [m]$ is the mapping from a point to the index of the fairlet to which it is mapped.

Then,

Definition 5
For a fairlet decomposition $\mathcal{Y}$, define its costs:

- *k-median cost* $= \sum_{x \in X} d\left(x, y_{\beta(x)}\right) =: \psi(X, \mathcal{Y})$
- *k-center cost* $= \max_{x \in X} d\left(x, y_{\beta(x)}\right) =: \phi(X, \mathcal{Y})$

Also, we say that a (b, r)-fairlet decomposition is optimal if it has minimum cost among all possible (b, r)-fairlet decompositions.

We shall now see how to reduce this fair clustering to a colorblind clustering. Recall that a (t, k)-fair clustering of X requires that $t \leq \text{balance}(X)$ To achieve this, we consider the **vanilla k-clustering of the centers of each fairlet** i.e. k-clustering of $\{y_1, \ldots, y_m\}$. Then we obtain a set of centers $\{c_1, \ldots, c_k\}$ and an assignment function $\alpha_Y : Y \to [k]$.

Define

$$\alpha(x) = \alpha_Y(y_{\beta(x)})$$

as the overall assignment function and denote $\mathcal{C}_\alpha$ as the clustering induced by α. Observe that balance $C_\alpha = t$. As for the cost, refer to the next lemma, which states that the cost is bounded by the sum of costs of vanilla k-clustering and the fairlet decomposition:

Lemma 3.1.1
Denote $\tilde{Y}$ as a multiset where each y_i appears $|Y_i|$ number of times. Then,

$$\psi(X, \mathcal{C}_\alpha) \leq \psi(X, \mathcal{Y}) + \psi(\tilde{Y}, \mathcal{C}_\alpha)$$

$$\phi(X, \mathcal{C}_\alpha) \leq \phi(X, \mathcal{Y}) + \phi(\tilde{Y}, \mathcal{C}_\alpha)$$

This lemma, along with previous reasoning, shows that the fair clustering problem can be reduced to

- Find a good fairlet decomposition (α-approximation)
- Solve the vanilla clustering problem on the centers of the fairlets (β-approximation)

, which is a $(\alpha + \beta)$-approximation in total!

3.2 Algorithms

3.2.1 (1, k)-fair center problem

Let us first consider the case when balance$(X) = 1$. To find a perfectly balanced clustering, we shall be utilizing a good $(1, 1)$-fairlet decomposition!

Below lemma tells us that finding an optimal $(1, 1)$-fairlet decomposition is not too heavy on computation:

Lemma 3.2.1
An optimal $(1, 1)$-fairlet decomposition for k-center can be found in polynomial time.

Recall that any fair solution induces a set of minimal fairlets. Thus, the cost of the fairlet decomposition found is at most *twice* the cost of an optimal solution to the clustering.

Lemma 3.2.2
Let $\mathcal{Y}$ be the partition found previously, and let ϕ_t^* be the cost of the optimal (t, k)-fair center clustering. Then,

$$\phi(X, \mathcal{Y}) \leq 2\phi_t^*$$

.

Remark
In the paper (Lemma 8), there is no 2 in the LHS of the inequality. [9] has observed that the fairlet decomposition cost should be bounded by the maximal diameter of a fair clustering, not the radius. He then gave an example where the above equality holds, thus disproving the original statement:

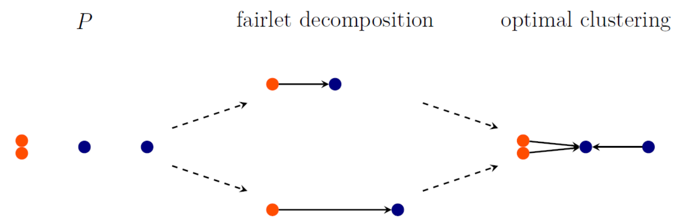


Figure 3.1: An instance where $\phi(X, \mathcal{Y}) = 2\phi_t^*$

As a final touch, let us utilize a result by Gonzalez for the vanilla k-center problem[4]:

Theorem 3.2.3 (Gonzalez, 1985)
There is an algorithm which, given a k-center instance $\mathcal{I}$, produces a 2-approximation solution to $\mathcal{I}$ in running time $O(kn)$

Combining above discussions, we have the following:

Theorem 3.2.4
The algorithm that first finds fairlets and then clusters them is a 4-approximation for the $(1, k)$-fair center problem.

3.2.2 (1/t', k)-fair center problem

Now let us consider the case when balance$(X) = t < 1$. For simplicity, assume that $t = 1/t'$ for some integer $t' > 1$ (as done in the paper) As a generalization of previous argument, we shall transform this problem into a **minimum cost flow problem (MCFP)**.

Definition 6
A flow network is a directed graph $G = (V, E)$ with a source vertex $s \in V$ and a sink vertex $t \in V$, where each edge $(u, v) \in E$ has capacity $c(u, v) > 0$, flow $f(u, v) \geq 0$ and cost $a(u, v) \in \mathbb{R}$

Now let us define what a MCFP is:

Problem (Minimum Cost Flow Problem (MCFP))
Input: *A flow network* $(G = (V, E), s, t, c, a)$ *(without the flow),* d
Constraints:

- *Capacity constraints:* $f(u, v) \leq c(u, v)$
- *Skew symmetry:* $f(u, v) = -f(v, u)$
- *Flow conservation:* $\forall u \neq s, t \ \sum_{w \in V} f(u, w) = 0$
- *Required flow from* s *to* t*:* $\sum_{w \in V} f(s, w) = \sum_{w \in V} f(w, t) = d$

Output: *Flow* $f(u, v)$ *such that* $\sum_{(u,v) \in E} a(u, v) f(u, v)$ *is minimized*

Let us construct an instance of MCFP corresponding to a clustering instance, with a parameter $\tau > 0$ that is to be determined later.

First, let us construct a directed graph $H_\tau = (V, E)$ such that:

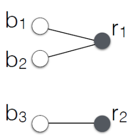
- Vertex set:

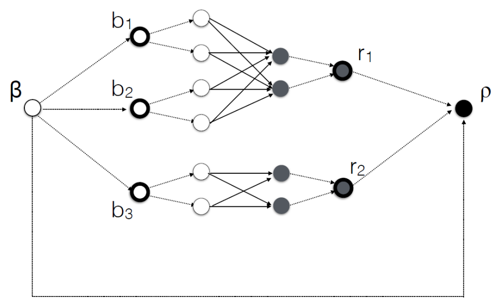
$$V = \{\beta, \rho\} \cup B(X) \cup R(X) \cup \left\{ b_i^j | b_i \in B(X) \right\}_{j \in [t']} \cup \left\{ r_i^j | r_i \in R(X) \right\}_{j \in [t']}$$

- Edge set:
  - (β, ρ) with cost 0 and capacity $\min(|B(X)|, |R(X)|)$
  - (β, b_i) and (r_i, ρ) for each $b_i \in B(X), r_i \in R(X)$, with cost 0 and capacity $t' - 1$
  - (b_i, b_i^j) and (r_i, r_i^j) for each $b_i \in B(X), r_i \in R(X), j \in [t']$, with cost 0 and capacity 1
  - (b_i^k, r_j^l) for each $b_i \in B(X), r_i \in R(X), 1 \leq k, l \leq t'$, with cost 1 if $d(b_i, r_j) \leq \tau$ and ∞ otherwise.

Define the supply and demand at every node as follows::

- Every node in $B(X)$ has a supply of 1
- Every node in $R(X)$ has a demand of 1
- β has a supply of $|R(X)|$
- ρ has a demand of $|B(X)|$
- Every other node has zero supply and demand





To relate this to our original fair clustering problem, we must be able to build a low cost $(1, t')$-fairlet decomposition, starting from a solution to the described MCF instance.

Lemma 3.2.5
By reducing the $(1, t')$*-fairlet decomposition problem to an MCFP, it is possible to compute a 2-approximation for the optimal* $(1, t')$*-fairlet decomposition for the* $(1/t', k)$*-fair center problem.*

Combining above with the result by Gonzalez gives:

Theorem 3.2.6
For any integer $t' \in \mathbb{N}$*, the algorithm that first finds fairlets and then clusters them is a 4-approximation for the* $(1/t', k)$*-fair center problem.*

3.2.3 (1/t', k)-fair median problem

Slight modification of previous argument gives us the (approx.) solution for (t, k)-fair median problem (with $t = 1/t'$).

For the perfectly balanced case, the modification changes to looking for a **perfect matching of minimum total cost** on the constructed bichromatic graph. To find $(1, t')$-fairlet decomposition for $t' > 1$, create an instance of MCF, *with (some of the) weights as the distances*

Let us utilize a result by Li & Svensson for k-median problem[6]:

Theorem 3.2.7 (Li & Svensson, 2013)
There is an algorithm which, given a k-median instance $\mathcal{I}$ and $\varepsilon > 0$, produces a $(1 + \sqrt{3} + \varepsilon)$-approximation solution to $\mathcal{I}$ in running time $O\left(n^{O(1/\varepsilon^2)}\right)$

Above discussions give the following result:

Theorem 3.2.8
For any integer $t' \in \mathbb{N}$, the algorithm that first finds fairlets and then clusters them is a $(t' + 1 + \sqrt{3} + \varepsilon)$-approximation for the $(1/t', k)$-fair median problem.

3.2.4 Hardness

We now have a theoretical framework and an actual algorithm for solving fair clustering problems. But by taking fairness into account, we have introduced some extra complexity to the classical clustering problems. As the next theorem shows, ensuring fairness (in clustering) introduces a computational bottleneck! (and a very narrow one, indeed.)

Theorem 3.2.9
For each fixed $t' \geq 3$,

- *Finding an optimal $(1, t')$-fairlet decomposition is NP-hard.*
- *Finding the minimum cost $(1/t', k)$-fair median clustering is NP-hard*

Chapter 4

Experiments

The goal of the experiments is two-fold:

- Show that the traditional algorithms for k-center and k-median tend to produce unfair clusters
- Show that the proposed algorithm outputs clusters that respect the fairness guarantees

4.1 Experiment Design

3 datasets from [7] were used:

- Diabetes (gender)
- Bank (married or not married)
- Sensus (gender)

The atttributes in the parenthesis are the protected attributes for each one. The unprotected attributes were chosen as the numeric attributes such as age, capital-gain...etc. (different for each dataset)

As for the algorithm, the flow-based fairlet decomposition algorithm (as discussed previously) was implemented. For the vanilla k-center clustering algorithm, the *greedy furthest point algorithm[4]* was used. For the vanilla k-median clustering algorithm, *single swap algorithm[1]* was used. The reason for this choice, even though it obtains 5-approximation in the worst case, is that it performs well in practice. (Refer to Kanungo *et al.*, 2002) for more information.)

In all cases, the experiment was done with $t' = 2$ i.e. aiming for balance of at least 0.5 in each cluster.

4.2 Results / Analysis

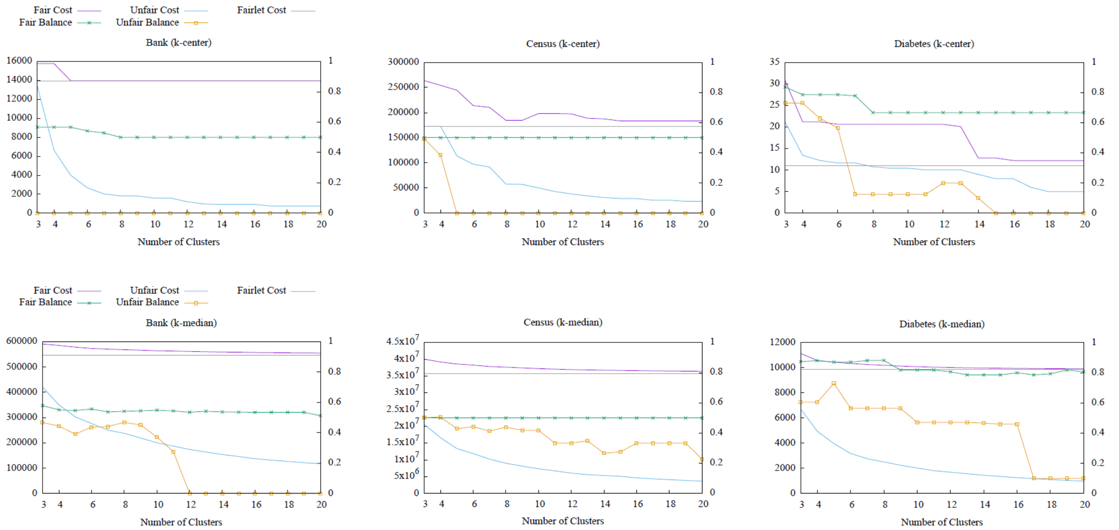


Figure 4.1: Empirical performance of the classical and fair clustering median and center algorithms on the three datasets. The cost of each solution is on left axis, and its balance on the right axis.

Observe that as expected, the balance of the solutions produced by the classical algorithms is very low. More than half of the cases show the phenomenon of balance being at 0 for high value of k.

On the other hand the fair clustering solutions maintain a balanced solution, regardless of the value of k. Not surprisingly, the balance comes with a corresponding increase in cost, and the fair solutions are costlier than their unfair counterparts. In all of the scenarios the overall cost of the clustering converges to the cost of the fairlet decomposition, which serves as a lower bound on the cost of the optimal solution.

Chapter 5

Conclusion

5.1 Summary

In summary,

- Reduction of fair clustering to classical clustering via fairlets
- Efficient approximation algorithms for finding fairlet decomposition
- Showed that fairness can introduce a computational bottleneck

5.2 Future research

Here are some possible directions for future research:

- Improve the approximation ratio of the decomposition algorithms
- Give stronger hardness results
- Extend to the case where the protected class is not binary, but can take on multiple values
- Improve the running time
  (Down to nearly linear time! *Scalable Fair Clustering (ICML 2019)*)
- New fairness constraint?
  (cf. *Proportionally Fair Clustering (ICML 2019)*)

Bibliography

[1] Vijay Arya et al. “Local Search Heuristics for k-Median and Facility Location Problems”. In: *SIAM J. Comput.* 33.3 (2004), pp. 544–562.

[2] Dan Biddle. *Adverse Impact and Test Validation: A Practitioner’s Guide to Valid and Defensible Employment Testing.* 2nd ed. Routledge, 2006.

[3] Michael Feldman et al. “Certifying and Removing Disparate Impact”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015.* 2015, pp. 259–268.

[4] Teofilo F. Gonzalez. “Clustering to Minimize the Maximum Intercluster Distance”. In: *Theoretical Computer Science* 38 (1985), pp. 293–306.

[5] *Griggs v. Duke Power Co., 401 U.S. 424 (Supreme Court of the United States 1971).*

[6] Shi Li and Ola Svensson. “Approximating k-median via pseudo-approximation”. In: *Symposium on Theory of Computing Conference, STOC’13, Palo Alto, CA, USA, June 1-4, 2013.* 2013, pp. 901–910.

[7] Mosche Lichman and Kevin Bache. *UCI Machine Learning Repository.* 2013. URL: `http://archive.ics.uci.edu/ml`.

[8] Ninareh Mehrabi et al. “A Survey on Bias and Fairness in Machine Learning”. In: *arXiv e-prints* (Aug. 2019). arXiv: `1908.09635 [cs.LG]`.

[9] Clemens Rösner. “Constrained Clustering Problems and Parity Games”. Dissertation zur Erlangung des Doktorgrades (Dr. rer. nat.) der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn. Universität Bonn, 2019.

[10] *Uniform Guidelines on Employee Selection Procedures.*

[11] Muhammad Bilal Zafar et al. “Fairness Constraints: Mechanisms for Fair Classification”. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA.* 2017, pp. 962–970.

[12] Richard S. Zemel et al. “Learning Fair Representations”. In: *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013.* 2013, pp. 325–333.

[13] Ziyuan Zhong. *A Tutorial on Fairness in Machine Learning.* July 2019. URL: `https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb`.