# Research Proposal

## Can we improve the automated decision-making systems to make public policies fairer and more efficient?

Yetong Xu
(1) 600 New Jersey Ave, NW, Georgetown University, Washington, DC,
(2) Georgetown University, Washington, DC

## Motivation and Objective

The use of big data in automated decision-making in public policy has gained popularity in recent years. While it offers many benefits, such as efficiency, consistency, and objectivity, it also has drawbacks, including the potential for bias and discrimination. As a result, people who need resources the most didn't receive them through those policies. As mentioned in the book “Automating Inequality” (Eubanks, 2018) and the article “Big Data Surveillance Policing” (Brayne, 2017), using big data in public policy automated decision-making has a lot of risks, including exacerbating existing social inequalities and the potential for errors or biases to be encoded into algorithms.

Virginia Eubanks' concept of the "digital poorhouse" further illustrates how these systems can disproportionately affect low-income populations. Also, she gives two examples: Los Angeles electronic registry for matching people to housing, and Allegheny County Pennsylvania Family Screening Tool for Child Protection. Another pertinent example is, when American universities use data to measure which students are granted student loan forgiveness, they may neglect the potential educational value that the education will generate in the future. As a result, this approach may contribute to more white or upper-middle-class students being favored (Looney, 2022).

The situations above underscore a critical need: to reassess whether we can and how we can improve automated decision-making systems to conduct public policies without hidden discrimination and deviations.

Consequently, **the objective of this study is to determine whether we can and how we can enhance automated decision-making systems to make public policies more equitable and efficient. To achieve this goal, I will conduct an experiment focusing on using machine learning algorithms to help medical centers identify high-risk populations and determine the most effective treatments that lead to significant health improvements.** We know that data in the medical and healthcare field is more private compared to other sectors, which means fewer people examine these data relatively. So, is it possible that the use of data-driven decision-making systems in the medical and healthcare field is more prone to bias and discrimination?

## Literature Review

Diverse perspectives on automated decision-making systems, views from both sides and specific applications on healthcare will be explored. While some people think that these systems lead to bias and discrimination, there are also some other positive perspectives. The article "Is an Algorithm Less Racist Than a Loan Officer?" (Miller, 2020) from the New York

Times cites Melany Anderson, who belongs to traditional underrepresented groups(as a black, divorced woman, also a contractor), as one of the examples. She has benefited from digital loan systems. For the healthcare topic I choose, Kee Yuan Ngiam(2019) explained how big data and machine learning algorithms can be applied to health-care delivery. He mentioned the advantage of machine learning algorithms is the ability to analyze diverse data types (eg, demographic data, laboratory findings, imaging data, and doctors' free-text notes) and incorporate them into predictions for disease risk, diagnosis, prognosis, and appropriate treatments.

The challenges that the application of machine learning algorithms on healthcare delivery may present will also be explored. I will find out how researchers have solved the challenges by data pre-processing, model training, or refinement of the systems with respect to the actual clinical problems.

## Data and Methodology

The data I'm going to use may come from public data sets, such as the 2020 Health and Retirement Study(National Institute on Aging, 2022), and private data sets collected by the medical centers. The potential variables may include: neighborhood, race, insurance status (Medicaid, private, uninsured), health status as indicated by preexisting conditions, and so on. For big data to be used in solving clinical problems, the clinical data must first be carefully labeled and curated. Medical data are diverse and can take the form of doctor's notes (in long-form text), clinical laboratory reports, clinical images, and information from medical devices. The labels used for these data must accurately reflect clinical reality because they will be used to train the machine learning algorithms. Any inaccuracy in labeling will severely limit the accuracy attainable by the machine learning algorithms, regardless of the effort invested in improving the algorithm. Thus, to discern data accurately, I'll invest substantial amounts of time and effort in ensuring the reliability of my data before I embark on model building. During this process, I'll also read the literature on healthcare disparities to determine: **Is there any factor that appears neutral could result in biased predictions for underrepresented groups?**

When it comes to the models, based on the data I have, I will decide whether to use PCA or other supervised learning techniques to reduce the dimensionality of the data. Then, I will decide which machine learning algorithms I should use to predict the high-risk populations and the most useful treatments. I may consider Random Forest, XG-Boost, and Neural networks.

**Random Forest** is an ensemble of decision tree classifiers, which contains many decision trees that have high prediction accuracy and weakly or even non-correlated. It can handle high-dimensional data. During the training process, it can identify variables that play an important role in classification and provide measures of the importance of these variables, **which is useful for understanding risk factors.**

**XG-Boost** is a type of boosting algorithm, which integrates many weak classifiers into a strong classifier. The basic idea of XG-Boost algorithm is to continuously generate trees and grow trees by continuously splitting feature variables. Each time a tree is generated, a new function is re-learned to fit the residual of the previous prediction, so as to continuously improve the learning quality and approximate the actual value.

**Neural networks** can handle linear and nonlinear data relationships, especially in nonlinear data fitting. Actually, I'll pay more attention to especially **convolutional neural networks**, which is good at image recognition (clinical images), and **recurrent neural network**, which is good at natural language processing (doctor's notes).

During this process, I must consider the advantages and disadvantages of these algorithms and whether their application to my healthcare data could lead to bias and discrimination. Also, my methodology is not static but will evolve as I engage more deeply with the data and continuously reassess my approach to ensure it aligns with both my research objectives and ethical standards.

## Discussion and Conclusions

In the end, I need evidence to support my conclusions. I will evaluate the models first. For example, we can use the Accuracy (ratio of correctly predicted observations to the total observations), Precision (ratio of correctly predicted positive observations to the total predicted), and Recall(ratio of correctly predicted positive observations to all observations in the actual class) to evaluate the machine learning algorithms. I may also use AUC-ROC to measure the models' ability to correctly classify outcomes across different thresholds, because it offers a more balanced perspective, especially in cases of class imbalance. What's more, I will also visualize the high-risk factors identified by the Random Forest and XG-Boost.

Then, I will collaborate with domain experts (e.g., Physicians, Nurses, Bio-Statisticians, Policy-makers, etc.) on the predictive results of my study (High-risk populations and the most effective treatments that lead to significant health improvements). Similar to the training of junior doctors, a clinical machine learning tool is best trained by incorporating real-world medical data into the model, then tuned by medical experts to improve its accuracy in predicting real cases. We'll discuss the variables I chose, whether they are reasonable and whether there are still some variables in my study that could cause hidden deviations and discrimination. **For example, we may assume that individuals who visit hospitals frequently might have higher risks. However, if there are no specific programs to assist people with lower incomes, even those at higher risk might not visit hospitals often.** Meanwhile, we will also determine whether the results I predicted make sense and accurate in the real-world. If my research encounters limitations, can they be addressed through data pre-processing or by refining the models? **Is there any hidden discrimination that we can't resolve by improving algorithms and models? And how discrimination that isn't captured by the model can be assessed and mitigated?**

If we can devise new health policies that effectively apply the most useful treatments to specific high-risk populations, according to my predictions, and if these predictions are highly accurate and applicable in real-world scenarios, then it can be affirmed that, at least within the healthcare domain of our experiment, we can improve automated decision-making systems to foster fairer and more efficient policies. Additionally, I will outline what adjustments I made for data pre-processing and application of machine learning algorithms to improve automated decision-making systems in healthcare.

If we find out there're some hidden errors, bias, or discrimination that can't be solved by data pre-processing or improving models, or if we encounter phenomena similar to those identified by Virginia Eubanks, then I can say at least within the healthcare domain, we can't foster fairer and more efficient policies **only** by applying automated decision-making systems. I have to explore further on the important roles that medical experts and administrators can play in policy-making and public services. **What part is it that big data and algorithms can't replace medical experts and administrators?**

## References

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Brayne, S. (2017). Big data surveillance: The case of policing. *American sociological review*, *82*(5), 977-1008.

Looney, A. (2022). Student loan forgiveness is regressive whether measured by income, education, or wealth. *Hutchins Center Working Papers*. Retrieved from https://www.brookings.edu/articles/student-loan-forgiveness-is-regressive-whether-measured-by-income-education-or-wealth/

Miller, J. (2020). Is an algorithm less racist than a loan officer. *New York Times*, 18. Retrieved from https://www.nytimes.com/2020/09/18/business/digital-mortgages.html

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), e262-e273. https://doi.org/10.1016/S1470-2045(19)30149-4

Weir, D. R. (2022). *Health and Retirement Study* [Data set]. The National Institute of Aging. Retrieved from https://hrsdata.isr.umich.edu/data-products/public-survey-data?_ga=2.160246261.910769648.1677873413-900753830.1677520506