Palladium
MAKE IT POSSIBLE

Embedding Equity in Big Data Analytics with DEIA-driven Analytics

Allison Fox

This is the third instalment in Palladium's Monitoring, Evaluation, Learning, and Analytics (MELA) Portfolio series on integrating diversity, equity, inclusion, and accessibility (DEIA) principles into monitoring, evaluation, learning, and analytics work in international development. This series is a follow-up to Promoting Equitable Outcomes in International Development.

Introduction

At Palladium, we use analytical methods such as inferential statistics, predictive analytics, artificial intelligence (AI) and machine learning (ML) models, natural language processing, and geospatial analyses to identify patterns in large and complex datasets. These methods capture causal relationships, predict outcomes, and generate prescriptive insights to support decision making in health and other programs.

Analysis of large amounts of data brings the potential to increase inclusivity of information and ensure that development projects are informed by data about those who are most in need of support and often have the least access to it. However, as we scale data, we also risk scaling biases in data, which in turn can lead to embedded and scaled biases in the analytical outputs. Populations who already face discrimination, such as gender minorities, have been shown to be susceptible to the consequences of these biases.

As the use of big data analytics, most notably AI and ML models, to make decisions increases, diversity, equity, inclusion, and accessibility (DEIA) must be prioritised. We stand at a crossroads; while the use of AI and other big data analytics has the potential to help us identify, understand, and address inequities in health and other sectors (e.g., by making use of more diverse and inclusive data), we must also recognise and counter the potential for these methods to overlook diversity among populations and exacerbate inequities.

Therefore, we practice DEIA-driven analytics, which utilises open-source tools that embed ethical considerations in the analytics process to ensure that results and their interpretation accurately reflect the situations and characteristics of the populations we aim to serve – and ultimately, lead to more equitable outcomes.

> "We must also recognise and counter the potential for these methods to overlook diversity among populations and exacerbate inequities."

Recognising and Addressing Bias

Over the last few years, researchers and activists have raised awareness of how AI models can discriminate based on sex, race, age, and other protected identities. MIT Professor Latanya Sweeney's 2013 study revealed that arrest record ads are more likely to be displayed when a Google search is performed for a name that is more often given to black babies than white babies, and Amazon's AI recruiting tool for reviewing resumes demonstrated a bias against women (Sweeney 2013, Dastin 2018).

Fairness metrics have recently been introduced as a tool for quantifying these biases present in machine learning models, so that steps can be taken to mitigate biases. One of the challenges with using fairness metrics is that there are many ways of defining fairness, across use cases and even for a given use case. USAID has issued a recent guidebook on Exploring Fairness in Machine Learning for International Development.

Fairness metrics provide a way of exploring how a model may be treating certain groups differently from others. Dozens of different fairness metrics exist, and these can be calculated by following formulas or by using open-source toolkits like AI Fairness 360 and Aequitas. AI Fairness 360 has code for calculating over 70 different fairness metrics. For a no-code approach, Aequitas contains a decision tree for guiding project teams to the metrics most appropriate for their use case, along with a Web Audit Tool where users can configure metrics and download a Bias Audit Report.

DEIA-Driven Analytics at Palladium

In the analytics work we do at Palladium, we use these fairness metrics when designing and deploying ML systems to make the analytic method more equitable.

For example, on the Tupime Kaunti Project in Kenya, funded by the United States President's Emergency Plan for AIDS Relief, ML techniques were used to identify patterns and predict which populations would most likely have malaria, then used the models to derive insights into who should be tested, and whether health facilities were testing the right patients.

To understand and address any bias that the model may introduce, we used fairness metrics to determine where bias may be introduced and the magnitude of potential bias. Results suggested that more females than males were tested at facilities, yet males were more likely to test positive for malaria than females.

From these data, the model predicted that males were more likely to test positive than females, but through conversations with our epidemiologists, we suspected that this prediction resulted from differences in care-seeking behavior between genders. Moreover, epidemiologists with local knowledge suspected that men were more likely to seek care only when they were severely sick.

Although the model made gender-biased predictions, it was still helpful as a clinical decision support tool at the health facility level, since it was effective in identifying which individuals were most likely to test positive.

However, this model would not be helpful for other levels of public health monitoring, such as home testing and community visits. We understood and socialised the limitations of the model and ran other analyses for these non-facility use cases that were more surveillance-based. As this DEIA-driven analytics example highlights, using fairness metrics and prioritising equity helped ensure that we thoroughly understood the use case and the application before deploying the analytics at scale.

We applied fairness metrics to a similar project in Kenya focused on HIV testing services (HTS). Palladium trained and deployed an HTS ML model that uses existing data from the Kenyan Health Management Information System's National Data Warehouse Repository, combined with geospatial datasets that describe the demographic and socio-economic context in which patients seek care, to predict the likelihood of a person testing positive for HIV during HTS client screening.

Applying this model to patient records collected since the start of 2018, we were able to sort patients by likelihood of testing positive such that the 15% of patients deemed at highest risk represented three quarters of all positive cases. We applied fairness metrics by prioritising a low False Negative rate, which means we were more comfortable with the model predicting testing for someone who does not have HIV (False Positive) than we were with the model predicting no testing for someone who does have HIV (False Negative).

Since the output of the model impacted whether people received HIV testing services, it was most important to ensure the model did not exclude people who needed these services from receiving them. If we had neglected fairness metrics and only looked at the overall accuracy of the model, we may have excluded certain populations from HIV testing, which would have been contrary to our overall goal of improving equity in access to HIV testing.

“ML models are susceptible to replicating existing biases and inequities, since they learn from the data, which itself is often biased and inequitable.”

As these examples highlight, ML models are susceptible to replicating existing biases and inequities, since they learn from the data, which itself is often biased and inequitable. Using fairness metrics, we can identify and quantify these inequities, then take steps to address them.

By applying DEIA principles and coupling them with the utilization of algorithmic fairness tools, we're actively addressing and mitigating bias and inequality in our MELA activities at Palladium.

Palladium

MAKE IT POSSIBLE

Palladium is a global impact firm, working at the intersection of social impact and commercial growth. For nearly 60 years, we've been helping our clients to see the world as interconnected – by formulating strategies, building partnerships, mobilising capital, and implementing programs that have a lasting social and financial impact. We simply call this "positive impact".

We work with corporations, governments, investors, communities, and civil society. With a global network operating in over 90 countries, Palladium is in the business of making the world a better place.

Contact info@thepalladiumgroup.com to learn more.

www.thepalladiumgroup.com