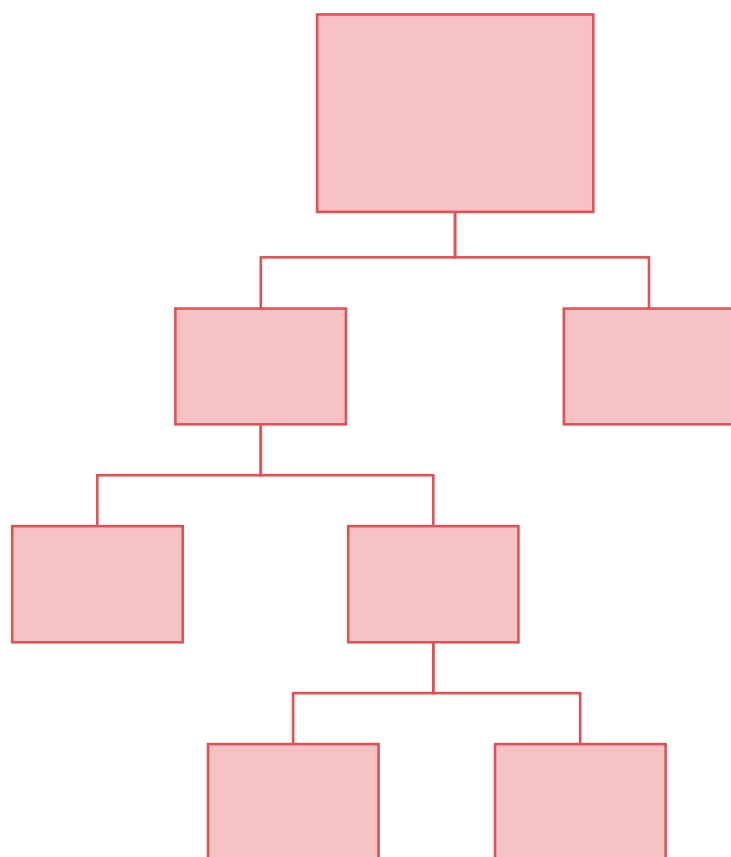# Policy Paper

No. 2020/01

Classification and Regression Trees (CART):

A User Reference Guide for Identifying Those Left Furthest Behind

Produced by UNESCAP Social Development Division

Sustainable Socioeconomic Transformation Section

10 REDUCED INEQUALITIES

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

Table of Contents

Introduction – Why LNOB?

The 2030 Agenda for Sustainable Development makes a call to "leave no one behind" and to "endeavour to reach the furthest behind first" on the path towards sustainable development. The Agenda reflects a consensus on the need to move towards more egalitarian, cohesive and solidarity-based societies, by promoting an inclusive development model leading to a sustainable future.

In practice, leaving no one behind (LNOB) means moving beyond assessing average and aggregate progress, towards ensuring progress for all population groups at a disaggregated level. The LNOB trees were developed with the aim of shedding light on how **various circumstances can intersect to create inequality** in access to basic opportunities.
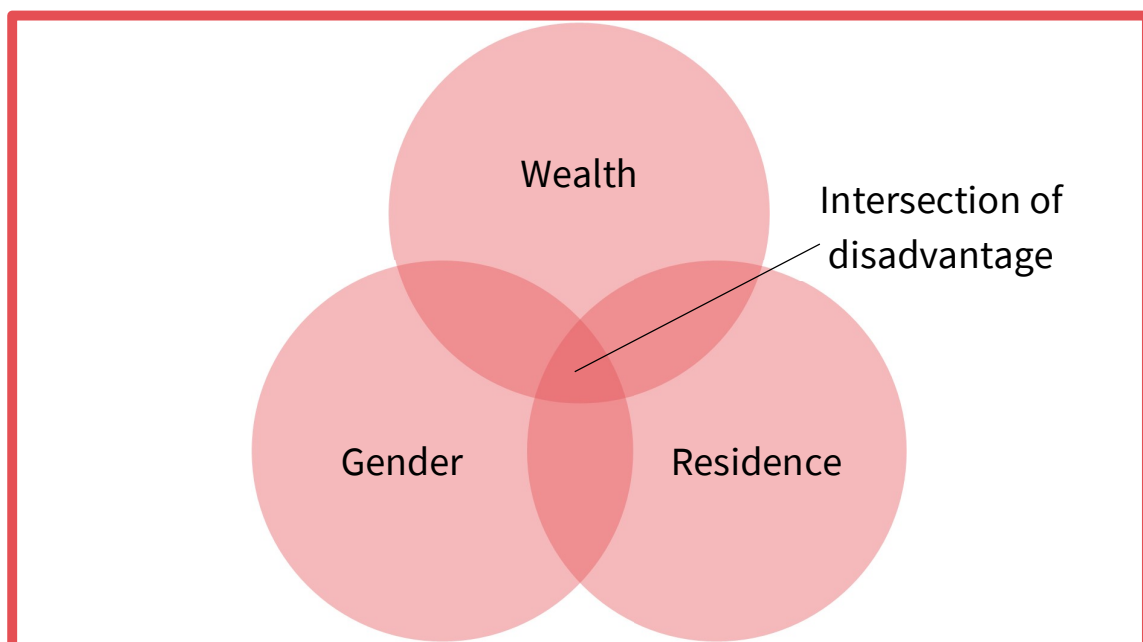
Basic Assumptions

1. Access to key services and opportunities should be universal.

In the LNOB trees, these services and opportunities are presented as **indicators**. Some indicators represent opportunities, such as access to electricity, while others represent barriers, such as childhood malnutrition.

2. Individual circumstances should not impact access.

To measure inequality in access, the LNOB trees disaggregate survey data into various groups defined by **circumstances**, such as age, sex, and area of residence. In doing so, we can better understand how different circumstances intersect to reveal disparities in the groups that are furthest ahead and furthest behind.

The Data

There are two sources of data for UNESCAP's LNOB Trees:

Multiple Indicator Cluster Surveys ("MICS")

Click to go to the MICS website

Demographic and Health Surveys ("DHS")

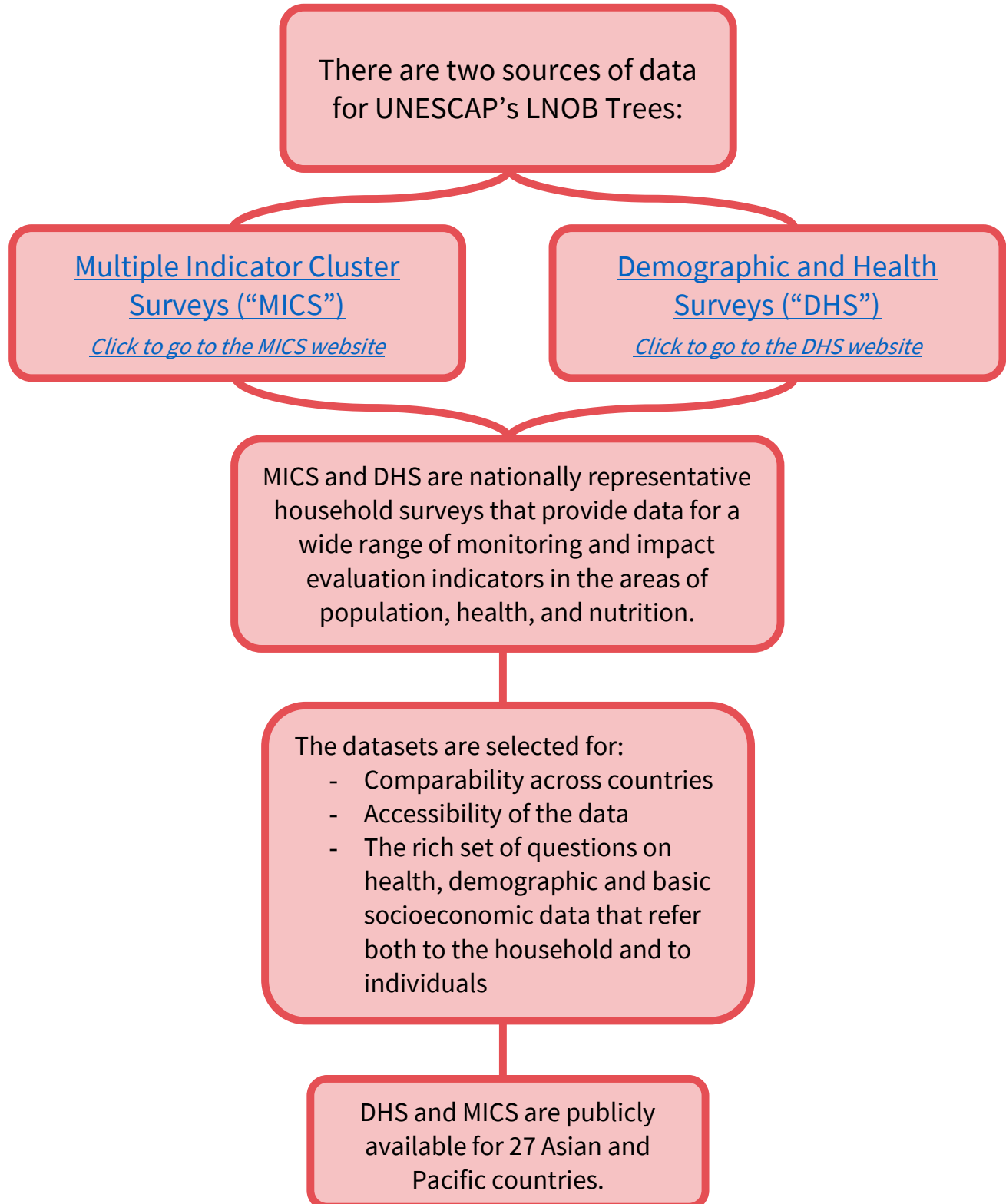*Click to go to the DHS website*

MICS and DHS are nationally representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition.

The datasets are selected for:
- Comparability across countries
- Accessibility of the data
- The rich set of questions on health, demographic and basic socioeconomic data that refer both to the household and to individuals

DHS and MICS are publicly available for 27 Asian and Pacific countries.

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

Interpreting Indicators

Each LNOB Tree is constructed around a single **indicator**, which is selected based on its theoretical relevance for a Sustainable Development Goal. For example, we might be interested in examining Access to Electricity as an indicator because Goal 7 of the UN SDGs calls for universal access to affordable and sustainable modern energy (Annex A). Indicators are used to measure the level of access to a basic opportunity or, alternatively, the prevalence of a barrier to sustainable development. Let's take a closer look at two permanent indicators available for UNESCAP's LNOB Trees: **access to electricity**, which is an opportunity; and **childhood stunting**, which is a barrier.

Indicator	Type	Data Source	Reference Group	Related SDG
Access to electricity	Opportunity	DHS/MICS	Household	7.1.1 Proportion of population with access to electricity
Childhood stunting	Barrier	DHS/MICS	Children aged 0-5	2.2.1 Prevalence of stunting among children under 5 years of age

Surveying Circumstances

People can be sorted into groups based on various **circumstances**, such as age, place of residence, or level of education. By creating groups in this way, the LNOB Trees can be used to reveal disparities between groups for various indicators. Let's examine some circumstances more closely:

Circumstance	Example Group Identities
Age	0-17; 18-29; 30-44; 45-65; 65+
Place of Residence	Urban; Suburban; Rural
Level of education	No formal education; primary education; secondary education; post-secondary education

By utilizing indicators and circumstances, the LNOB Trees make it easy to identify the social groups that are behind in access to basic opportunities. This information can be used by policymakers to design social protection programmes that are effective at closing gaps and achieving sustainable development goals. Annex A contains a complete list of the indicators and circumstances utilized in UNESCAP's LNOB Trees.

COVID-19 Adaptation

The CART methodology is highly flexible and can be molded to include new indicators and circumstances as necessary. UNESCAP has adapted the methodology to include indicators relating to the ongoing COVID-19 pandemic to help policymakers identify the layers of vulnerability in protecting against the virus. If left unaddressed, vulnerabilities can exacerbate transmission and multiply the impacts of the crisis. A more detailed understanding of the groups left furthest behind in ability to protect themselves strengthens the knowledge and evidence base for more informed policy responses.

Using MICS and DHS data, ESCAP has developed a composite indicator for COVID-19 protection comprising five key measurements in the categories of access to information, access to preventive measures and access to social distancing.[1]

Category	Measurement
Ease of access to information	Mobile phone ownership or access to internet, TV, or radio
Ease of access to preventive measures	The individual resides in a household that has: • water pipes into the dwelling or yard, or other private water source • a handwashing facility on premises with soap and clean water available • a toilet not shared with other households
Ease of access to social distancing	• No more than two people per sleeping room

---

[1] These measures are broadly drawn from research on key protection indicators for vulnerable groups, including from the paper "**Can the World's Poor Protect Themselves from the New Coronavirus".** Caitlin S. Brown, Martin Ravallion, Dominique van de Walle. (NBER Working Paper 27200)

The lack of access to information or preventive and social distancing measures intensifies vulnerability to COVID-19 and can create environments where community transmission is more likely to occur. It is thus vital that policymakers prioritize these vulnerable groups in their social protection responses. For example, governments can help to ensure access to clean water, soap, and medicines for the most vulnerable groups. Policies aimed at closing the digital divide can also ensure vulnerable populations have access to the most up-to-date information and recommendations concerning the pandemic. In addition, efforts should be made to ensure universal access to health services, such as mobile services for remote populations. Continued income support, such as cash transfers, paid sick leave, or unemployment benefits, can also help ensure compliance with quarantines or temporary lockdowns, reducing the risk of further contagion.

By adding to the evidence base of which groups are most likely to be left furthest behind, the LNOB trees and the CART methodology can help policymakers develop more comprehensive, inclusive, and effective social protection responses. The success of these responses will be vital both for protecting those most vulnerable to the ongoing COVID-19 pandemic as well as fostering societies that can be more resilient in the face of future crises.

Reading an LNOB Tree
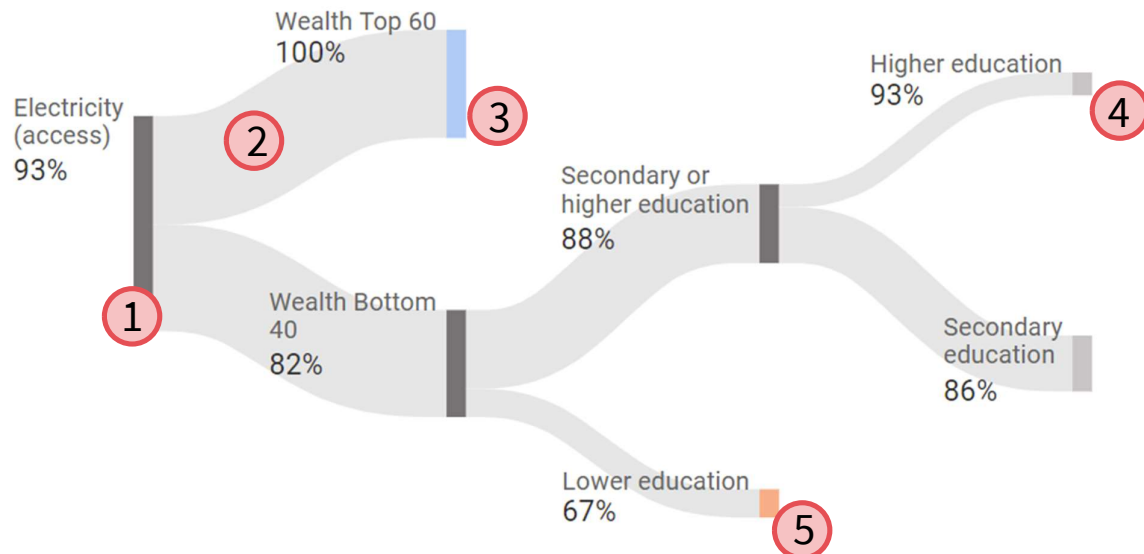
1. This is a **node**. It represents a certain social group. The first node in the tree is called the "root node" and represents the average access across the entire sample. For this tree, we can see the average rate of access to electricity is 93% for all people who were surveyed.

2. This is a **branch**. Each branch splits the sample of survey respondents by a different circumstance – in this case, household income. The branch width varies based on the proportion of the sample that has been split into a new group.

3. This is the group that is **furthest ahead**. Their average rate of access to electricity is **100%**.

4. Trees are built recursively, with each new split building upon the previous splits. This node shows the average rate of access to electricity is **93%** for people in the bottom 40 percent of household income and with higher education.

5. This is the group that is **furthest behind.** Their average rate of access to electricity is **67%**.

Understanding CART

LNOB trees use the **Classification and Regression Tree (CART)** methodology to determine the best split at each node. For each node, an algorithm evaluates all possible splits to determine which circumstance explains the most variance in our indicator. Each node is split in this way until further splits cease to explain enough variance, or the sample size becomes too small.

The logic of CART closely imitates how most people intuitively make decisions. For example, let's think about how we choose which bunch of bananas to buy. We might prioritize bananas that have as few blemishes as possible. Perhaps we have other preferences – such as the size of the individual fruit or the number of bananas in the bunch – but no matter what, we're disinclined to buy a bunch of bananas if the fruit are bruised. If we constructed a decision tree of our banana selection process, the number of blemishes would be the first split in our tree, with other characteristics being ranked lower but perhaps still influencing our final selection.
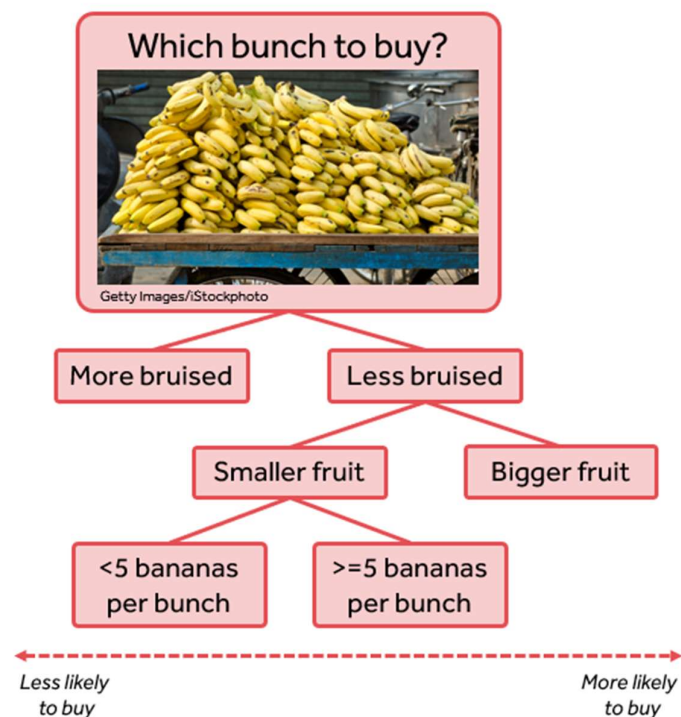
CART In Practice

Let's look at how the CART methodology works in practice with LNOB trees. If our indicator was "Access to Electricity", the CART algorithm uses a **splitting criterion** to determine which circumstance explains the greatest amount of variance in which groups of people have access to electricity.

To the right, we can see two possible splits for our indicator: the top splits on "education" (less vs. more), while the bottom splits on "residence" (urban vs. rural). The "residence" split results in groups that are more homogenous – in other words, the bottom split satisfies our splitting criterion by explaining the most variance between our groups. "Residence" would become the first circumstance in our LNOB tree.
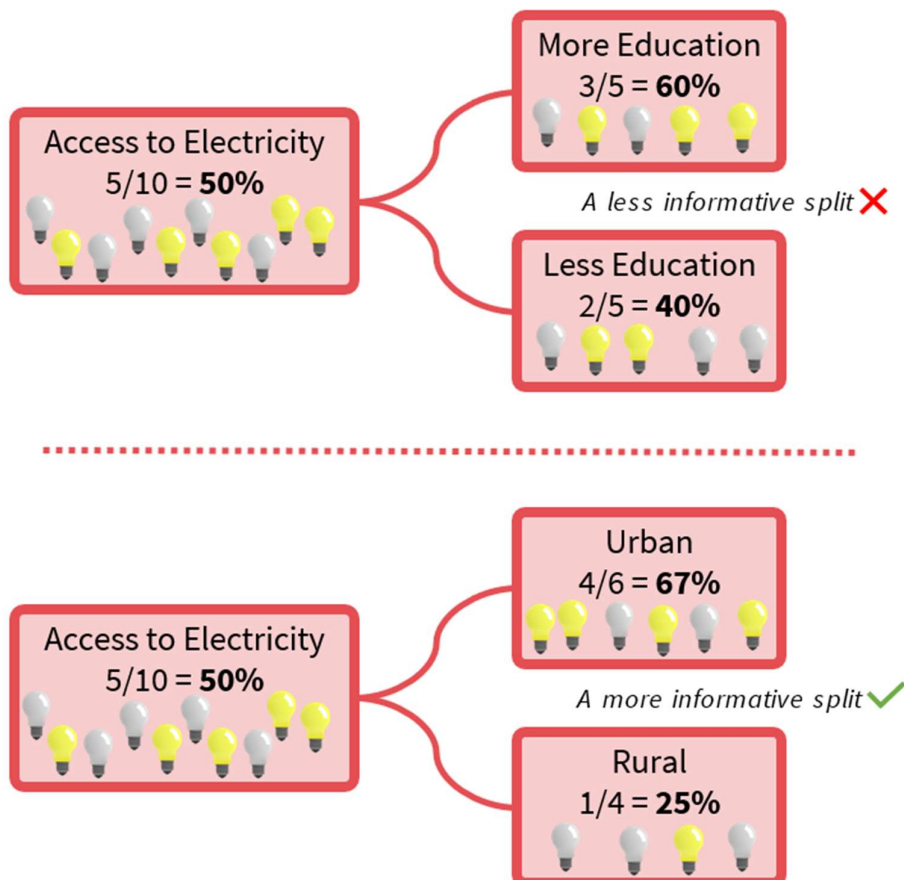
Building Trees with CART
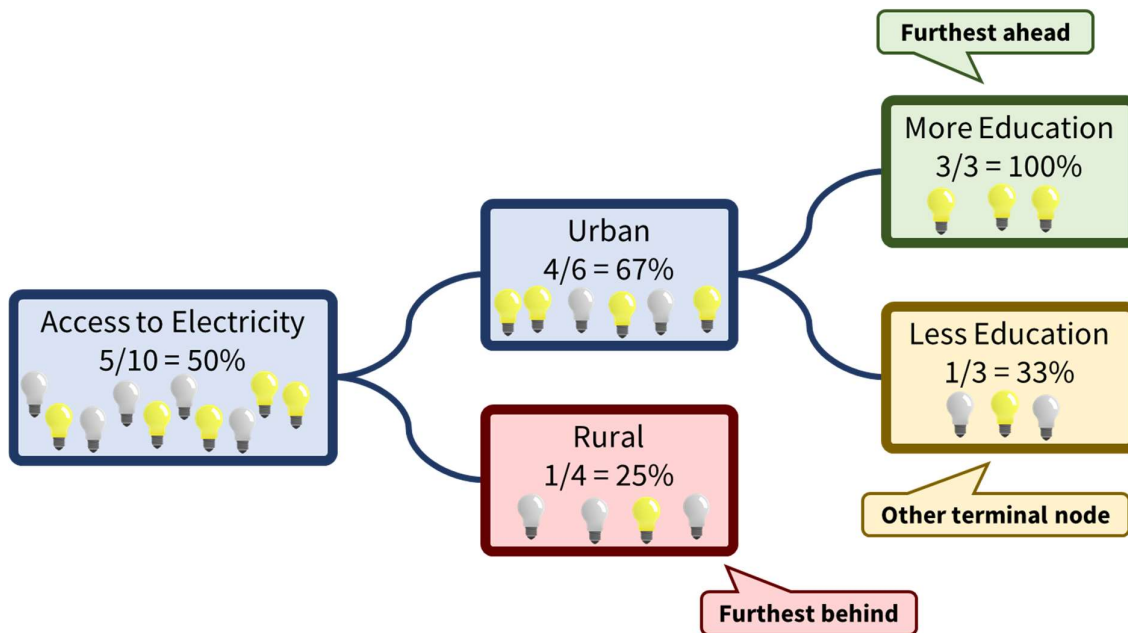
CART analysis doesn't stop at the first split – it analyzes subsequent nodes to see if more variance can be explained from further splits. In the example above, the "urban residence" node is further split by "education" (more vs. less). We can see that people who reside in urban areas with more education are the group furthest ahead, with 100% access to electricity.

The CART algorithm did not split the node for "rural residence." Either additional splits didn't satisfy the splitting criterion, or the sample size was too small to be split further. This means that people living in rural areas are the group furthest behind, with only 25% access to electricity.

D-Index Explained

D-Index is a way to measure inequality of access across all groups in a sample. Let's look at the countries in the example below: the average rate of access to electricity in each country is 65%, but they have different rates of access for different social groups. D-Index values range from 0-1: a higher number indicates more inequality, while a lower number indicates less inequality.[2] D-Index is comparable to the Gini coefficient, which is frequently used to measure household-level economic inequality.

Access to Electricity		
Country A	**Country B**	**Country C**
Average rate of access: 65%	Average rate of access: 65%	Average rate of access: 65%
Urban: 88%; Rural: 51%	Urban: 73%; Rural: 53%	Urban: 68%; Rural: 62%
Wealthy: 92%; Impoverished: 43%	Wealthy: 76%; Impoverished: 60%	Wealthy: 66%; Impoverished: 65%
D-Index: **0.73**	**D-Index:** **0.41**	**D-Index:** **0.08**

While LNOB trees are useful for comparing rates of access across groups, the D-Index provides a single number that can summarize inequality of access for all groups in a sample. D-Index facilitates easy comparisons between samples, such as between provinces, countries, or groups of countries.

---

[2] D-Index has been adapted so that the value of a barrier (e.g. childhood malnutrition) still has the same interpretation as that of an opportunity: the lower the D-index the lower the inequality. In general, the D-Index measures the distribution of a positive outcome. Malnutrition is not a positive outcome, but rather a barrier for a child's development prospects. To calculate the D-Index for this barrier, while keeping the same interpretation as for other positively defined indicators (opportunities), the absence of stunting is first calculated. The remaining calculations follow the same formula as for standard positively defined indicators.

Specifications and Limitations

LNOB Trees can…

- Identify the groups that are the furthest behind based on their shared circumstances
- Reveal which circumstances are associated with the biggest gaps in access to basic opportunities
- Help policymakers understand whom to prioritize for interventions to reduce inequality

LNOB Trees can't…

- Explain causal relationships between circumstances and outcome
- Predict who will be the furthest behind in the future or in a different sample
- Be easily compared to one another over time, as splits may involve different circumstances

UNESCAP LNOB Tree Specifications

- Each tree is restricted to a depth of six or fewer levels of circumstances
- End nodes must have no fewer than 49 observations (absolute criterion) and no less than 9% of weighted sample size (relative criterion) unless otherwise specified
- Indicators and circumstances are selected based on theoretical relevance for the 2030 UN SDGs as well as data availability

Beyond the Trees

The Sustainable Socioeconomic Transformation Section (SETS) at UNESCAP SDD can provide LNOB analysis beyond the trees. These services include:

- Subnational LNOB analysis
- Regional comparisons by D-Index value
- Adoption of new indicators to measure alternative dimensions of inequality
- Inclusion of specific circumstances, such as being a migrant, depending on availability of DHS and/or MICS data

Example: Province-level analysis of skilled birth attendance in Afghanistan

Province Analysis

Skilled birth attendance during childbirth
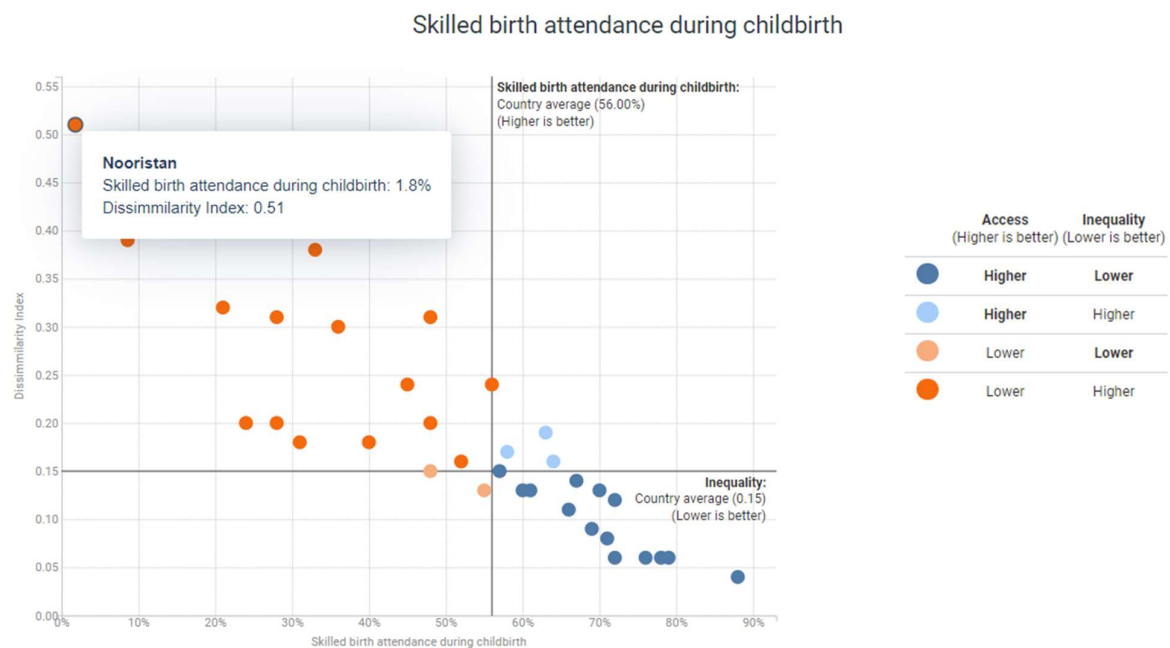
For inquiries on custom LNOB analysis, send an email with the nature of your request to: escap-sdd@un.org

Creating Custom LNOB Trees

UNESCAP's LNOB trees are intended to be both a tool for policymakers and a methodological proof of concept. UNESCAP SDD accepts requests for custom analyses of inequality in countries in Asia and the Pacific, but we encourage other parties to adopt this methodology for their own research on inequality and sustainable development.

For a more detailed resource on UNESCAP's methodological approach to LNOB trees, see:

Savic, Predrag and Yichun Wang. "Leaving No One Behind: A Methodology To Identify Those Furthest Behind In Accessing Opportunities In Asia And The Pacific." UNESCAP Social Development Working Papers, No. 2019/06, published 27/12/2019: https://www.unescap.org/sites/default/files/LNOB_AP.pdf

LNOB tree checklist

To create your own LNOB trees, you will need…

- ☐ An indicator that represents either access or a barrier to a basic opportunity (such as financial inclusion or childhood malnutrition)

- ☐ A data source that can be disaggregated by theoretically relevant circumstances, such as wealth, gender, or area of residence

- ☐ Proficiency with statistical analysis and statistical software such as R, Stata, or SPSS. UNESCAP LNOB trees are created in R, a free and open-source software.

Annex A – UNESCAP LNOB Indicators, part 1

	Opportunities/Barriers (response variable)			Circumstances used to determine the furthest behind/ best-off groups (independent variables)								Closest SDG indicator reference
	Indicator	Survey used	Reference population in survey	Wealth: Bottom 40- Top 60	Residence: Urban - Rural	Education: No/Primary - Secondary - Higher	Sex: Male- Female	Children: Yes-No, Number	Age: 15-24, 25-34, 35-49	Marital status (Single, currently/ formerly married or in a union)	Household access to electricity: Yes-No	Related SDG Indicator
1	Completion of secondary education	DHS/MICS	Household member aged 20-35	Wealth	Residence	n/a	Woman/Man	n/a	n/a	n/a	n/a	4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex
2	Completion of higher education	DHS/MICS	Household member aged 25-35	Wealth	Residence	n/a	Woman/Man	n/a	n/a	n/a	n/a	4.3.1 Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex
3	Stunting in children under 5 years of age	DHS/MICS	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children under 5 years of age	n/a	n/a	n/a	2.2.1 Prevalence of stunting (height for age <-2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age
4	Overweight in children under 5 years of age	DHS/MICS	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children under 5 years of age	n/a	n/a	n/a	2.2.2 Prevalence of malnutrition (weight for height >+2 or <-2 standard deviation from the median of the WHO Child Growth Standards) among children under 5 years of age, by type (wasting and overweight)
5	Wasting in children under 5 years of age	DHS/MICS	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children under 5 years of age	n/a	n/a	n/a	2.2.2 Prevalence of malnutrition (weight for height >+2 or <-2 standard deviation from the median of the WHO Child Growth Standards) among children under 5 years of age, by type (wasting and overweight)
6	Demand for family planning satisfied with modern methods	DHS/MICS	Women between 15-49 currently in union	Wealth	Residence	Respondent's education	Only Woman	Number of children under 5 years of age	Age group	n/a	n/a	3.7.1 Proportion of women aged 15-49 years who have their need for family planning satisfied with modern methods
7	Skilled birth attendance during childbirth	DHS/MICS	Women between 15-49 ever given birth in the last 5 years	Wealth	Residence	Respondent's education	Only Woman	Number of children under 5 years of age	Age group	Marital status	n/a	3.1.2 Proportion of births attended by skilled health personnel
8	Access to basic drinking water	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	n/a	6.1.1 Proportion of population using safely managed drinking water services

Annex A – UNESCAP LNOB Indicators, part 2

	Opportunities/Barriers (response variable)			Circumstances used to determine the furthest behind/ best-off groups (independent variables)								Closest SDG indicator reference
	Indicator	Survey used	Reference population in survey	Wealth: Bottom 40- Top 60	Residence: Urban - Rural	Education: No/Primary - Secondary - Higher	Sex: Male- Female	Children: Yes-No, Number	Age: 15-24, 25-34, 35-49	Marital status (Single, currently/ formerly married or in a union)	Household access to electricity: Yes-No	Related SDG Indicator
9	Access to basic sanitation services	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	n/a	6.2.1 Proportion of population using safely managed sanitation services, including a hand-washing facility with soap and water
10	Access to electricity	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	n/a	7.1.1 Proportion of population with access to electricity
11	Access to clean fuels	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	n/a	7.1.2 Proportion of population with primary reliance on clean fuels and technology
12	Ownership of bank account	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	n/a	8.10.2 Proportion of adults (15 years and older) with an account at a bank or other financial institution or with a mobile money-service provider
13	Internet use	DHS/MICS	All households	Wealth	Residence	Highest Education in household	n/a	n/a	n/a	n/a	Yes/No	17.8.1 Proportion of individuals using the internet
14	Sexual or physical violence against women	DHS/MICS	Ever married women	Wealth	Residence	Respondent's education	Only Woman	Number of children under 5 years of age	Age group	n/a	n/a	5.2.1 Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age
15	Early childhood education	DHS/MICS	Child aged 3-5 years old	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children under 5 years of age	n/a	n/a	n/a	4.2.2: Participation rate in organized learning (one year before the official primary entry age), by sex
16	Ability to protect from COVID-19	DHS/MICS	Household member	Wealth	Residence	Highest Education in household	Woman/Man	n/a	Age group: 0-24, 25-59, 60+	n/a	n/a	This index assesses the percentage of individuals living in a household with: 1) access to the internet, TV, phone, mobile phone or radio; 2) water pipes into the dwelling or yard or other private water source; 3) a handwashing facility on premises with soap and water available; 4) two or fewer individuals per sleeping room; and 5) a toilet which is not shared with other households