Original Research Paper

Volume - 12 | Issue - 03 | March - 2022 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Computer Science

META-ANALYSIS OF USAGE OF MACHINE LEARNING TO IDENTIFY AND ASSESS POVERTY

Anish Roy

ABSTRACT Poverty is defined as a person's state of survival without the adequate financial resources for a minimum standard of living. Poverty is assesed through multidimensional factors like access to clean drinking water, access to electricity, access to quality education, child mortality rates, nutrition, and so on. On 1st January 2016, the United Nations Economic and Social Affairs department published a list of 17 Sustainable Development Goals (SDGs), and "No Poverty" was the first goal on the list and the purpose was to eradicate all forms of poverty from all corners of the globe by the year 2030. The first step in the goal is to identify poverty in all its forms and not just the income level. The initiative of identifying poverty is a humongous task in itself and various researchers, academicians, statisticians and computer scientists had proposed several methods of identifying poverty in all it's forms. This paper unifies and consolidates several machine learning techniques proposed previously to theoretically formulate a new, robust methodology to identify, validate and assess poverty which would be the first step towards sustaibale development.

KEYWORDS : machine learning, poverty indicators, sustainable development, theoretical framework

1. INTRODUCTION

1.1 United Nations Sustainable Development Goals (SDGs)

The United Nations adopted an initiative which was "The 2030 Agenda for Sustainable Development" a list of 17 Sustainable Development Goals (SDGs) in September 2015, which was officially ratified on the 1st of January, 2016. These 17 SDGs all form a socio-economic ecosystem where each goal has an effect on the other. These goals were created to ensure that by the year 2030, the whole planet would live in peace and prosperity through stable social, economic and environmental sustainability.

(SDGs, n.d.)

1.2 No Poverty - The First Goal

The first goal among the global SDGs is eradicating poverty in all its manifestations by the year 2030. The member states of the United Nations and other organisations strived towards achieving the first goal. Although the global extreme poverty rate declined, it slowed down a bit in 2019. In 2020, things took a turn for the worse when the world was hit by the coronavirus pandemic which only exacerbated the issue further. About 119-124 million people were pushed into extreme poverty in 2020, seeing a spike in the number of people living in extreme-poverty in a generation. Poverty is inextricably linked with the rising numbers of unemployment, homelessness, starvation and so on. This is a global issue, and eradicating poverty is one of the most collosal challenges ever faced by mankind.

1.3 What Is Poverty?

(Alcock, 1997) As encapsulated beautifully in this paper
Poverty means going short materially, socially and emotionally.

Poverty is a complex phenomena which goes beyond the income of a person. Poverty has social, environmental and educational aspects. As explained later, a multidimensional poverty index was created, which gives a concrete estimate of the poverty of a country through multiple perspectives. Poverty eradication is a humongous task which can be undertaken in several domains. In this paper we use the basic principles of machine learning for this purpose.

1.4 Multidimensional Poverty Index

Poverty goes beyond the income level of a person. Poverty has 10 indicators spread across 3 dimensions - health, education and living standards. This robust methodoly was developed by the Oxford Poverty and Human Development Initiative and the United Nations to give a firm estimate of people living in poverty and to show the depth of interconnection among the three dimensions of deprivation.

Dimensions of Poverty	Indicator	Deprived if living in the household where...	Weight
Health	Nutrition	An adult under 70 years of age or a child is undernourished.	1/6
	Child mortality	Any child under the age of 18 years has died in the five years preceding the survey.	1/6
Education	Years of schooling	No household member aged 10 years or older has completed six years of schooling.	1/6
	School attendance	Any school-aged child is not attending school up to the age at which he/she would complete class 8.	1/6
Standard of living	Cooking Fuel	The household cooks with dung, wood, charcoal or coal.	1/18
	Sanitation	The household's sanitation facility is not improved (according to SDG guidelines) or it is improved but shared with other households.	1/18
	Drinking Water	The household does not have access to improved drinking water (according to SDG guidelines) or safe drinking water is at least a 30-minute walk from home, round trip.	1/18
	Electricity	The household has no electricity.	1/18
	Housing	Housing materials for at least one of roof, walls and floor are inadequate: the floor is of natural materials and/or the roof and/or walls are of natural or rudimentary materials.	1/18
	Assets	The household does not own more than one of these assets: radio, TV, telephone, computer, animal cart, bicycle, motorbike or refrigerator, and does not own a car or truck.	1/18

(Human Poverty Index, n.d.)

1.5 Identifying Poverty - The First Step

Identification of poverty is not an easy job. As explained earlier in the paper, poverty has multiple dimensions and it goes beyond the income of a person. For eradicating poverty, identification of poverty is the first step. Human surveying and judgement is time consuming and inefficient. Human surveying to measure the income level of a person and comparing it with the poverty threshold to identify poverty may not be as fulfilling, adequate or concise. Poverty is not one object. It is a complex phenomena, and poverty needs to be identified from the grassroot and it needs to be viewed from multiple perspectives, before we can even think about eradicating poverty. Most countries do not have the proper methodology to identify poverty from different perspectives and they stick to the basic questions. This paper brings a fresh perspective of viewing and identifying poverty from multiple dimensions.

2. Literature Review

2.1 (Hu et al., 2022)

This paper provides a unique perspective of assessing poverty in a Chinese village by implementing machine learning algorithms and using high resolution images and geospatial data. The model was able to detect village-level poverty with an accuracy of 54%. First some explanatory variables were generated using geospatial data. This study implemented machine learning models and used the random forest algorithm to identify and assess 3 factors affecting poverty - accessibility to healthcare and education, agricultural production conditions and socio-economic conditions. The study consists of 3 main parts - (i)interprating HRI based on eCognition Developer, (ii)calculation of time cost to nearest facilities combining HRI, POI and OSM, (iii)measuring dispersity of village settlements using Thiessen polygon.

2.1.1 Accesibility to Facilities and Services
Education and healthcare are two impostant indicators of poverty. Combining road networks from OSM, location of schools and healthcare facilities from POIs and land use patterns of the region, time costs to nearest schools and hospitals were measured to assess the ease of access to these basic resources.

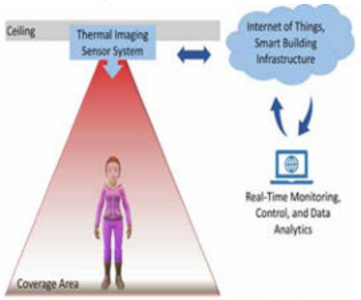
2.1.2 Agricultural Production Conditions
Agricultural Production is a key component of rural economy. The study selected a proportion of cropland as an indicator of agricultural production. Then the average elevation and the proportion of land with slopes exceeding 25 degrees were used to measure the terrain conditions. The lands with steep slopes and high elevation were unsuitable for agriculture and had negative impact on agricultural yeild.Also, highly forested areas implied isolation and less agricultural production, which is corelated to poverty.

2.1.3 Socio-Economic Conditions
The geospatial arrangements of village settlements indicate the access to natural resources. The study used the mean area of villages and the coefficient of variation of the Thiessen's polygons to measure the dispersity of villages. The densely dispersed settlements would naturally have more access to resources and would achieve high levels of socio-economic conditions.

2.2 (Jean et al., 2016)
This paper demonstrates a scalable and accurate way to identify poverty using machine learning. The region of study includes 5 African countries - Nigeria, Tanzania, Uganda, Malawi and Rwanda. This paper uses a convolutional neural network to identify upto 75% of local economic levels. Basically, this paper uses machine learning data and high-resolution sattelite imagery to compare daytime and nightime lightings across countries. Countries with poorer lighting indicates less access to electricity which is an important indicator of poverty.

2.3 (Kshirsagar et al., 2017)
This is a unique paper which offers a fresh perspective into poverty - through the poverty probability index, which is a methodology to construct a proxy-means test (PMT). This paper suggests a machine learning model to assess household poverty. Instead of the methodology of step by step regression and human judgement for variable selection, this model uses a coventional mode employing cross-validation and parameter regularization techniques to model both selection and estimation. According to this paper a PMT can be implemented using existing national surveys. It suggests that a subset of variables from these surveys be utilised to predict poverty at household level using machine learning models, and translate the fitted models into scorecards and validate the scorecard using field testing.

2.4 (Chidurala & Li, 2021)
This paper provides in-depth techniques and knowledge about occupancy estimation using thermal imaging and machine learning algorithms. The study uses low-resolution thermal imaging sensors for real-time non-intrusive occupancy estimation. The paper discusses the uses of various thermal imaging sensors and focuses on sensor classification and also discusses about different machine learning models. It presents a strong unified processing algorithm pipeline for occupancy estimation. Different algorithms are proposed for pre-preprocessing of sensor data, feature extraction and fine-tuning of occupancy estimation algorithms. The results show that occupancy estimation can be done with about 99% occupancy. This study can be used to estimate dispersion of settlements around resources and real-time settlement monitoring.



Deployment structure of the occupancy estimation system.

(Occupancy Estimation Algorithm, n.d.)

3. Theoretical Framework of Proposed Methodology
Based on previously done work, a coalesced, consolidated and robust methodology can be concocted, which implements all the above studies, with new ideas supporting the underlying skeletal framework. The individual studies referred to above do not have the required accuracy or depth for sequential identification, validation, classification and communication of poverty in all its forms. The methodology proposed in this paper would be a 4-fold plan of action with each step covering identification, validation, classification and communication. Machine learning techniques would be used in each of these steps to make the overall process smmother, fast, data-driven, robust and accurate.

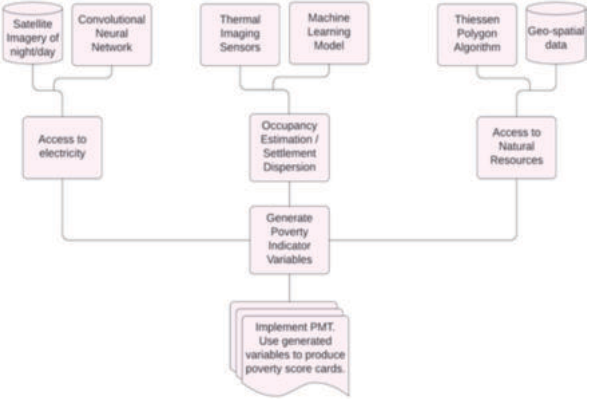
3.1 Step 1 - Identification
The 4-fold methodology suggested in this paper starts with identification. The identification process consists of 3 consists of 3 components based on previous research methodologies, to identify and assess the scale of poverty in a region.

This step is the unified efforts of three components steps which are as follows:

3.1.1 Identifying Poverty Indicators from the 3 Dimensions of Poverty
As stated above poverty has 3 dimensions and 10 indicators under them. Measuring just one indicator would neither be adequate, nor accurate. The first step of identification is to measure the indicators of poverty like access to electricity, natural resources like water, healthcare, education, etc. This methodology implements the different methodologies suggested by previous researchers as components to provide an even stronger methodology. First the satellite imagery and convolutional neural networks can be used to study the day/night images of different regions to identify areas with poor lighting, which implies lower access to electricity. Then thermal imaging sensors can be used along with machine learning algorithms to estimate the population density. Finally, High-resolution geospatial data can be used along with machine learning algorithms to calculate time costs to healthcare facilities, educational institutions, etc.

3.1.2 Generating Poverty Indicator Variables
From the above step, we get a clear idea about the demography and geography of an area, and we can use them to generate variables like percentage of population living near natural resources, number of healthcare facility units or educational institutions available in an appropriately selected square units of area, and so on.

3.1.3 Creating Poverty Scorecards
The above generated data can be used to implement PMTs and generate poverty scorecards based on these variables. The scorecards can be created on a global, regional, national, or local scale.

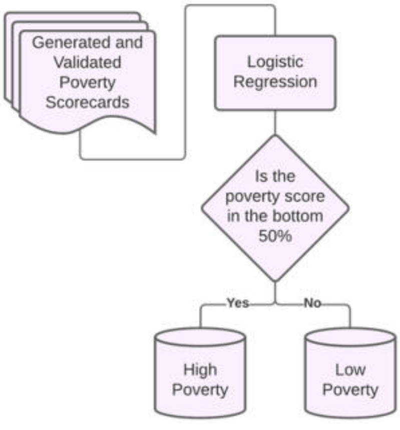


Step 1- Identification (created by author)

3.2 Step 2 - Validation
In the previous steps the methodology suggests creating poverty scorecards to give the statistics of the poverty indicators. The second step would be to validate and verify the data. The data needs to be compared to the survey data acquired through human survey, and all bias in the data needs to be removed. Any wrong data should be corrected to remove hindrance.

3.3 Set 3 - Classification
In the previous steps the methodology was implemented to identify

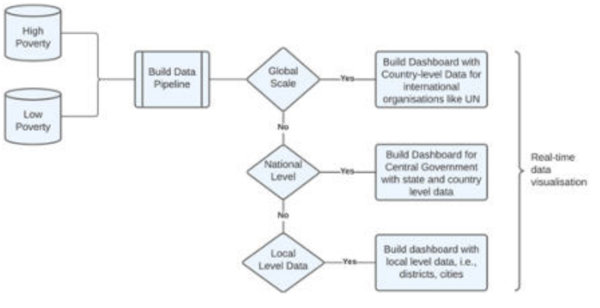
poverty and verify it. Now, the regions need to be classified on the basis of high/low poverty. The generated scores from the respective scorecards can be classifed through machine learning models implementing logistic regression. The classified scorecards with the scores and all the indicators should be stored in separate databases for further analysis and insights.



Step 3 - Classification (created by author)

3.4 Step 4 - Communication

The results of the previous steps, after being stored in a database, needs to be shared with the governmental bodies. Use of real-time, interactive dashboards connected to these data sources, to graphically convey the huge amounts of data is suggested. Real-time, interactive dashboards can be created based on the scale of the data. Scorecards of countries can be conveyed graphically to international organisations like the United Nations. On the other hand, state level/local level data can be shared through dashboards to central governmental bodies.



Step 4 - Communication (created by author)

4. CONCLUSIONS

Identification and assessment of poverty is a huge task, but it can be broken down into components for simplicity. Furthermore, using machine learning would make the task even easier, more accurate and fast. This paper formulated a firm and robust, data-driven methodology implementing numerous machine learning algorithms for identifying and assessing poverty through a multitude of indicators from all the three dimensions.

The proposed theoretical methodology should be highly accurate and adequate. It would give a clear idea about the poverty of different regions/countries/states depending upon the use. The interactive dashboard would give the governmental departments the clarity, to assess the aspect they are lagging in, to implement methodologies for further improvement. By applying this methodology, if a real-world pipeline is made, that would be a huge leap towards the first sustainable development goal - No Poverty. The sustainable development goals form an ecosystem, where one factor affects the other. So, by implementing this methodology, not only would poverty be identified and assesed, many other aspects of sustainable development would be affected as well.

5. Limitations and Future Scope of Work

5.1 Limitations

The whole framework proposed in the paper is purely theoretical. The success of this methodology would only be known after the practical application of the proposed method.

5.2 Future Work

With the current advancement in technology, it would not be long before a researcher, computer scientist, economist, or statistician would use the proposed methodology for identification and assessment of poverty across multiple dimensions. The machine learning models could be used with geospatial data available to identify poverty and classify them and visualise them for different governmental and/or international organisations.

6. Funding

This paper received no funding.

REFERENCES

1. https://i1.wp.com/www.un.org/sustainabledevelopment/wp-content/uploads/ 2015/ 12/ english SDG 17goals poster all languages_with_UN_emblem 1.png? fit= 728 % 2C451 &ssl=1. https://i1.wp.com/www.un.org/ sustaina bledevelopment/wp-content/ uploads/ 2015/ 12/english SDG 17goals poster all languages with UN emblem 1.png? fit= 728 % 2 C451&ssl=1
2. Understanding poverty. (1997). Macmillan International Higher Education.
3. https://communitymedicine4asses.files.wordpress.com/2019/12/multidimensional-poverty-index-mpi-_-human-development-reports.png?w=768. https://communitymedicine4asses.files.wordpress.com/2019/12/multidimensional-poverty-index-mpi-_-human-development-reports.png?w=768
4. Village-level poverty identification using machine learning, high-resolution images, and geospatial data. (2022). International Journal of Applied Earth Observation and Geoinformation, 107, 102694.
5. Combining satellite imagery and machine learning to predict poverty. (2016). Science, 353(6301), 790–794.
6. Household poverty classification in data-scarce environments: a machine learning approach. (2017). ArXiv Preprint ArXiv:1711.06813.
7. Occupancy estimation using thermal imaging sensors and machine learning algorithms. (2021). IEEE Sensors Journal, 21(6), 8627–8638.
8. https://ieeexplore.ieee.org/ielx7/7361/9357819/9314180/graphical_abstract/jsen-gagraphic-3049311.jpg. https://ieeexplore.ieee.org/ ielx7/7361/ 9357819/ 9314180/ graphical abstract/jsen-gagraphic-3049311.jpg