ISSN 2311-3782

NIGERIA POVERTY ANALYSIS, PREDICTION USING MACHINE LEARNING METHODS AND DEEP LEARNING

Danbirni Mustapha Ibrahim
Zhejiang University of Science and Technology School of Science
No 318 Liuhe Road, Hangzhou,310023, **CHINA**
Email: mustaphaibrahimdanbirni@yahoo.com

Dongxiao Ren
Zhejiang University of Science and Technology School of Science
No 318 Liuhe Road, Hangzhou,310023, **CHINA**
Email: Rendx29@163.com

ABSTRACT

Poverty refers to a lot of lacking which also include lacking enough resources to provide the necessities of life such as food, clean water, shelter and clothing which in today world it, include access to health care, education and even transportation. A lot method was given to Nigeria government which fails to work the way it supposed to work. The predications are not accurate, and the traditional way of prediction in Nigeria include site survey which is expensive and labor intensive which tend to be waste of time before getting the real result. Lack of reliable poverty data in Nigeria is the major obstacle for making informed policy decisions and allocating resources effectively in those areas that need help the most. In this paper we will first of all carry out multidimensional poverty analysis using multiple correspondence analysis to fully understand the cause of poverty then well make predictions using three different machine learning method then finally based on previous research that is satellite images processed through convolutional neural networks, we will also use that to gauge poverty levels. This paper attempts to extend on past work by comparing the simple machine learning methods to the complex deep learning method to see which is more suitable to best understand and predict poverty in Nigeria.

Keywords: Poverty Prediction, Machine Learning, Convolutional Neural Network, Decision Tree, Binary Logistic Regression, Random Forest, Multiple Corespondent Analysis, K-Mean.

1. INTRODUCTION

Poverty is a big problems in Nigeria that has multidimensional problem that has a variety of definitions. For some authors, it is measured by income while other researchers include also health, education, social status and political rights into the picture. However, what connects all researchers is their work in the field of identification of factors that cause poverty, classification of population according to different views of poverty and prediction of future poverty levels [1]. Our motivation for this paper is to use different statistical, machine learning methods to address the problem of poverty in Nigeria using household survey dataset, then create three simple poverty prediction models and then also use satellite images of two distinct states; Lagos and Jigawa to create a complex convolutional neural network CNN model and finally compare them to see which best describes poverty in Nigeria. The machine learning methods we will be using to create our models are logistic regression, decision tree and random forest. Our paper is divided in to sections. In the first section we will talk about multidimensional poverty analysis. The second section describes our methodology, it talks about the methods we used. The third section describes our results and findings and then finally the fourth section we make conclusions.

2. LTERATURE REVIEW

In this chapter we will briefly talk about the literature review of the study. First, we will discuss about poverty measurement and multidimensional poverty analysis. Secondly, we will talk about the theoretical aspect of the statistical techniques used in performing poverty analysis and then finally we will discuss the theoretical aspects of our machine learning and deep learning methods we used for this work. This section gives more insight on what various methods and algorithms we use are.

2.1 Poverty Measurement and Analysis

A poverty line is calculated in order to determine the well-being of households. Those whose expenditure (or income) falls below the line can be categorized as poor. There are three methods to construct that measurement: the cost of basic needs, food energy intake, and subjective evaluations [2]. As the cost of living across the world varies, the World Bank set the international poverty line $1.90 per person per day as a global threshold. Those who have the average spending per capita per day below the threshold are classified as poor [2]. According to Oxford Poverty & Human Development Initiate (OPHI), Poverty should not be defined as just lack of income as conventionally done by most countries, focusing alone on one factor is not enough as people's experience with poverty is broad, multidimensional poverty analysis/measure can be used to create a more comprehensive picture. The analysis reveals who is poor and how they are poor [3]. According to UNDP the Multidimensional Poverty Index (MPI) identifies multiple deprivations at the household and individual level in health, education and standard of living. It uses micro data from household surveys, and—unlike the Inequality-adjusted Human Development Index—all the indicators needed to construct the measure must come from the same survey. Each person in a given household is classified as poor or non-poor depending on the weighted number of deprivations his or her household, and thus, he or she experiences. These data are then aggregated into the national measure of poverty. The MPI reflects both the incidence of multidimensional deprivation (a headcount of those in multidimensional poverty) and its intensity (the average deprivation score experienced by poor people). It can be used to create a comprehensive picture of people living in poverty, and permits comparisons both across countries, regions and the world and within countries by ethnic group, urban or rural location, as well as other key household and community characteristics. The MPI offers a valuable complement to income-based poverty measures [4].

2.2 Statistical Methods For Poverty Analysis

This section gives an overview of the various statistical analysis techniques used for the research.

2.2.1 Multiple correspondence analysis

Multiple correspondence analysis (MCA) is a proper technique for the construction of poverty indices [1]. The technique comes from the family of 'factorial' techniques which aim to decrease the dimension of the dataset and understand the data patterns. The idea is the same as Principal Component Analysis (PCA). While PCA works well with numerical data, MCA is a proper technique to use for categorical and ordinal data [1].

2.2.2 K-means clustering

The method divides and partitions the data into K clusters which are not overlapping. The criterion for good clustering output is when the within-cluster variation is as small as possible.

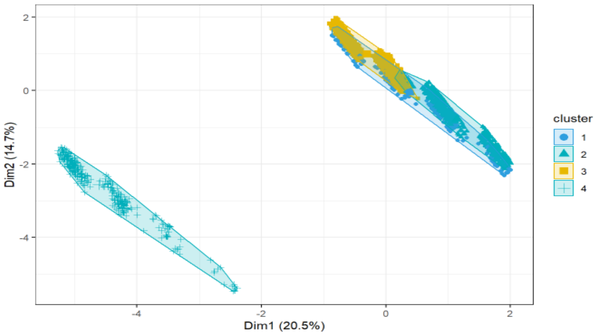

Figure 1: cluster plot

As we can notice on the above plot for our data set, there is a large decrease of variance already by choosing 3 clusters but the function suggests 4 therefore we have chosen 4 groups.

2.3 Machine Learning And Deep Learning Methods

Basically, machine learning is a computer algorithm that learns the data without having to be programmed explicitly. Simply put, given a model, or structure, that are determined by some parameters, the algorithm will optimize those parameters using the training data. Then, we can use that model to predict new data [5].

Machine Learning

Machine learning provides effective solutions for understanding data patterns and predictions. The main goal of prediction is to accurately determine the outcome based on the causality relationships in the data [6]. In our case, we wanted to use the household data from Nigeria and satellite images to determine and predict poverty levels.
There are two main groups of machine learning algorithms that are frequently used to provide insights on similar problems:
-unsupervised learning algorithms
- supervised learning algorithms

Unsupervised Learning

The unsupervised machine learning uses all available data to understand the patterns and relationships in the data [7]. A typical example is clustering the data into objects that have some common features or compressing the data set into a lower dimension while still keeping as much information as possible. The first part of our analysis was done using unsupervised learning, the multiple correspondence analysis and the k-means are all unsupervised algorithms.

Supervised Learning

Supervised learning algorithms are suitable for predictions. Generally, there are multiple choices for prediction models widely used by statisticians and data scientists [8]. In our case we choose four supervised machine learning methods to make our predictions. The description on how we used these algorithms and summary of their outputs is explained in the consequent part of this paper.

2.3.1 Logistic regression

Logistic regression is a convenient method to use when we have a qualitative dependent variable. The requirement for this method is that the dependent variable is coded as binary, i.e., reaches values of 0 when the characteristic is not present and 1 when the individual has the characteristic. The desired outcome for such scenario would be the probability with which the person has the characteristic. The logistic regression model is thus based on the logistic function:

$$P(X) = \frac{e^{B0+B1x1+\cdots Bnxn}}{1+^{e^{B0+B1x1+\cdots Bnxn}}}$$

Where p(X) is the probability that a person has the characteristic and Bi are the unknown parameters. The best estimate for the unknown parameters is such that will associate a low probability to respondents which do not possess the feature and vice versa will assign high probability to individuals who have the feature [7].

2.3.2 Decision tree

There are two types of trees that can be constructed - regression and classification trees. Classification trees are suitable to predict a qualitative outcome, as our dependent variable 'poverty'. Therefore, we will limit ourselves to defining these. The way how we grow the tree is called recursive binary splitting. First, we take the whole dataset which represents the root of the tree. Based on the first predictor we split the dataset into two regions. Typically, the decision is based on some threshold criteria and depending on the value of the feature we either move to the right or to the left node. We consider binary predictors as can be found in our dataset. Therefore, we move to the right when the answer is 1 or to the left when it is 0. In the same manner we continue until we have built the tree [9].

2.3.3 Random forest

As the name suggests, random forest is a technique that analyzes many decision trees. The disadvantage of the single decision tree is its high variance and dependency on the training dataset. This can be avoided with the increased number of trees that are considered. In a fantasy world we could draw many samples, grow decision trees for each sample and base our conclusions on some averaged value obtained from the individual trees. In reality, this is not feasible since we typically do not have resources for drawing many samples. The technique called 'bagging' simulates this scenario using only our original sample. We provide a brief overview derived from [9].

2.3.4 Convolutional nueral network

In machine learning, artificial neural network is a model that inspired by biological neural networks in the brain. It has two main components that follow the idea of how the brain works. The first is a neuron (or node) that is like a biological neuron, they are stimulated by inputs. These neurons will pass information they receive to other neurons, often with transformations.

The second component is the signal. The artificial neurons will be trained to pass forward the useful signals to achieve the larger goals of the brain [10].

3. Related Work

There are a lot of research on predicting poverty using machine learning but there's almost none on Nigeria. Asselin and Ahn wrote a paper on multidimensional poverty and multiple correspondence analysis [1]. Ezzari and Verme wrote a paper titled multiple correspondence analysis approach to the measurement of multidimensional Poverty in Morocco [11]. In 2016 Pluliková wrote a thesis on poverty analysis using machine learning methods [12]. Following the explosion of deep learning, Oquab et al. apply transfer learning with CNNs to classify images in the PASCAL VOC by reusing layers trained on ImageNet [13]. Neal Jean uploaded his research on Github on how to combine satellite images and machine learning to predict poverty [14].

4. METHODOLOGY

In this section we will describe the data used for the work, how we obtained and pre-processed it, and then we will present an overview of the approach used to achieve the goal of this work.

4.1 Data

Nigeria Living Standards Survey (2018)

The data used for training and scoring the machine learning models was obtained from the world bank Living Standard Measurement Study (LSMS), the survey was carried out based on various aspects of Nigerian standard of living in 2018-2019, all surveys were conducted by the world bank in collaboration with the national bureau of statistics (NBS). The survey contains information on households and communities the household surveys were conducted on various household where questions were asked about the household and its members; and it covers sectors like health, education, assets, housing, labor, income and so on. The survey was carried out for 116,321 household but lot of answers were missing. After cleaning we were left with information for 8,229 households. The description of the data is presented in the table 1 below.

Variable name	Values	Frequency	Description
School	Yes NO	6991 1238	Has the person attended any school
Randweng	Yes NO	4570 3659	Can the person read and write in English
Randwothr	Yes NO	3214 5015	Can the person read and write in any language
Healthy	Yes NO	806 7423	Did the person went to see a doctor recently
Employed	Yes NO	564 7665	Is the person employed

Assets	Yes NO	2381 5848	Does the household have any form of assets
Savings	Yes NO	564 7665	Does the household have any form of savings
Tapwater	Yes NO	1916 6313	Does the household have tap water from any source
Owntoilet	Yes NO	3089 5140	Does the household own flushable toilet
Electricity	Yes NO	5318 2911	Have electricity

Table 1. Data Description

Daytime Satellite Image

The primary data used in training our neural networks was obtained from google maps, we collected imagery data of states in Nigeria. Daytime satellite imagery is extracted from Google Static Maps API. By providing an API key, we are able to generate the high-resolution images given the geolocation information and the zoom level. The geolocation consists of latitude and longitude value indicating a location of a place in the real world. Meanwhile, the zoom level of Google Maps ranges from 0 to 19 describing the map scale. Google Maps is built on a 256X256 pixel tile system where zoom level 0 is a 256X256 pixel image of the whole earth. A 256 256 tile for zoom level 1 enlarges a 128X128 pixel region from zoom level 0 [15].

4.2 Algorithms and Methods

Here we will describe all the methods and algorithm used in this research and how we used them to carry out the research.

4.2.1 Multiple Correspondence Analysis

The first step of our research was to use multiple correspondence analysis to carry out multi dimensional poverty analysis on our living standard dataset. We performed MCA on the variables from in the dataset that were all coded as factor variables with two levels. The MCA was performed on the dataset of 8,229 respondents (households). The results show that dimension 1 explains 26.6% of total inertia. We visualize the output in figure 1. The chart represents the scatterplot where on x-axis we show Dimension 1 and on y-axis we show Dimension 2. The values for these dimensions represent the distances from the columns i.e., the principal axes. After visualizing and inspecting the plot we can see that there is a pattern in the data the categories associated with non-poverty are clustered in the left-hand side of the plot. We used MCA to determine if the household was poor or not by analyzing the various indicators of each household and thus implementing multidimensional poverty analysis. We used the package FactoMineR and function MCA.

The syntax for the function is the following:

MCA(X, ncp = 5, ind.sup = NULL, quanti.sup = NULL, quali.sup = NULL, *graph = TRUE, level.ventil = 0, axes = c(1,2), row.w = NULL, method = " Indicator", na.method = "NA", tab.disj = NULL)*

The output of the MCA is illustrated below.

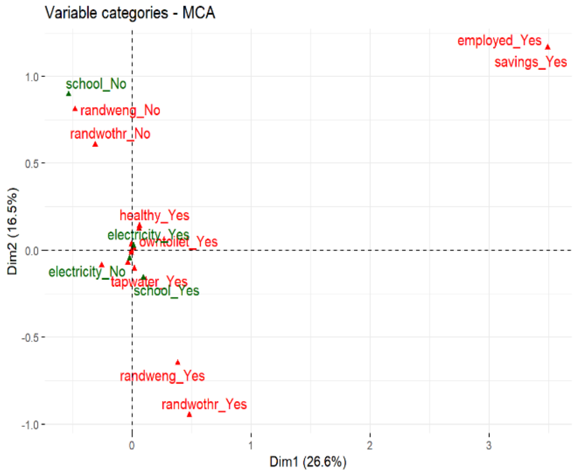

Figure 2: MCA

The important part of the output includes the factor loadings for each category and the total inertia captured by our model. The full output from the MCA analysis is not included in the paper. To compute the poverty index, we used the factor loadings in the following equation:

$$PI = \sum_i (XijWj),$$

where PI stands for Poverty Indicator and Wj stands for the weight of the jth variable obtained from MCA. The weight is actually a factor loading from Dimension 1.
The MCA output represents the weights that we give to each category. We have seen on the variable map that the negative scores are associated with the poor and the positive scores with the non-poor.

4.2.2 K-Means Algorithm

The second step of our analysis is to determine the poverty levels. K-means clustering is a suitable solution for this problem. One problem with this algorithm is that we need to determine the number of clusters. This is not something that we are always aware of right from the beginning. One way to resolve this problem is to make a scree plot with different choices of K and their respective within cluster variance [7]. The plot below shows different choices of number of K.

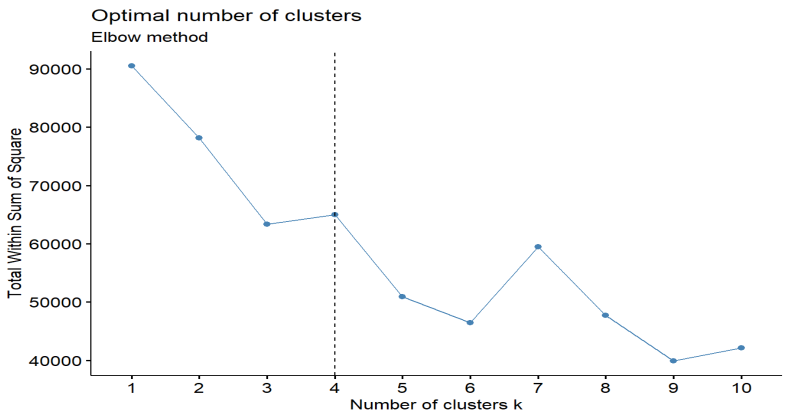

Figure 3: Plot for number of K

for our dataset, there is a large decrease of variance already by choosing 3 clusters but the function suggests 4 therefore we have chosen 4 as our number of K. We used the K-Means Clustering algorithm in R. The function syntax is the following:

kmeans(x, centers, iter.max = 10, nstart = 1, algorithm = c("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen"), trace=FALSE)

4.2.3 Machine Learning Models

After creating classes out of our dataset using the K-means algorithm our next step was to create a target column and bind every household to its target that is if its poor or not. Afterwards we split the dataset into training and testing 70% of the dataset was used for training while 30% was used for testing. Three models were created based on logistic regression, decision tree and random forest methods all the models will be discussed individually below.

Logistic Regression

We fit the logistic regression method to our model with the following R syntax:

multinom(formula, family = gaussian, data, weights, subset, na.action,
start = NULL, etastart, mustart, offset, control = list(...), model =
TRUE, method = "multinom.fit", x = FALSE, y = TRUE, contrasts = NULL, ...)

After training the model then we used the test data to make predictions which was used to test and score the performance of the model the output or result of the models shows that it took 15.11 seconds to make the predictions and it has a 99% specificity and also 99% percent sensitivity and overall accuracy 98% which show the model performed very well in making predictions.

Decision Tree

Next, we fit our second model using the decision tree method, fitting the decision tree method in R is supported by many packages but we chose to use the rpart package and the syntax used is:

rpart(formula, data, weights, subset, na.action = na.rpart, method,
model = FALSE, x = FALSE, y = TRUE, parms, control, cost, ...)

After fitting we also made predictions using the training data and the output of the model showed its too 1.09 seconds to make the predictions, it has 69% sensitivity and 94% specificity with overall accuracy of 86%, looking at the other model we can see that our model didn't perform as well as the first one so we tried to improve the model by tweaking some hyperparameters using the control function and we afterwards got a better model with accuracy of 95%.

Random Forest

We used the R package RandomForest to grow the trees for our model, the syntax used is:

randomForest(x, y=NULL, xtest=NULL, ytest=NULL, ntree=500, mtry=if (!is.null(y) && !is.factor(y)) max(floor(ncol(x)/3), 1) else floor(sqrt(ncol(x))), replace=TRUE, classwt=NULL, nodesize = if (!is.null(y) && !is.factor(y)) 5 else 1, importance=FALSE, ...)

The output of our random forest-based model shows it has a sensitivity of 99% and specificity of 99% it also has an overall accuracy of 98% while it took 33.34 seconds to make its predictions. While the slowest of the three it made a very good model and it's the best of the three models.

4.2.4 Comparison Of The Three Models

The random forest-based model performed similarly to the logistic regression model while the decision tree model which has the least accuracy performed low until some of its hyperparameters were tweaked to get a batter performance but even after that its still the least performed of the three we summarized the three models with the figures below.

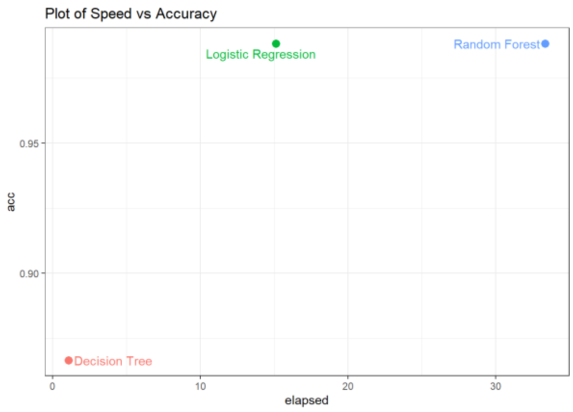

Figure 4: Speed versus Accuracy

Convolutinal Neural Networks Cnn

This satellite image are images of Earth that are collected by imaginable satellites which is operated by governments and difference businesses around the world. We will make used of this daytime image in predicting and detecting the area that are in poverty in Nigeria. We downloaded 5000 images data from Google earth where we select two state in Nigeria. We access along with 300 square tiles of the daytime satellite image of jigawa within 5km radius, and almost 70 square tiles of Lagos within the 3 km radius. Every single cluster centroid was been surrounded by tile that corresponds to pixel in visible imaging radiometer suite dataset. Each cluster location defined by latitude and also longitude of the households which is labeled by a point to highlight the area differences. we selected this because one area is very poor and the other place is rich, we run and train 500 images from the poor region and 500 from the rich places. Given the corresponding daytime satellite imagery, we want to accurately predict daytime light. Static Map API provides images on a purpose; we used 15, 10, level of zoom and have pixel level of resolution of about 1 meter. In this zoom level, 400 of 400-pixel images compress with the 0.5km solution for daylights data. An example of daylight satellite image as shown in Fig. 4 and 5 below.

Figure 5: Example of Jigawa Satellite Image

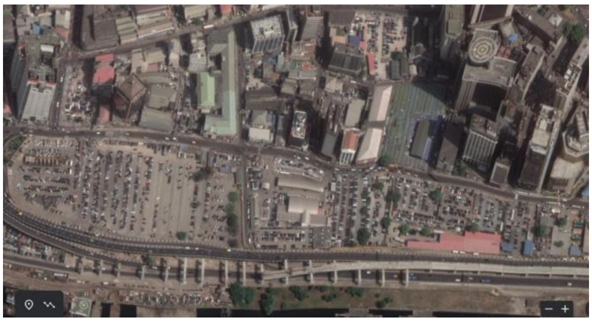

Figure 6: Example of Lagos satellite image

The output of the neural network daytime satellite image poverty prediction model shows it has sensitivity of 99% and its 100% percent accurate which is one of the advantages of CNN it has the maximum accuracy. On the other hand, its drawback lies in the interpretability. It is almost impossible to connect the input variable to the outcome. We cannot say that by changing values of one predictor we will increase or decrease the output variable by a certain value. This might be problematic to a problem of poverty prediction where this connection might be interesting to the development agencies or governments. The second problem is connected to random initialization of the model. Since we need to run it multiple times, this can be computationally expensive, especially with large datasets.

5. DISCUSSION

The aim of the paper was to analyze poverty in Nigeria using Nigeria living standard survey data and then predict poverty with the same data using simple machine learning technique, then use daytime satellite image to predict poverty using a complex deep learning technique and then compare both and see which helps and which is best for understanding poverty in Nigeria. After the experiments and after taking a good look at both results, we can see that they both have their advantages and disadvantages, more so, they can both be used simultaneously to understand and predict poverty, we can even go ahead and create a model that uses both data types to predict poverty. When we look at the machine learning methods, we can see that they are simple classification methods and they worked well with our dataset and thus the neural network worked well with our satellite images. Our research show how unsupervised learning algorithms can be used on the survey data to make poverty analysis and then we can used supervised learning algorithms to predict poverty, our research will help government and other parties to understand household poverty as well as predicting poverty by looking at satellite imagery. Finally, our research also suggests that using both techniques could a strong tool to help in understanding and predicting poverty.

6. CONCLUSION

To achieve the aim of the research, a lot of steps were carried out which include data sourcing and preparation, even though we had issues with the data obtained but we made it suitable for research and used different functions to make it suit all the methods and algorithms used for the research. In the light of this we used Multiple Correspondence analysis to work with categorical variables and created a poverty index that is a numerical representation of information extracted from our categorical descriptor variables. We chose categorical variables that captured information about household conditions, education and health of the household. The choice of these descriptors was justified by the theoretical framework that represents our view of poverty as a multidimensional concept. Consequently, we have applied the K-Means algorithm on the newly created numeric index of poverty to cluster our respondents into four groups. With this new variable we could proceed with predictive modeling. Models we used belong to the group of supervised learning algorithms. As the name suggests, these algorithms require that both the predictors and the outcome are included in the model. The learning then takes place which represents the process of tuning the parameters of an algorithm.

To conclude, we claim that Neural networks though requires very high computational power is the algorithm with the greatest predictive ability. This was also supported when we fed the algorithms satellite imagery. We also claimed that used both techniques simultaneously could create a great tool to help reduce poverty. However, with the results we obtained, we cannot claim that we have created a proper model for poverty prediction. Some revision has to take place to introduce more or better predictors. We recommend that further research takes place in this direction.

REFRENCES

[1] Asselin, L.-M. And Anh, V.T. (2008): Multidimensional Poverty Measurement with Multiple Correspondence Analysis. In: Kakwani, N. And Silber, J. (eds.) Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave Macmillan.
[2] Introduction to poverty analysis (English). World Bank Group, 2014.
[3] https://ophi.org.uk.
[4] https://feature.undp.org/multidimensional-poverty/.
[5] E. Alpaydin, Introduction to machine learning. MIT press, 2014.
[6] Shalev-Schwartz, S. and Ben-David, S. (2014): Understanding Machine Learning :From Theory to Algorithms. Cambridge University Press.

[7] Ng, A. (2015) : Machine learning [course]. Lecture materials: https://www.coursera.org/learn/machine-learning. Stanford University: 2015.

[8] Breiman, L. (2001): Statistical Modeling: The Two Cultures. Statistical Science. Vol. 16, No. 3, 199-231.
[9] James, G; Witten, D., Hastie, T., Tibshirani, R. (2013): An Introduction to Statistical Learning with Applications in R. New York: Springer.
[10] J. Patterson and A. Gibson, Deep Learning: A Practitioner's Approach. " O'Reilly Media, Inc.", 2017
[11] Abdeljaouad Ezzrari and Paolo Verme A Multiple Correspondence Analysis Approach to the Measurement of Multidimensional Poverty in Morocco, 2001-20071.
[12] Pluliková poverty analysis using machine learning methods 2016.
[13] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1717–1724, 2014.
[14] https://github .com/nealjean/predicting-poverty.
[15] (2018). Google maps platform - maps javascript api. (accessed July 18, 2018), [Online]. Available:https://developers.google.com/maps/documentation/javascript/maptypes# World Coordinates.
[16] Beck, Marcus W (2016) : nnet_plot_update.r. Retrieved 22-04-2021 from https://gist.github.com/fawda123/7471137.
[17] Everitt, B.S. and Hothorn, T. : Handbook of Statistical Analysis Using R. Retrieved 27-04-2021 from https://cran.r-project.org/web/packages/HSAUR/vignettes/ Ch_logistic_ regression_glm.pdf.
[18] FactoMineR in R – documentation. Retrieved 22-04-2021 from https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf.
[19] Ggplot2 in R – documentation. Retrieved 10-02-2021 from http://docs.ggplot2.org/current/.
[20] K-Means Clustering in R – documentation. Retrieved 28-04-2021 from https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html.
[21] Rpart in R – documentation. Retrieved 25-04-2021 from https://cran.r-project.org/web/packages/rpart/rpart.pdf.
[22] M. Pinkovskiy and X. Sala-i-Martin, \Lights, camera. . . income! illuminating the national accounts-household surveys debate", The Quarterly Journal of Economics, vol. 131, no. 2, pp. 579{631, 2016.
[23] A. Head, M. Manguin, N. Tran, and J. E. Blumenstock, \Can human development be measured with satellite imagery?", in Proceedings of the Ninth International Conference on Information and Communication Technologies and Development, ACM, 2017.
[24] A. Perez, C. Yeh, G. Azzari, M. Burke, D. Lobell, and S. Ermon, \Poverty prediction with public landsat 7 satellite imagery and machine learning", arXiv preprint arXiv:1711.03654, 2017.
[25] R. Engstrom, J. S. Hersh, and D. Newhouse, \Poverty from space: Using high-resolution satellite imagery for estimating economic well-being", 2017.
[26] B. Klemens, A. Coppola, and M. Shron, \Estimating local poverty measures using satellite images: A pilot application to central america", 2015.
[27] A. Castrounis. (2016). Arti cial intelligence, deep learning, and neural networks, explained. (ac-cessed march 25, 2021), [Online]. Available: https://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neural-networks-explained.html.
[28] T. Keenan. (2017). Neural networks demystified. (accessed march 25, 2021), [Online]. Available: https://www.upwork.com/hiring/data/neural-networks-demystified/.
[29] D. Gupta. (2017). Fundamentals of deep learning { activation functions and when to use them? (accessed march 28, 2021), [Online]. Available:

https://www.analyticsvidhya.com/blog/2017/10/fundamentals-deep-learning-activation-functions-when-to-use-them/.
[30] K. Nogueira, O. A. Penatti, and J. A. dos Santos. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61:539–556, 2017.
[31] National bureau of statistics website.
[32] World bank living standard survey website.