Poverty Prediction Through Machine Learning

Huang Zixi
UWC Changshu China
Changshu, China
zxhuang19@uwcchina.org

***Abstract*—Poverty elimination stands as an inevitable process in human development, with predicting poverty being the first and one of the essential steps. The paper considers poverty as an outcome of multidimensional factors, and offers various practical models for such prediction using machine learning, none of which accounts for the whole, while some factors may outweigh others. Thereby, an integrated approach of prediction is needed by combining the data from Poverty Probability Index and Oxford Poverty & Human Development Initiative. Through applying linear regression model, decision tree, random forest model, gradian boosting model, and neural network to analysis existing data, the paper assesses respectively the extent to which the factors matter and the efficacy of each model. Final advancing employs cross validation and grid research. Through analysis and comparison, the paper concludes that generally, gradient boosting is the model with the highest accuracy for predicting poverty and education as the most influencing factor. The finale finishes upon the possible reason behind the factors.**

Keywords-Poverty; Prediction; Multidimensional; Machine Learning; Models

I. Introduction

A. Research Background

Listed as the first goal for sustainable development, elimination of poverty speaks its own importance in the course of human development. [1] According to World Bank, in 2015, 11% of the world population live in extreme poverty or less than $1.9 per day and struggling subsistence. [2]

Over the past decade, researchers kept seeking ways to both eliminate and predict poverty. Elimination of poverty, through the collaborative efforts of local government and international NGOs, has been largely improved, with the poverty rate fallen by 3 percentages since 2014. [3,4] However, Poverty prediction has yet been adequately studied. two questions are pondered: 1, Who is in poverty? 2, Who is likely to be in poverty? the answers enable accurate and ahead intervention. Thus, researches on poverty prediction are worth spending effort researching.

B. Literary Review

Looking through today's findings, existing models pave the way for new researches. Expertise France analyses a variety of objective and subjective single indicators and concludes that the objective indicators are better predictors and the indicators related to income, benefits, and debts show more correlation. [5,6] The finding also suggests that none of the single indicator can be responsible for risks of poverty, since all the accuracy is relatively low and many factors are, in fact, interdependent and interconnected. Therefore, the following studies should take an integration approach. However, the limitation for the finding is the number and dimension of the data chosen, as such social research should also incorporate other dimensions, such as health and education.

More recently, satellite images have become a popular instrument. A team led by Marshall Burke, an economist at Stanford University, predicts poverty in Africa with satellite images and machine learning. The model was highly efficient——without extra efforts for data —— and, more importantly, accurate, with the daytime image prediction reaching 90% accuracy. [7,8] Nonetheless, the method may satisfy regional but not specific household analysis. Thus, it is only good for part of the study. In 2018, a paper published by Sustainable Architecture and Construction Research Group provides a possible method. In the paper, the authors applied linear regression model and neural network to predict the risks of fuel poverty with a R^2 coefficient of 99.6%. Therefore, this method may also be applicable to the issue discussed in this paper.

II. Methodology

The main approach used is machine learning, which is a method that mainly focuses on the development of computer programs that access data and learn automatically. In particular, this paper uses python as programming language. We applied several models like linear regression, decision tree, neural network, and so on. The detailed process is described below.

A. Data collection

The data used in this paper is merged from two sources, Oxford Poverty & Human Development Initiative and Poverty Possibility Index. Poverty Possibility Index provides data from individuals, including information about education level, internet accessibility, and so on. Oxford Poverty Human Development Initiative presents multidimensional poverty index for different countries. It is calculated through the Function (1); H is the portion of people that is unable to access certain benefits (e.g. the people who cannot have access to clinics), and A is the average deprivations a person suffers. Therefore, we take the M as the value of indicators in our research.

$$M_0 = H \times A \tag{1}$$

B. Data observation

Before applying the models, we first make some pre-observation of the distribution of our data. The data, in total, contains 59 features for 12600 individuals. We specifically look through the distribution of poverty probability distribution, as well as the distribution of factors like age, education.

Figure 1 is the distribution of the poverty probability. Although a large amount of the data appears within the range of 0.8 to 1.0, it is wildly distributed throughout 0 to 1, showing that the prediction will have reliability for predicting people with different situations. However, the concentration demonstrates

978-1-6654-3873-5/21/$31.00 ©2021 IEEE
DOI 10.1109/ECIT52743.2021.00073

that the model might be more accurate for predicting people with high risks of poverty since the data is more abundant.

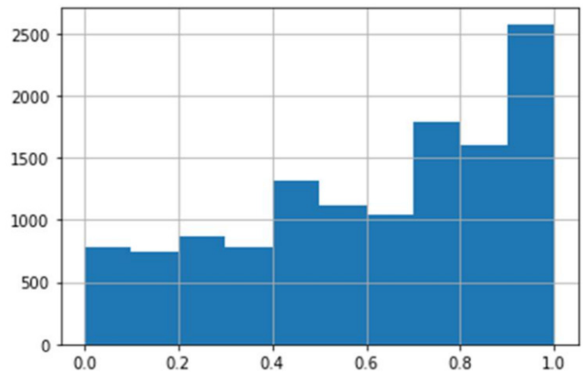

Figure 1. The distribution of poverty probability

Then, we look through the distribution of age:

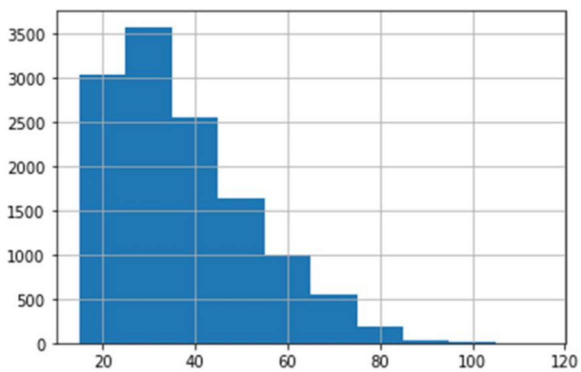

Figure 2. The distribution of age

The data we see is negatively skewed, with over 3500 concentrates around thirdly, a relatively young age. This shows that the prediction mainly focuses on young adults who normally just finish their school and start working. This information is important, since their age and working experiences might influence their job position, income and other factors.

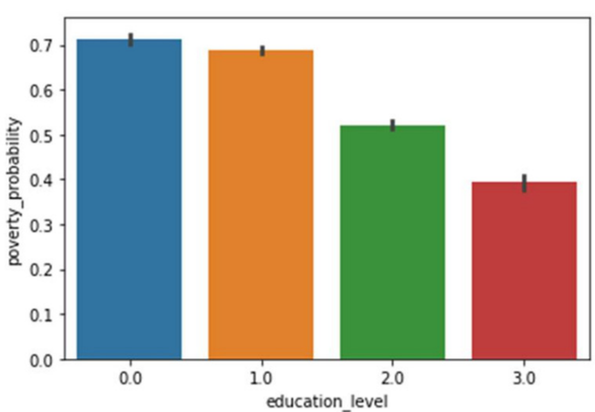

Figure 3. The graph of education level-poverty possibility

Above is the graph of education level-poverty possibility. There is a negative relationship between the education level and poverty possibility, indication that higher education level is likely to have a higher poverty rate.

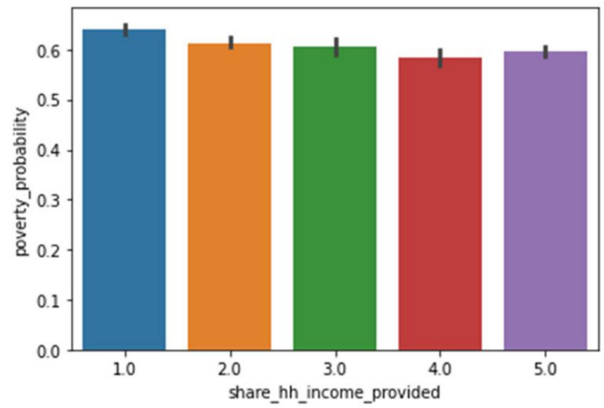

Figure 4. The relationship between household income and poverty possibility

We also look at the relationship between household income and poverty possibility. However, as is shown above, the relationship is less obvious. Although there is still a negative relationship, the differences are relatively smaller. Hence, household income might not be a significant factor for prediction.

C. Preprocessing

Two mean steps in preprocessing are filling the missing value and transforming the data type.

Firstly, looking through the data set, we can find that two features have missing values, educational values and shared income provided. Hence, we look through the distribution of the data. The distribution of education level is shown Figure 5.

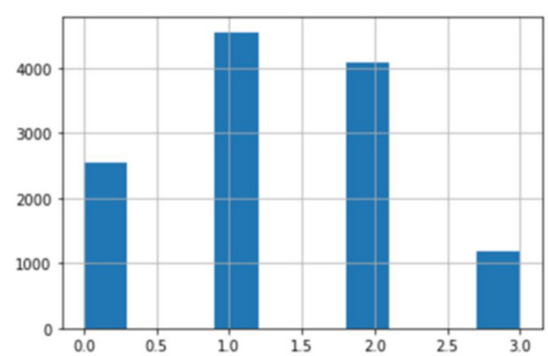

Figure 5. The distribution of education level

From the graph, we can clearly see the data mostly concentrate on 1 and 2. Therefore, we take the median for supplementing the missing value. Similarly, we also use the median to fill out the missing value of shared household income.

Next, some type of data is not suitable for fitting in the models. The data type "object" is not value or "true and false." Instead, they are letters in category. Among our data the features country, religion, relationship to household head, and employment category last year need to be changed. Hence, we use get_dummies function for changing them from "ABCD" to the form of "0" and "1".

D. *Training for Regression*

The next step is fitting the data into different models. For regression, our goal is to calculate the relationship between factors and poverty possibility. To test the accuracy of each model, we use mean squared error and mean absolute error.

Mean absolute error is the average absolute error of each estimation, as shown below:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{2}$$

Mean squared error is the measurement of the average squared differences between the estimated value and actual value. The closer the value is to zero, the better the model is at estimating the result. The basic function is shown below. Comparing with MAE, MSE is better for noting out the large error, because squaring can magnify the differences.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \tag{3}$$

Four models in total have been applied to calculate the continuous relationship.

1) *Linear regression*

Linear regression is a parametric regression. It is used for the prediction of the continuous relationship between independent variables and dependent variables. The result of linear regression shows the possible relationship between forecasting and variables. The basic form is shown below:

$$y = \sum_{i=1}^{n} a_i x_i + b + e \tag{4}$$

However, regression is highly likely to involve the problem of multilinearity, which means the variables are interrelated. For example, in poverty prediction, job position might directly link to income and savings. To solve the problem, lasso regression is used. Lasso regression adds a penalty to the loss function of regression. The improved function is shown below. Therefore, with lasso, we can only keep the factors that influence the result to a large extent. If not, a large number of variables would make the model more too complex while many factors are actually meaningless.

$$\hat{\beta}^{lasso} = \underset{\beta \in \mathbb{R}_p}{\operatorname{argmax}} ||y - X\beta||_2^2 + \lambda||\beta||_1 \tag{5}$$

Lasso regression is very efficient at identifying the important features by turning the parameters of less important features to zero as the graph shows. The intercept of the model is on the y-axis.

We first apply linear regression to all the features to calculate the relationship. However, the performance is not well as expected, the adjusted R-squared reaches merely 0.362 due to the problem of multilinearity. Therefore, we apply lasso regression to improve the performance with alpha=0.01 and the result is summarized in the following chart. Lasso regression successfully solve the problem of multilinearity, and cut off the insignificant features. For instance, whether a person can calculate, do adding, and do dividing, is closely related with their education level, while education level is a more comprehensive feature in terms of one's literacy. Thereafter, lasso regression decreases the coefficient of those kinds of insignificant factors to 0. The features with a coefficient larger than 0.01 are summarized in the TABLE I:

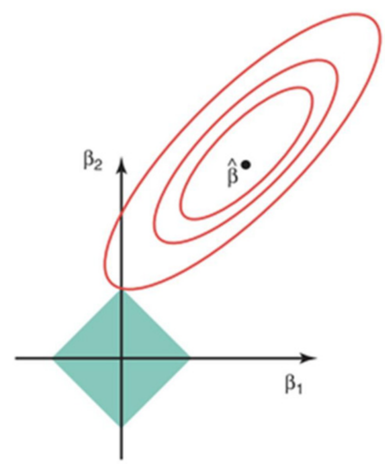

Figure 6. Lasso Regression

TABLE I. THE FEATURES WITH A COEFFICIENT LARGER THAN 0.01

Variables	Coefficient
is urban	-0.0393
age	-0.0166
education level	-0.0459
number of shock last year	0.0636
num financial activity last year	-0.0166
phone technology	-0.0641
can use interest	-0.0124
country A	0.0385
country D	0.0502
country F	-0.0171
country I	-0.0203
country J	-0.0228

We also apply lots of methods to exam the result. One of them is the adjusted coefficient of determination or adjusted R^2, shown below, where n is the size of the sample, and p is the number of independent variables. The use of n and p prevents overfitting.

$$Adjusted\ R^2 = 1 - \frac{\sum(y-\hat{y})^2/(n-p-1)}{\sum(y-\bar{y})^2/(n-1)} \tag{6}$$

Another important method is t-test. T-test is used to determine whether the parameters are accurate and significant. It works by assuming the parameter $\hat{a}$ that follows student's t distribution to be 0, and calculates the possibility P through the distribution. If P≥0.05, the assumption stands, and the parameter is not significant. If P≤0.05, the assumption is not true, and the parameter is significant. Therefore, we also apply this method to test whether the features of poverty prediction are significant.

2) Decision Tree

First created by J. Ross Quinlan at the University of Sydney, decision tree is another model for regression and classification. It is built through splitting data set to form different branches through determining whether they fulfill the statement. Thus, it will form a flowchart-like structure. Gini index is the measure of impurity with a value from 0 to 1. A higher value indicates higher impurity. Gini index is shown below:

$$\sum_{m=1}^{|T|} qm \sum_{k=1}^{K} \hat{p}\, mk(1-\hat{p}mk) \quad (7)$$

We adjust the model by changing the max depth and derive the mean squared error, mean absolute error, and accuracy to evaluate. Figure 7 is a demonstration of the decision tree when max depth equals to 3. The model uses education level, country, and urban as features for splitting the data.

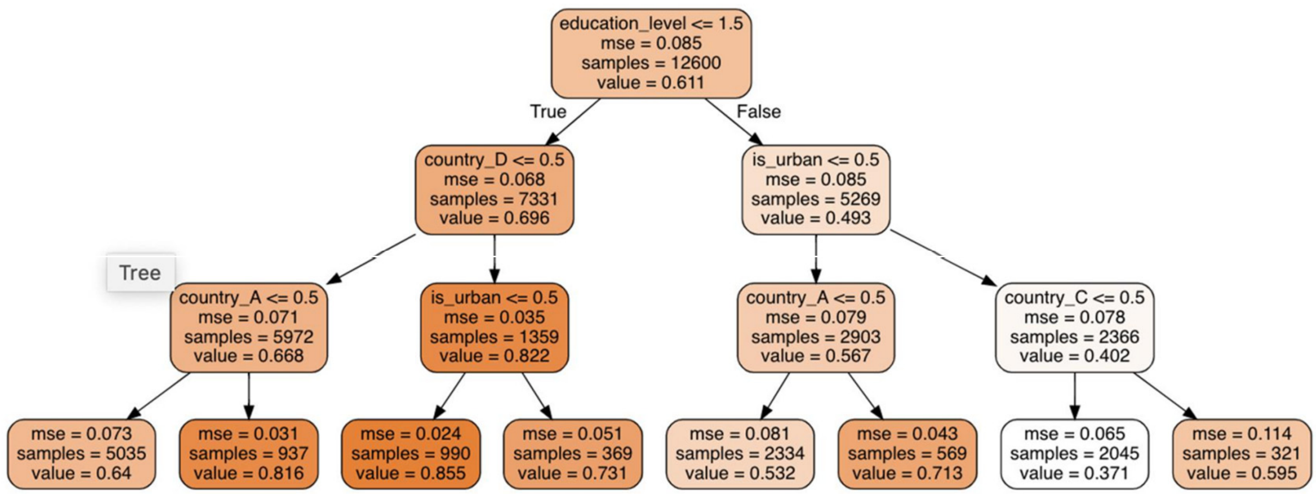

Figure 7. The decision tree

However, decision tree suffers the problem of overfitting, since the more branches it has, the more accurate it will become. Therefore, we further expend the model into random forest and gradient descending decision tree.

3) Random Forest

Bagging is a model averaging approach to reduce the variance. It works through training parts of the data to fit the model respectively, and averaging the result of all the models to derive the best result. In addition, the chance of overfitting is reduced through bagging.

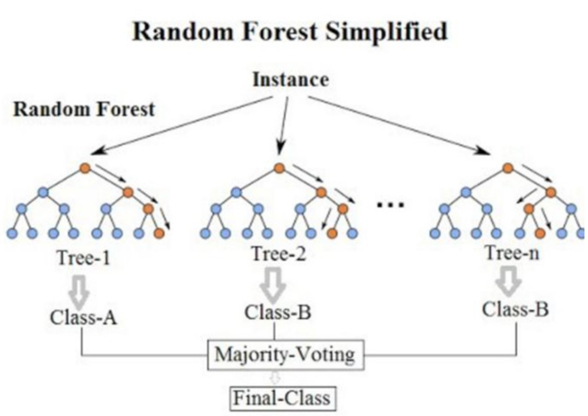

Figure 8. Random Forest

Random forest model is the application of bagging on decision tree to minimize variance and prevent overfitting. It is done through randomly training a fraction of the sample, and then take the mean score for regression or the majority for classification. The general process is demonstrated below.

Moreover, in order to find out the best parameters, cross validation is used. K-fold cross validation is a method to evaluate the accuracy of model and find out the best parameter, α. Different values of α will be given and the data set will be split into k fold. For a single split, a fold will be the test data while the remaining are training data. Therefore, it ensures that all the data are used using CV. Then, the (k-1) fold will be used to fit the model while the other one for evaluating. Therefore, through averaging the cross-validation score, we find out the best value for α.

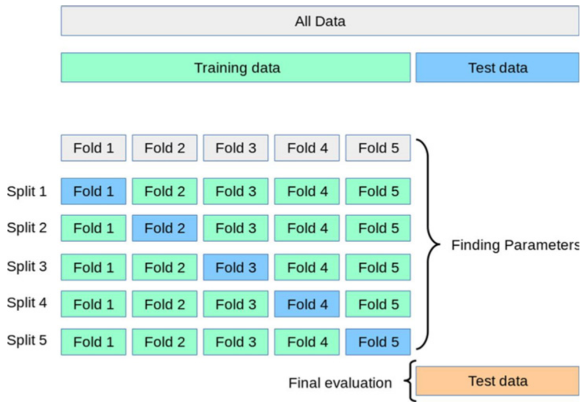

Figure 9. Cross Validation

Hence, we try out the following pairs of parameters. The result of cross validation is shown in the TABLE II:

TABLE II. THE RESULT OF CROSS VALIDATION

Max Depth	**CV Score**
3	0.0499
5	0.0499
7	0.0499
9	0.0500
11	0.0499

From the table, we can clearly see that the CV score reaches the highest when max depth equals to 9.

TABLE III. THE RESULT OF CROSS VALIDATION

Max features	**CV Score**
5	0.04985974849413673
10	0.049894834429343754
15	0.050005428284897534
20	0.04997544065899696

For max features, cv score reaches the highest at 15.

Therefore, we use " max_depth=9, max_features=25, n_estimators=50" in the model.

4) *Gradient Boosting*

Boosting is a method to reduce the bias in training the model. It works through adjusting the weight on the error in the previous model when training for another one. Therefore, it is able to reduce the bias in training the data.

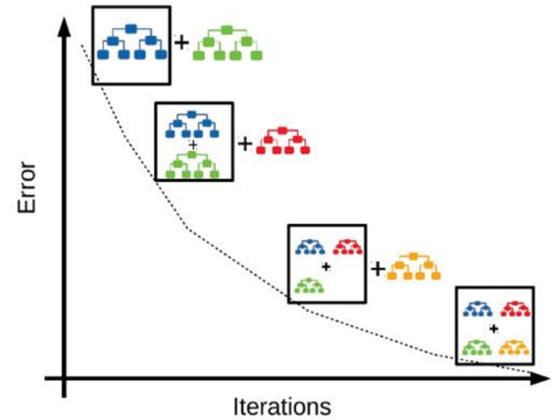

Figure 10. Gradient Boosting

Gradient boosting is the application of boosting to decision tree to minimize bias in the model and hence increasing the accuracy. As is shown in Figure 10, the model works by gradually reduce the error at every step. However, overfitting is likely to happen due to its mechanism. To prevent overfitting, we use a fraction of data and a portion of features for model training every time.

To find out the best parameter, we use grid search this time. Grid search is an optimization algorism for selecting the best parameter. It is guided by performance metrics, usually by cross validation introduced before. It calculates pairs of best parameters at the same time through doing cross validation overtimes. Hence, we use grid search as an efficient tool to find out the parameters with best accuracy. We use the parameters shown in the TABLE IV.

TABLE IV. THE PARAMETERS

Max Depth	3,5,7,9
N Estimator	20,30,40,50
Max Features	10,15,20,25

The result turns out that "max_depth=7, max_features=20, n_estimators=50" is the best pair of parameters. Hence, we apply the pair of parameters to the model.

E. *Training for Classification*

The goal for classification is to predict only whether or not the person will be in poverty in the future. Statistical relationship will not be calculated. We use accuracy to determine the performance of the models.

1) *Decision Tree*

We apply decision tree classifier to test the classification relationship. To find out the best parameter, we also apply cross validation:

TABLE V. THE RESULT OF CROSS VALIDATION

Max Depth	**CV Score**
5	0.0968
7	0.0963
9	0.0966

11	0.0972
13	0.0970
15	0.0968

CV score reaches the best when max depth equals to 11. Hence, we use "max_depth=11" in the model.

2) Random Forest

We employed grid search to find out the best pair of parameters:

TABLE VI. THE PARAMETERS

Max Depth	3,5,7,9
N Estimator	20,30,40,50
Max Features	10,15,20,25

"Max_depth=9, max_features=20, n_estimators=40" becomes the best pair. Hence, we apply the pair in our model.

3) Gradient Boosting

Similarly, grid search shows "max_depth=7, max_features=20, n_estimators=50" as the best pair of parameters. Therefore, we employ this group in the model.

TABLE VII. THE PARAMETERS

Max Depth	3,5,7,9
N Estimator	20,30,40,50
Max Features	10,15,20,25

4) Neural Network

Neural network is a complex model for classification, similar to a multi-layered logistic regression. Linear function, $y = \sum_{i=1}^{n} a_i x_i + b$, is used from one layer to the other. However, each layer requires an activation function. Commonly, ReLu is used in the hidden layer in the middle and Sigmoid function is for the last one. An activation function is used to transfer the linear relationship into classification, mapping the result between 0 and 1. The more layers it contains, the more complex the model will be. Usually, a common neural network would contain three layers, which is also used in this research. Neural network is trained through the loss function as shown below.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \times \log\big(p(y_i)\big) + (1 - y_i) \times \log\,(1 - p(y_i)) \quad (8)$$

Learning rate is also very important for reaching the optima. A small learning rate is time-consuming, requiring many updates, whereas a large learning leads to divergent behavior. Therefore, it is important to choose an optimal learning rate.

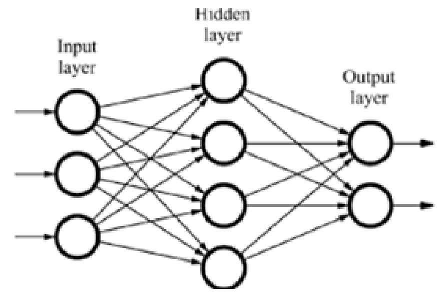

Figure 11. Neural Network

Hence, before fitting the data into the models, we first normalize the input to mean 0 and standard deviation of 1. By doing so, we ensure the learning rate to be optimal, neither too quick nor slow. Additionally, we employ Adam as an optimizer. Adam is one of the most commonly used optimizers for computing the learning rate. It applies momentum, adding the gradient calculated together with the weighted value of the previous value. Therefore, it can ensure the learning rate to stay within an effective speed. The function of the optimizer is shown below.

$$v = \beta_1 v + (1 - \beta_1) dw \quad (9)$$

$$s = \beta_2 + (1 - \beta_2) dw^2 \quad (10)$$

$$w = w - \alpha \frac{v}{\sqrt{s + \epsilon}} \quad (11)$$

We set the following parameters:

TABLE VIII. THE PARAMETERS

Batch Size	50
Epoch	20
Verbose	1
Validation Split	0.1
Dropout Rate	0.3
Optimizer	Adam
Loss Function	Binary cross entropy
Activation Function	Relu, sigmoid

In addition, early stopping is employed, which is a method to prevent overfitting. Early stopping stop training the model when the performance of the model on the validation data set starts to decrease. By setting the patience equaling to 2, the model will automatically stop training after the value accuracy keeps decreasing for 2 epochs.

III. RESULT

A. Results of Regression

1) Linear regression

TABLE IX is a summary of the result of normal linear regression.

TABLE IX. A SUMMARY OF THE RESULT OF NORMAL LINEAR REGRESSION

Dep. Variable:	poverty_probability	**R-squared:**	0.365
Model:	OLS	**Adj. R-squared:**	0.362

Method:	Least Squares	**F-statistic:**	103.0
Date:	Sun, 13 Sep 2020	**Prob (F-statistic):**	0.00
Time:	13:04:43	**Log-Likelihood:**	519.32
No.Observations:	12600	**AIC:**	-896.6
Df Residuals:	12529	**BIC:**	-368.3
Df Model:	70		
Covariance Type:	nonrobust		

TABLE X shows the result of lasso regression.

TABLE X. THE RESULT OF LASSO REGRESSION

Dep. Variable:	poverty_probability	**R-squared:**	0.792
Model:	OLS	**Adj. R-squared:**	0.792
Method:	Least Squares	**F-statistic:**	3432.
Date:	Thu, 03 Sep 2020	**Prob (F-statistic):**	0.00
Time:	12:13:16	**Log-Likelihood:**	-3061.4
No.Observations:	12600	**AIC:**	6151.
Df Residuals:	12586	**BIC:**	6255.
Df Model:	14		
Covariance Type:	nonrobust		

We can clearly see that the r-squared improved to a large extent with lasso. This is because lasso regression successfully solves the problem of multilinearity which decreases the r-squared. Result of t-test is shown in the following form:

TABLE XI. THE RESULT OF T-TEST

	std err	**t**	**P>|t|**
is urban	0.006	-6.282	0.000
age	0.000	49.162	0.000
education level	0.004	11.259	0.000
num of shock last year	0.003	34.939	0.000
num funancial activity last year	0.002	-9.834	0.000
phone technology	0.005	-14.776	0.000
can use interest	0.008	0.819	0.000
Country A	0.009	36.900	0.000
Country d	0.009	39.048	0.000
Country f	0.009	8.283	0.000
Country i	0.009	14.588	0.000
Country J	0.009	2.443	0.015

All the p values are smaller than 0.05, suggesting that all the factors are significant.

Conclusion:

For linear regression, the most important features that are number of shock last year, education level, phone technology, and country. For a person, larger number of shock (like economic shocks, natural disasters, etc.), low education level, inaccessibility to phone and internet, and living in a less developed country increase the possibility of poverty. However, factors like religion are less significant.

2) Decision tree

We tried out different parameters in the model, and the result is summarized in TABLE XII. As is shown, although the mean squared and absolute error keep decreasing, accuracy reaches the highest when max depth equals to 5. Hence, we conclude that the decision tree works best when max depth equals to 5 with an accuracy of 0.8159.

TABLE XII. THE RESULT OF THE MODEL

Max depth	**Mean Squared error**	**Mean absolute error**	**accuracy**
3	0.0654	0.2126	0.7134
5	0.0645	0.1972	0.8159
7	0.0514	0.1819	0.7612
9	0.0445	0.1655	0.7856

3) Random Forest

Cross validation shows" max_depth=9, max_features=25, n_estimators=50 " as the best of parameters in random forest. We calculate the mean square error equals to 0.0499 and the mean absolute error is 0.1828. The top 16 feature importance is shown in Figure 12:

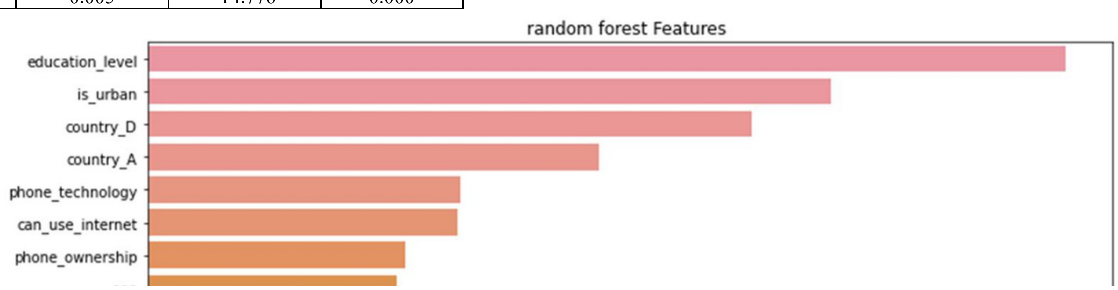

Figure 12. Random forest Features

Conclusion:

For random forest, the most important features are education level, urbanization, country, phone technology, internet accessibility, phone ownership, age, whether can text, number of financial activities, formal savings, and financial shocks. In particular, education level stands out as the most influential factor, with importance more than 0.095.

4) Gradient Boosting

Using ""max_depth=7, max_features=20, n_estimators=50", we derive the mean square error equals to 0.0484. Mean absolute error equals to 0.1771. The top 16 importance is shown in Figure 13:

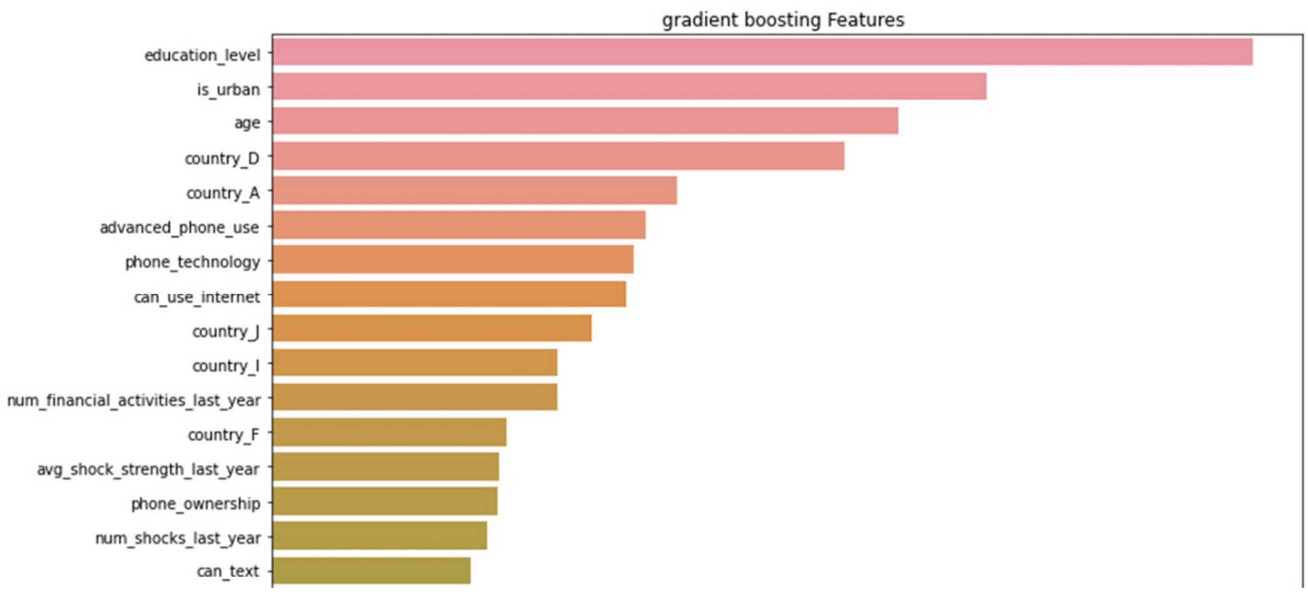

Figure 13. Gradient boosting Features

Conclusion:

Similar to random forest, education level stands out to be the most important factor. Other significant features include: whether is urban, age, country, advanced phone use, phone technology, internet accessibility, number of financial activities, number of shocks and their average strength, phone ownership, whether can text. Factors related to technology are all shown to be important in gradient boosting model.

Summary for the evaluation of all the models in regression:

TABLE XIII. SUMMARY FOR THE EVALUATION OF ALL THE MODELS

Models	Mean Square Error	Mean Absolute error
Decision Tree	0.0574	0.1945
Random Forest	0.0499	0.1828
Gradient Boosting	0.0484	0.1771

For lasso regression, the adj. r-squared is 0.792, showing that the correlation is relatively strong. Its advantage is that lasso regression is simple, intuitive, and easy to understand. However, the simplicity also restricts its ability to form a more complex and accurate model. A single decision tree, although easy to interpret, is also relatively inaccurate, with large possibility of errors. Random forest and gradient boosting improve the accuracy to a large extent, with gradient boosting having the highest performance.

B. Result for Classification

1) Decision Tree

Using different max depth, we calculate the mean squared error, absolute error, and accuracy. Although cross validation identifies "max_depth=11" as the best parameter. We can clearly observe that the errors keep decreasing and the accuracy keeps increasing due to overfitting. Therefore, a single decision tree is not accurate enough.

TABLE XIV. THE RESULT OF THE MODEL

Max depth	Mean Squared error	Mean absolute error	accuracy
3	0.0723	0.2113	0.7204
5	0.0670	0.2021	0.7418
7	0.0646	0.1974	0.7662
9	0.0639	0.1959	0.7957
11	0.0351	0.1399	0.8210
13	0.0246	0.1071	0.8666

Therefore, we turn to random forest and gradient boosting model.

2) Random Forest

Using "max_depth=7, max_features=20, n_estimators=50", the accuracy reaches 0.7575. The top 10 feature importance is shown in Figure 14:

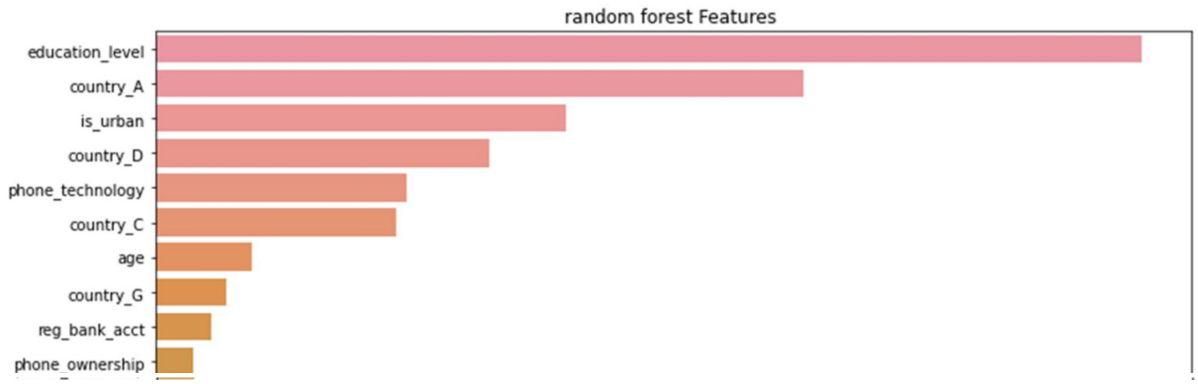

Figure 14. Random forest Features

Conclusion:

For random forest classifier, the most important features are education level, country, urban, phone technology, age, bank account, and phone ownership.

3) Gradient Boosting

Similarly, grid search shows "(max_depth=7, max_features=20, n_estimators=50" as the best pair of parameters. Hence, we calculate the accuracy 0.7853. The top 10 feature importance is shown Figure 15:

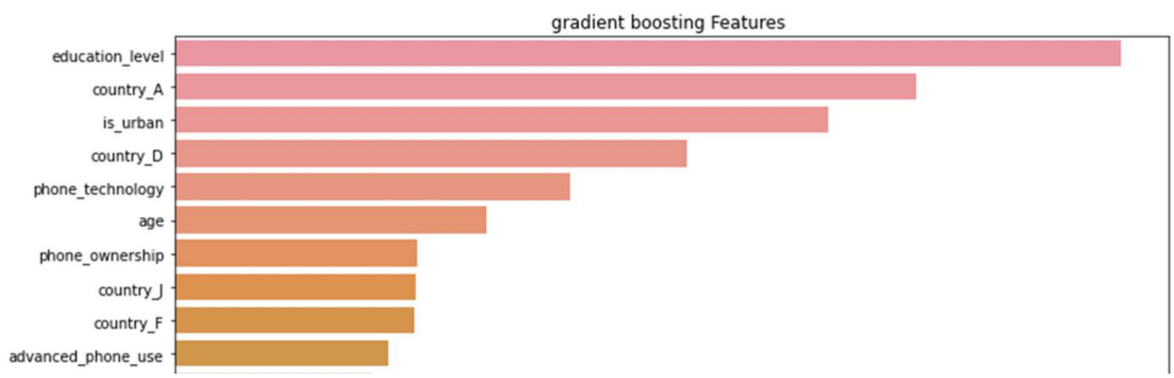

Figure 15. Gradient boosting Features

Conclusion:

Similar to gradient boosting regressor, factors related to technology play a large role. In addition, education level, country, and urban are among the top important features.

4) Neural Network

The following is a summary of neural network result in training:

```
array([[0.7846341 ],
       [0.7084472 ],
       [0.98313475],
       ...,
       [0.5342681 ],
       [0.7586027 ],
       [0.550282 ]], dtype=float32)
```

Using the parameters and method introduced TABLE XV, we derived the following result:

TABLE XV. THE PARAMETERS AND METHOD

Epoch	Loss	Accuracy	Value Loss	Value Accuracy
1	0.5406	0.7303	0.4825	0.7627
2	0.3476	0.7609	0.4711	0.7786
3	0.4756	0.7708	0.4610	0.7794
4	0.4630	0.7779	0.4691	0.7770
5	0.4563	0.7778	0.4708	0.7698

During the third epoch, neural network reaches its highest accuracy, 0.7794.

Summary for classification models:

TABLE XVI. SUMMARY FOR CLASSIFICATION MODELS

Model	Accuacy
Decision Tree	0.8210
Random Forest	0.7575
Gradient Boosting	0.7853
Neural Network	0.7794

Although decision tree model has the highest accuracy, the reliability is diminished by overfitting. In addition, it is less complex than the other models. Therefore, we conclude that overall, gradient boosting classifier performs the best with the highest accuracy. Neural network is also very accurate. However, Although it is praised for the complexity, the interpretability is lower than the other models.

IV. DISCUSSION

A. *Factors Lead to Poverty*

Among all the models, almost all of them identify educational level as the most important factor. The reasons behind the factor are interconnected. The loss of education leads to poor knowledge and skills, which will in turns result in unemployment or low-paying job. Furthermore, education expends the horizon and changes the mindset, while uneducated people may have less awareness of certain things that prevent them from falling into poverty. A simple example is the idea that uneducated adults are less likely to send their children to school. ChildFund shows that 30% of the children from poor families do not complete basic high school education in the USA. (Poverty and Education | ChildFund) Therefore, the lack of education actually formed a vicious cycle and may even lead the person into a poverty trap, as shown in Figure 16. A poverty trap is defined as the mechanism that forces the person to remain in poverty. In other words, it refers to the diminishing return of income day by day. For a poverty trap of education, the poor skill hinders a person from developing, hence trapping them into a cycle of poverty.

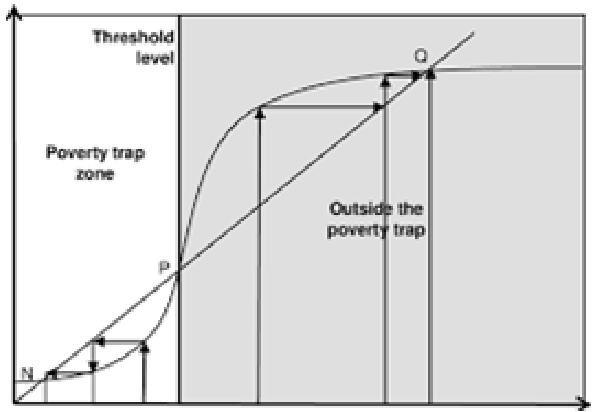

Figure 16. Poverty trap "The s-shape curve and the poverty"

Another significant factor is the country they live in. This means that the countries' economic status, policy, overall trend of poverty, technological development, and so on are tightly connected with the fate of a single individual. People in developing countries are far more likely to be in poverty. In a country with a less stable economy, a person is more likely to experience economic shock due to the volatility. However, in a developed country, various government program provides subsidies and even free education, thus providing opportunities for a person to avoid poverty. For instance, in 2016, the United States has a GDP per capita of 52.6k and an extreme poverty rate of only 2%. By contrast, Zimbabwe has a GDP of 1220. However, the extreme rate reaches 80.1%.

Similar to this factor is urbanization. All the models identify that living in urban area decreases the chance of poverty. This is proved by data from united nation, over 80% of the people of extreme poverty live in rural area worldwide. (Opening Remarks at "Renewing Efforts to Address Rural Poverty to Meet SDG 1 by 2030" | Under-Secretary-General Liu Zhenmin - United Nations Department of Economic and Social Affairs) Limited chances for high salary job, inconvenient transportation, poor facilities might be the reason behind the chance. Even welfare programs, which are supposed to target the poorest people, benefit the residents who live or adjacent to the urban area with higher and better access. This creates an unfair disadvantage among the rural dwellers. Their living location posts a constraint on their development.

In addition, factors related to technology are essential. Phone accessibility, Internet usability, and advanced phone technology are all among the top of the factors. The reason might be that technology adds to the potential of a person's development. The world right now is indispensable to technology. Technologies like phones increase efficiency, provide advanced tools, and open up new opportunities. Processing the technology, people can be connected to the economic and advanced world, thus decreasing the chance of falling into poverty.

Number of shocks is also identified as one of the most important factor. Shocks such as environmental disaster, political conflict, health problem, criminality can increase the vulnerability of a person, and can even have the ability to completely turn the economic situation of a person. For instance, during the 2008 crisis, the poverty rate reaches 31.9%

percentage due to unemployment. ("New 2008 Poverty, Income Data Reveal Only Tip of the Recession Iceberg") Therefore, the more shocks the person experienced, the more likely the person will fall into poverty.

B. *Factors Less Significant*

Surprisingly, many economic factors related to wealth does not influence the possibility to a very large extent many previous study suggests. In the t-test for regression for instance, the possibility for factors like income are well above 0.05 suggesting that these factors do not do not actually have a great influence on predicting the result. The possible reason might be that such indicators meature the current economic situation of a person. As most of the age groups are concentrated around 20-30, an age when people just start off finding jobs and working. Therefore, their income and savings may be relatively lower for a certain period of time. However, the number of financial activities stand out as one of the influencing factor. The reason might be financial activities not only manifest the financial situation of a person, but also the access to banking activities, enough educational background to perform the action, and the awareness of investment and banking.

Religion is also identified as a less influential factor. In lasso regression model, all the religions have a coefficient of 0, meaning that the religion a person belief cannot increase or decrease the chance of being in poverty.

Relationship with families and others are also less significant, showing that social factors are not very useful for predicting poverty. This is also reasonable, since relationship seems to be a less minor thing that influences the life of a person compared with education or technology.

V. Conclusion

The research applies multiple models for the prediction of poverty. For regression and classification, we conclude that gradient boosting model, with "max_depth=7, max_features=20, n_estimators=50" works the best with high accuracy and interpretability. We also conclude education, country, urban, age, shocks, and technology as the most important factors determining whether a person will be poor in the future.

This research again confirms the important role of improving education level and accelerating urbanization in poverty eradication. E-government is one of the most novel and appealing way to benefit both fields. On the one hand, E-government fill gaps in access to knowledge and information. On the other hand, it promotes the improvement of infrastructure in Information and Communication Technologies (ICTs) which have repeatedly demonstrated its potential for alleviating poverty in developing countries. Further, imaginative use of E-government can serve as an enabler of people-centered development and reduce poverty. Some examples about identifying optimal financial resource and risks using e-government are worth learning.

However, there are still some limitations. Firstly, the data can be more diverse and detailed in order to further improve the accuracy and enable a deeper explanation. Secondly, the models are not perfect. Through finding out better parameter and improve the mechanism of the model, higher accuracy can be obtained.

In the future, more study needs to be done on analysis of how exactly the factors have a role in influencing the poverty possibility, and a more detailed relationship between the factors and poverty possibility.

Acknowledgment

I would like to extend my special thanks to my supervisor, Feiran Zhang, who provides me with theoretical study of economics and modeling, innumerable times of assistance, sufficient resources, and helpful advice for revising the thesis. I would also like to express my thanks to my teacher Zhu, who taught me python and useful knowledge of machine learning for completing the program. Moreover, I am also very grateful for my parents, who emotionally support me when facing challenges and difficulties.

References

[1] BohannonAug. 18, John, et al. "Satellite Images Can Map Poverty." Science | AAAS, 18 Aug. 2016. www.sciencemag.org, https://www.sciencemag.org/news/2016/08/satelliteimages-can-map-poverty.

[2] Bureau, US Census. "Income and Poverty in the United States: 2018." The United States Census Bureau. www.census.gov, https://www.census.gov/library/publications/2019/demo/p60-266.html. Accessed 4 Aug. 2020.

[3] Goal-1.Pdf. https://www.un.org/sustainabledevelopment/wpcontent/uploads/2018/09/Goal-1.pdf. Accessed 4 Aug. 2020.

[4] "New 2008 Poverty, Income Data Reveal Only Tip of the Recession Iceberg." Economic Policy Institute, https://www.epi.org/publication/income_picture_20090910/. Accessed 10 Sept. 2020.

[5] Opening Remarks at "Renewing Efforts to Address Rural Poverty to Meet SDG 1 by 2030" | Under-Secretary-General Liu Zhenmin - United Nations Department of Economic and Social Affairs. https://www.un.org/development/desa/statements/mrliu/2019/10/opening-efforts-to-address-rural-poverty-sdg.html. Accessed 10 Sept. 2020.

[6] Poverty and Education | ChildFund. https://www.childfund.org/poverty-and-education/. Accessed 10 Sept. 2020.

[7] Research Findings_ Indicators for Assessing the Risk of Poverty.Pdf. https://mdomsp.gov.hr/UserDocsImages//Twinning01//Research%20findings_%20Indi cators%20for%20assessing%20the%20risk%20of%20poverty.pdf. Accessed 5 Aug. 2020.

[8] Sustainable Development Goals. "Goal 1: End Poverty in All Its Forms Everywhere." United Nations Sustainable Development. www.un.org, https://www.un.org/sustainabledevelopment/poverty/. Accessed 4 Aug. 2020.