Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

# Poverty Prediction using Random Forest based Machine Learning Technique

## Based on Multidimensional Poverty Concept

Daniel D
PG Student
Department of Computer Science,
Impact College of Engineering & Applied Sciences,
Bangalore, India

Rekha M S
Assistant Professor
Department of Computer Science,
Impact College of Engineering & Applied Sciences,
Bangalore, India

***Abstract*—Poverty is an heterogeneous problem and it varies according to time and geographical location. Our study focuses on (1) a method based on multidimensional concept to predict poverty by taking various household characteristics. (2) a novel feature extraction frame work to find a feature that put household in a specific class of poverty. (3) Defining four classes of poverty instead of two traditional levels (poor/non poor). We make use of random forest machine learning algorithm for more accuracy and we will divide data sets into multiple individual data sets.**

***Keywords*—Random forest, multidimensional poverty, poverty levels, data sets***

## I. INTRODUCTION

Poverty prediction and classification is tough, expensive and time consuming. Achieving accuracy is complicated because of data scarcity and security. It may still be hard to define poverty even when various different data are collected from households. Measurement of poverty has two separate complications, (i) Poverty identification (ii) Creation of an index to measure poverty. Income is classically used to overcome the first problem, but the second part is long debated by researchers and practitioners.

Based on the multidimensional poverty concept we predict poverty level. Multidimensional poverty index algorithm is responsible for analysis of different data given by the algorithm executor in order and then determines the level of poverty the user is having. .Randomized Forest algorithm divides the entire data set into set of multiple independent rows. For each of the independent dataset the C4.5 algorithm is executed. The 5 different independent C4.5 algorithms are executed and then class label is generated. After the set is formed the maximum count of class label is found out and then class is determined.

## II. EXISITING SYSTEM

Multidimensional poverty index constitutes the first implementation of the direct method to measure poverty for over 100 developing countries. Multidimensional Poverty Index (MPI), a measure of *acute poverty*, understood as a person's inability to meet minimum international standards in indicators related to the Millennium Development Goals and to core functioning's. The MPI offers a reliable framework that can complement global income poverty estimates.

Disadvantage

1. the method consider the family income of the end user, computes the average income and if it is below the threshold set by MPI review the user is classified into Poverty and Non-Poverty.

## III. PROBLEM DESCRIPTION

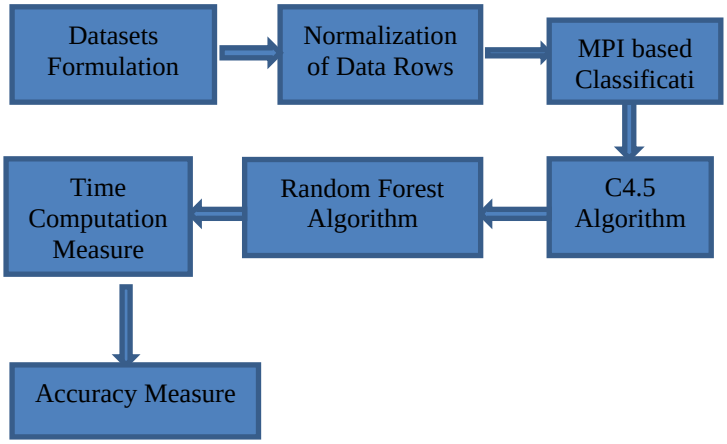


Figure 1: problem description

*Normalization of Data Rows*

This module is responsible for dividing the actual row data with the value highest among all the rows specific to each of the columns.

*MPI based Classification*

This algorithm is responsible for analysis of different data given by the algorithm executor in order and then determines the level of poverty the user is having. The detailed steps can be found as below:

1) obtain the list of attribute1 from the previous history data set for users who have the poverty level label
2) obtain the list of attribute2 from the previous history data set for users who have the poverty level label
3) Compute the summation of list of attribute1
4) Compute the summation of list of attribute2
5) Compute the mean of attribute1
6) Compute the mean of attribute2
7) Compute the standard deviation of list of attribute1
8) Compute the standard deviation of list of attribute2
9) Compute the probability of attribute1

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

$$P_{attribute} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{(\mu-T)^2}{2\sigma^2}\right)}$$

$Where,$

$\sigma = s\tan dard\ \ deviation$

$u = mean$

$T =$

T=current value

10) Compute the probability for attribute2 in the same way
11) Compute the total probability that the patient will have the disease

$$P_{havepoverty} = \frac{1}{2} * \sum_{i=1}^{Natt} p(att \mid have\Pr operty)_i$$

12) obtain the list of attribute1 from the previous history data set for users who do not have the poverty
13) obtain the list of attribute2 from the previous history data set for users who do not have the poverty
14) Compute the summation of list of attribute1
15) Compute the summation of list of attribute2
16) Compute the mean of attribute1
17) Compute the mean of attribute2
18) Compute the standard deviation of list of attribute1
19) Compute the standard deviation of list of attribute2
20) Compute the probability of attribute1

$$P_{attribute} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{(\mu-T)^2}{2\sigma^2}\right)}$$

$Where,$

$\sigma = s\tan dard\ \ deviation$

$u = mean$

$T = currentval\,ue$

21) Compute the probability for attribute2 in the same way.
22) Compute the total probability that the user will have the disease

$$P_{donothave} = \frac{1}{2} * \sum_{i=1}^{N} p(havingproperty \mid ai)\ 2$$

23) Compute the Average Probability from the two classes.
24) P(class1) = p(class1)/(pclass1+pclass2)
25) P(class2)=p(class2)/(pclass1+pclass2)
26) In a similar fashion if there are N classes repeat for N class.
27) Find the maximum value of P.
28) The class to which maximum value of P belongs to is the final class.

*Randomized Forest*

Randomized Forest algorithm is responsible for dividing the entire data set into set of multiple independent rows. For each of the independent dataset the algorithm is executed. The 5 different independent algorithms are executed and then class label is generated. After the set is formed the maximum count of class label is found out and then class is determined.

*Architecture*

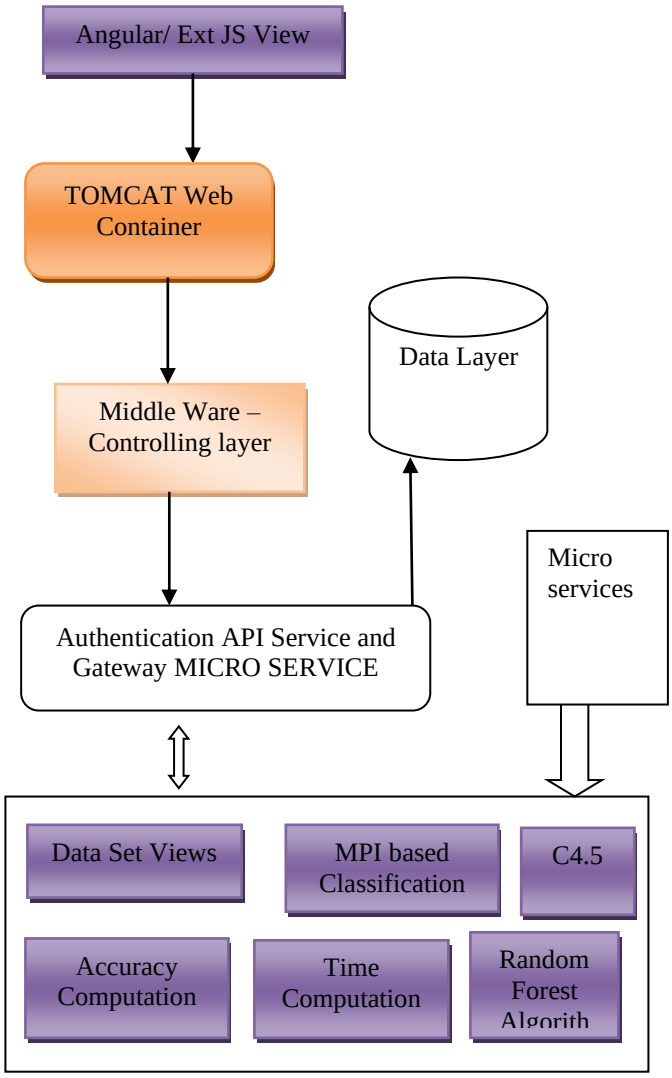


Figure 2: system architecture

IV IMPLEMENTATION

Software development which can be delivered fast, quick adaptation to requirements and collecting feedback on required information. The agile software methods and development is practices based approach empowered with values, principles and practices which make the software development process easier and in faster time.

Agile methods which encompasses individual methods like Extreme programming,

Feature Driven Development, Scrum, etc. are coming into the commercial and academic worlds.

Agility refers to the quality of being agile. Internet software industry and Mobile and wireless application development industry are looking for a very good approach of software development. Conventional software development methods

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

have completely closed the requirements process before analysis and design process. In contrast to the conventional approaches, agile methods allow developers to make late changes in the requirement specification document.

The focus of the agile software development as given by "Agile Software Development Manifesto" is presented in the following:

- Individuals and interactions over processes and tools.
- Working software over comprehensive documentation.
- Customer collaboration over contract negotiation.
- Responding to change over following a plan.

1) There is vital importance of communication between the individual who are in development team, since development centers are located at different places. The necessity of interaction between Individuals over different tools and different versions and processes is very vital.
2) The only objective of software development team is to continuously deliver the working software for the customers. New releases must be produced for frequent intervals. The developers try to keep the code simple, straight forward and technically as advanced as possible and will try to lessen the documentation.
3) The relationship between developers and the stakeholders is most important as the pace and the size of the project grows. The cooperation and negotiation between clients and the developers is the key for the relationship. Agile methods are using in maintaining good relationship with clients.
4) The development team should be well-informed and authorized to consider the possible adjustments and enhancements emerging during the development process.

*Implementation architecture*

The user interface is designed in the HTML/JSP pages and then the request goes to the web container and web container verifies the request in the web.xml file by looking first into the url pattern and then it goes to the servlet name and then it searches for the corresponding servlet name in the servlet tag and looks into the servlet class and creates an object of Action Servlet and then the action servlet will delegate its job to Request Processor.

The request processor will look for the action to which must be called in looked up in the stucts-config.xml and corresponding action form is called and then the action is called. The action class will then call the delegate, then the delegate calls the service and service calls the Data Access layer and results goes exactly in the opposite way and the resultant JSP page is loaded.

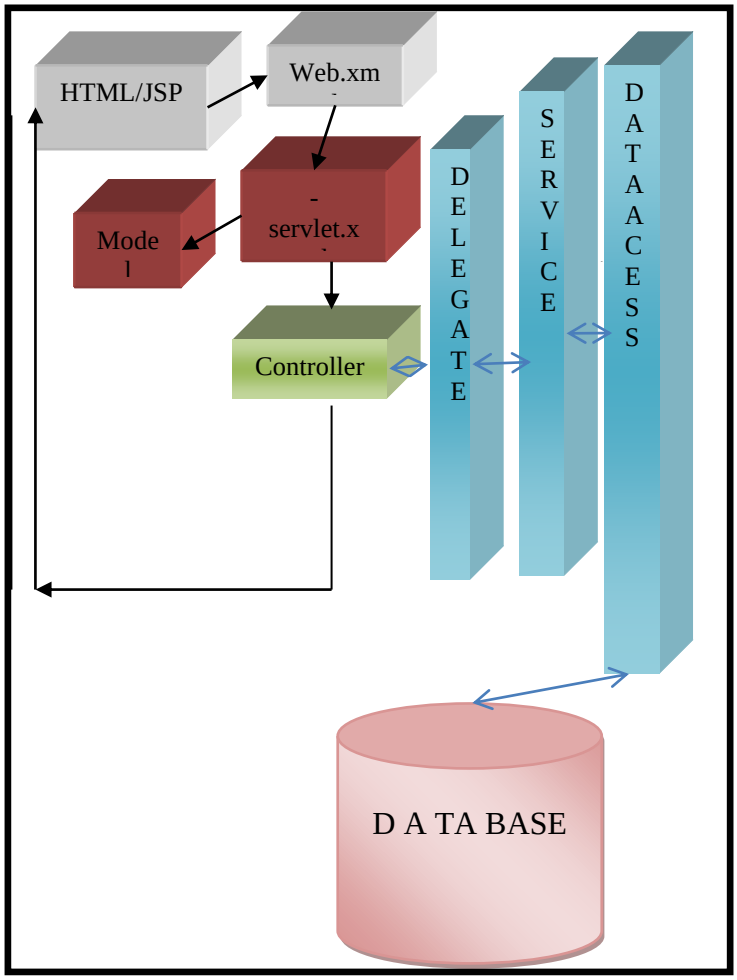


Figure 3: implementation architecture

## V RESULTS

The MPI Classification input in which there are multiple values provided by the end user in order determine the category in which the following attributes

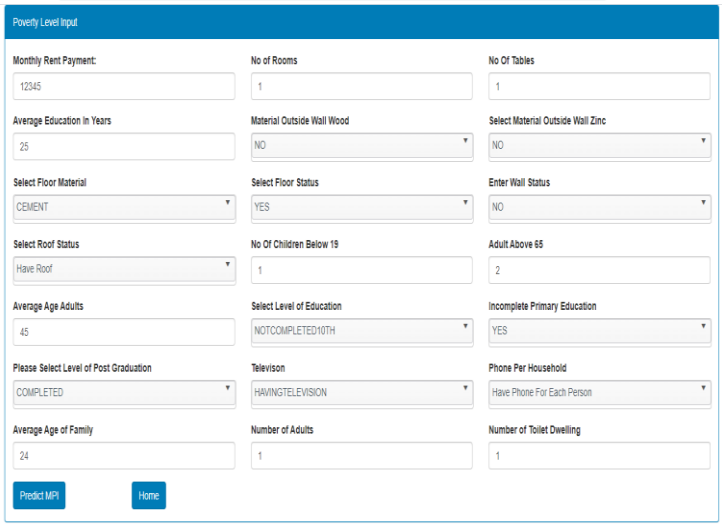


Figure 4: MPI Classification

The classification result for the MPI algorithm. As shown in the fig cluster number which is predicted is 4 and the class

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

label is NONVULNERABLE. The details why class is predicted as NONVULERABLE is based on highest probability

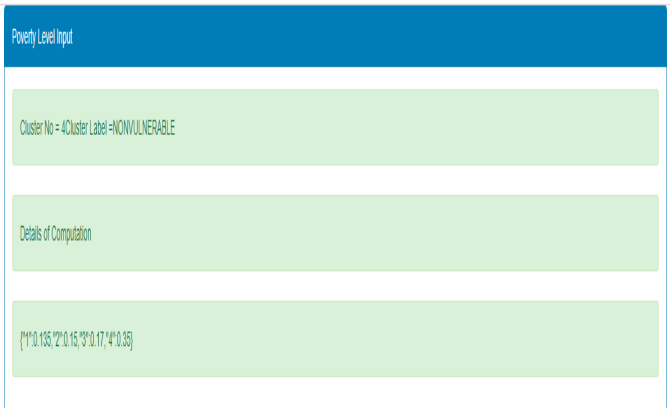


Figure 5: MPI classification result

The random forest algorithm input. In this algorithm the various attributes of the algorithm are discussed. The input attributes are taken into consideration which is responsible for analysis of the data and then executing the random forest algorithm.

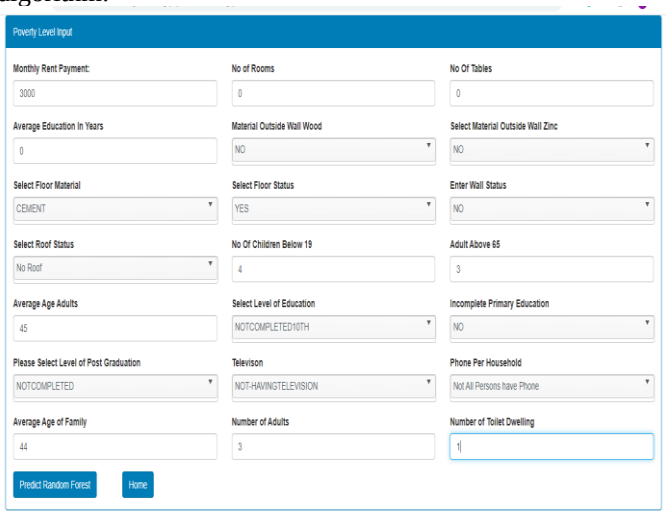


Figure 6: Random forest input

The time taken for various algorithms The Random Forest algorithm will have lowest time taken across all the iterations as compared to MPI method.

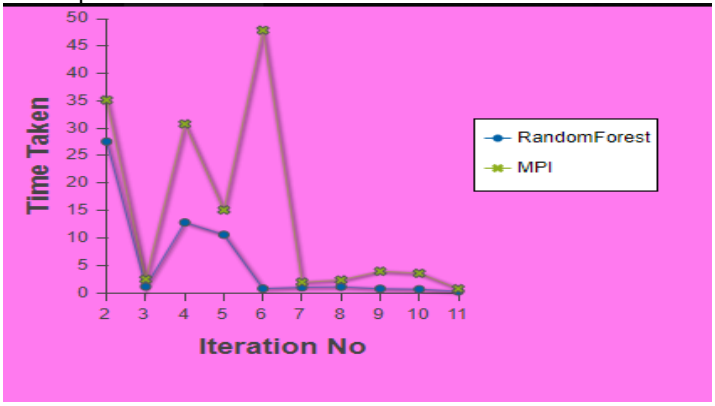


Figure 7: Time taken by algorithm

As shown in the fig the accuracy of the proposed method is 100% as compared to previous method whose value is 92.85%.

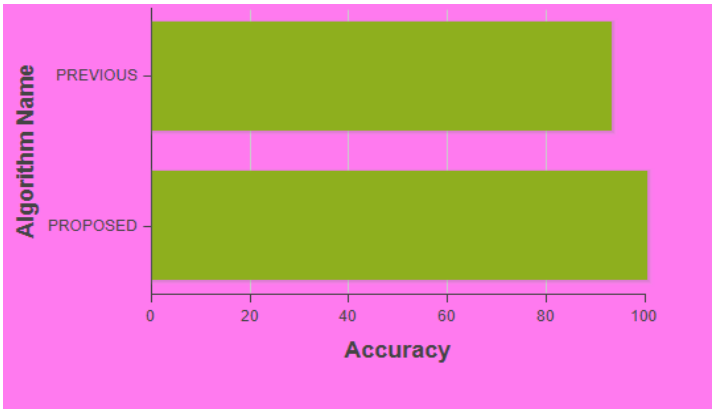


Figure : Accuracy

## VI CONCLUSION

The data sets are divided into multiple independent data sets. For the MPI algorithm all the data rows will act as an input and then prediction of MPI class label is obtained. The random forest algorithm will have divided the entire data sets into multiple independent data sets. From each dataset the output class label is determined if each of the decision tree, the process is repeated for the remaining decision trees. The count of output class label is taken into consideration and the actual class is determined. The comparison of MPI method with Random Forest is compared across all the iterations and the time taken by proposed Random Forest will be lesser than that of

MPI method. The accuracy of Random Forest algorithm is compared with MPI. The accuracy of the proposed method is always higher.

## VII BIBLIOGRAPHY

[1] A. Sen., Poverty: An Ordinal Approach to Measurement, Econometrica, vol. 44, no. 2, p. 219, 1976.

[2] S. Alkire and M. E. Santos, "*Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index*," World Dev., vol. 59, pp. 251274, 2014.

[3] F. Bourguignon and S. R. Chakravarty, "The Measurement of Multidimensional Poverty," J. Econ. Inequal, vol. 1225, no. February, pp. 4142, 2003.

[4] S. Alkire and M. E. Santos, "Multidimensional Poverty Index," Oxford Poverty Hum. Dev. Initiat., no. July, pp. 18, 2010.

[5] N. Nari and N. Quinn, "Alkire-Foster Method The Global MPI Policy Use Public Communication The Global Multidimensional Poverty Index," no. November, 2017.

[6] L. McBride and A. Nichols, "Improved poverty targeting through machine learning: An application to the USAID Poverty Assessment Tools," p. 24, 2015.

[7] N. S. Sani, M. A. Rahman, A. A. Bakar, S. Sahran, and H. Mohd, "Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification," vol. 8, no. 4, pp. 16981705, 2018.

[8] S. Narendranath, S. Khare, D. Gupta, and A. Jyotishi, "Characteristics of Escaping and Falling into Poverty in India: An Analysis of IHDS Panel Data using machine learning approach," 2018 Int. Conf. Adv. Comput. Commun. Informatics, pp. 13911397, 2018.

[9] World bank, "Measuring income and poverty using Proxy Means Tests."

[10] B. B. Pineda-Bautista, J. A. Carrasco-Ochoa, and J. F. Martnez-Trinidad, "General framework for class-specific feature

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

selection," Expert Systems with Applications, vol. 38, no. 8. pp. 1001810024, 2011.

[11] A. Roy, P. D. Mackin, and S. Mukhopadhyay, "Methods for pattern selection, class-specific feature selection and classification for automated learning," Neural Networks, vol. 41. Elsevier Ltd, pp. 113129, 2013.

[12] A. M. P. Canuto, K. M. O. Vale, A. Feitos, and A. Signoretti, "ReinSel: A class-based mechanism for feature selection in ensemble of classifiers," Applied Soft Computing Journal, vol. 12, no. 8. Elsevier B.V., pp. 25172529, 2012.