IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Poverty Level Characterization via Feature Selection and Machine Learning

[1]MAHEK SABHA, [2]SOUNDARYA R, [3]SPOORTHY BALAJI, [4]NIRIKSHA, [5]ANKITHA SHETTY

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor

[1,2,3,4,5] Department of Computer Science and Engineering

[1,2,3,4,5] Alva's Institute of Engineering and Technology, Mijar, India

*Abstract:* Poverty is a persistent socio-cultural problem that necessitates precise definition in order to develop well-designed intervention policies. Unfortunately the poverty-wealthiest scale is not used to categorize people. Surveys are a simple way to establish this. The population is quite large. Subjective opinions are frequently skewed, and the data available is limited. Poverty is a complex issue that varies with the passage of time and geographical location. Our research focuses on (1) a method for predicting outcomes based on a multidimensional concept by taking into account numerous household variables, poverty can be reduced. (2) a storey To find a feature, use the feature extraction framework .a household in a certain poverty level (3) Establishing four distinct classes instead of poverty. For more accuracy, we will divide data sets into numerous distinct data sets using the random forest machine learning technique.

***Index Terms*** **– Random Forest, Multidimensional Poverty, Poverty levels**

## I. INTRODUCTION

Predicting and classifying poverty is difficult, expensive, and time-consuming. Because of data scarcity and security, achieving accuracy is difficult. It may still be difficult. Elucidate poverty even when a variety of data is collected from private residences Poverty measurement is divided into two categories.(I )Identifying poverty (ii) Creating an emergency fund. The first problem is traditionally solved through income, but the second portion has long been a source of disagreement among researchers and practitioners. We forecast poverty levels using the multidimensional poverty concept. The multidimensional poverty index algorithm is in charge of analysing various data provided by the user. After that, the algorithm executor determines the level of difficulty. The user is in a state of destitution. The Randomized Forest technique splits the data set into numerous independent rows. The C4.5 algorithm was used for each of the separate datasets is carried out .The C4.5 algorithms are made up of five separate algorithms that are all independent of one another. After that, a class label is formed. When the set is finished, the maximum number of class labels is calculated, and after that, the class is determined.

## II. Existing System

For more than 100 developing nations, the Multidimensional Poverty Index is the first direct means of measuring poverty. Poverty is multifaceted. MPI is a measure of extreme poverty that is calculated as a percentage of the total population an individual's failure to achieve international minimum requirements in indicators pertaining to the Millennium Development Goals (MDGs) and the Sustainable Development Goals (SDGs) to the fundamentals of the human body's operations. The MPI provides a dependable methodology for estimating global income poverty.

## III. Disadvantage

The approach considers the end user's family income, computes the average income, and classifies the user if it falls below the MPI review level. Poverty and Non-Poverty are two types of poverty.

## IV. Problem Description

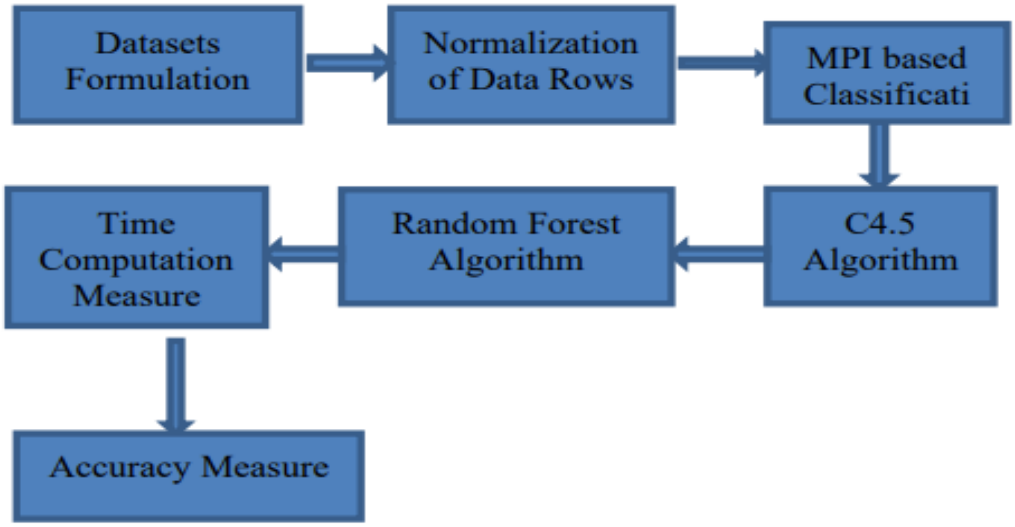


Figure 1: Problem Description

- Normalization of Data Rows

This module is in charge of dividing the actual row data by the highest value among all the rows for each of the columns.

### 1.1 MPI based Classification

This algorithm is in charge of analysing various data provided by the algorithm executor in order to assess the user's level of poverty. The following are the detailed steps:

1.For users with the poverty level label, get a list of attribute1 from the previous historical data set.
2. Get the attribute2 list from the preceding history. Users with the poverty level badge have their own data collection.
3. Compute the total of the attribute1 list
4. Compute the total of the attribute2 list.
5. Calculate the attribute1 mean.
6. Calculate the attribute2 mean.
7. Determine the standard deviation of the attribute1 list.
8. Determine the standard deviation of the attribute2 list.
9. Calculate the likelihood of attribute1 being true.

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

Where,
μ=mean
$^{\upsilon}$=standard deviation
T=current value

10. Do the same thing with attribute2's likelihood.
11. Compute the total probability that the person will have the poverty.
$P_{havepoverty}=0.5*\Sigma p(att|havePoverty)_i$
12. For users with no poverty level label, get a list of attribute1 from the previous historical data set.
13. Get the attribute2 list from the preceding history. Users with the poverty level badge have their own data collection.
14. Compute the total of the attribute1 list.
15. Compute the total of the attribute2 list.
16. Calculate the attribute1 mean.
17. Calculate the attribute2 mean.
18. Determine the standard deviation of the attribute1 list.
19. Determine the standard deviation of the attribute2 list.
20. Calculate the likelihood of attribute1 being true.
21. Do the same thing with attribute2's likelihood.
22. Compute the total probability that the person will have the poverty.
$P_{donothave}=0.5*\Sigma p(havingPoverty|ai)^2$
23. Compute the Average Probability from the two classes.
24. P(class1) = p(class1)/(pclass1+pclass2)
25. P(class2) = p(class2)/(pclass1+pclass2)
26. If there are N classes, repeat the process for each class.
27. Determine P's maximum value.
28. The class in which P has the highest value. This is the last class.

### 1.2 Randomised Forest

The Randomized Forest technique divides the full data set into a set of many independent rows. The algorithm is different for each independent dataset executed. After that, a class label is formed. When the set is finished, the maximum number of class labels is calculated, and after that, the class is determined.

### 1.3 Architecture

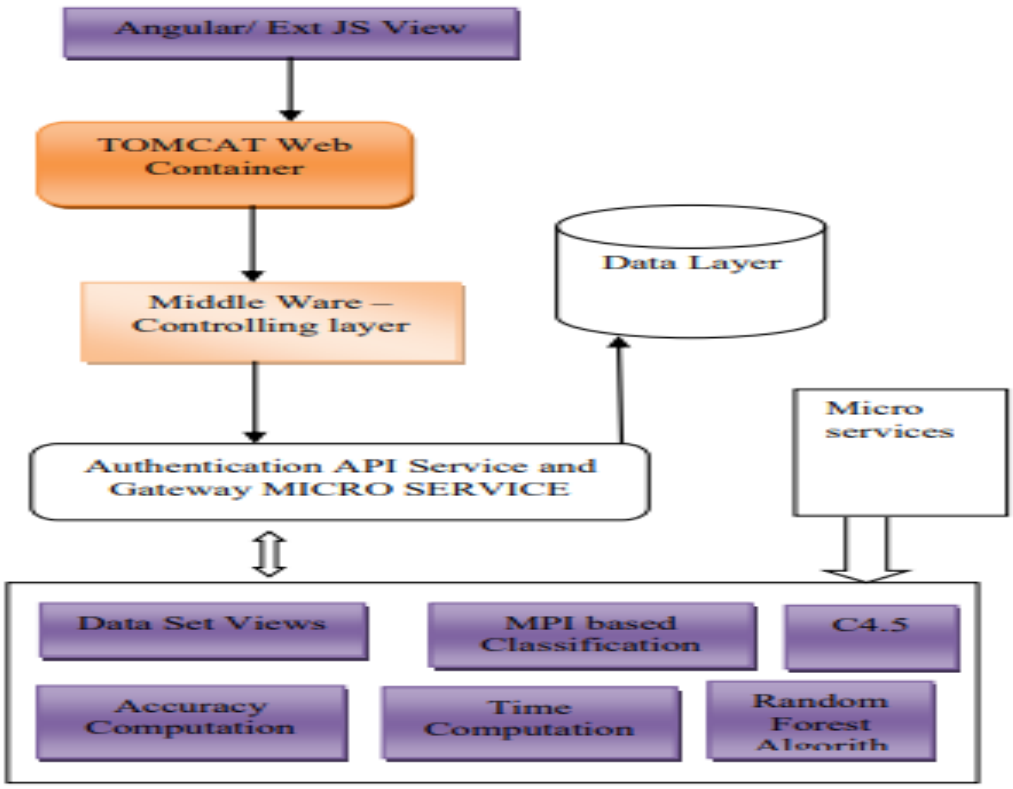


Figure 2: System Architecture

### 1.4 Implementation

Software development that can be produced quickly, with rapid adaptability to needs and feedback on required data. The methods of agile software development . Development is a method based on practises that is aided by technology. The software's ideals, concepts, and practices making the development process easier and faster .Individual methodologies such as Extreme programming, Feature Driven Development, Scrum, and others are being incorporated into Agile methods the business and academic spheres. The trait of being agile is referred to as agility. Software for the internet growth of mobile and wireless applications in the industry. The software sector is seeking for a very good technique development. Methods used in traditional software development Prior to the analysis and design process, the requirements process must be entirely closed. Agile methodologies, in contrast to traditional processes, allow engineers to make late alterations to the document containing the requirement specifications according to Agile Software Development, the goal of agile software development is the “Manifesto for Agile Software Development” is presented in the following:

- People and their interactions, rather than processes and instruments.
- Useful software trumps extensive documentation.
- Contract negotiation is prioritised over customer collaboration.
- Responding to a changeover in a planned manner.

[1] Because development centre are located in different locations, communication between individuals in the development team is critical. Interaction between individuals is essential over a variety of tools and versions and processes are really important.

[2] The software development team's sole goal is to consistently provide working software to clients. For this reason, new releases are required at regular intervals The creators make every effort to preserve the environment as natural as possible. Simple, straightforward, and technically sound code as far as practicable, and will make every effort to reduce the risk documentation.

[3] As the project's pace and size increase, the interaction between developers and stakeholders becomes increasingly critical. Collaboration and the client-developer negotiation is the most important part of the process crucial to the connection. In Agile methodologies are used keeping a positive client connection.

[4] The development team should be well-informed and empowered to evaluate any changes or improvements that may arise during the development process.

### 1.5 Implementation Architecture

Many employers conduct background checks on job candidates through third-party corporations that verify the candidate's background by confirming with the past executive department or university and visiting home address to verify the residence. Some employers conduct checks when they need to be employed associate workers. large cash is spent by the corporate throughout this background verification method. So, there square measure tons of physical document checks while not knowing it's legit or authentic and it takes a large quantity of your time for verification. the most aim is to scale back of these large tasks and third-party involvement which can compromise the system to straightforward, direct, and secure interaction between an organization and also the candidate certificate.

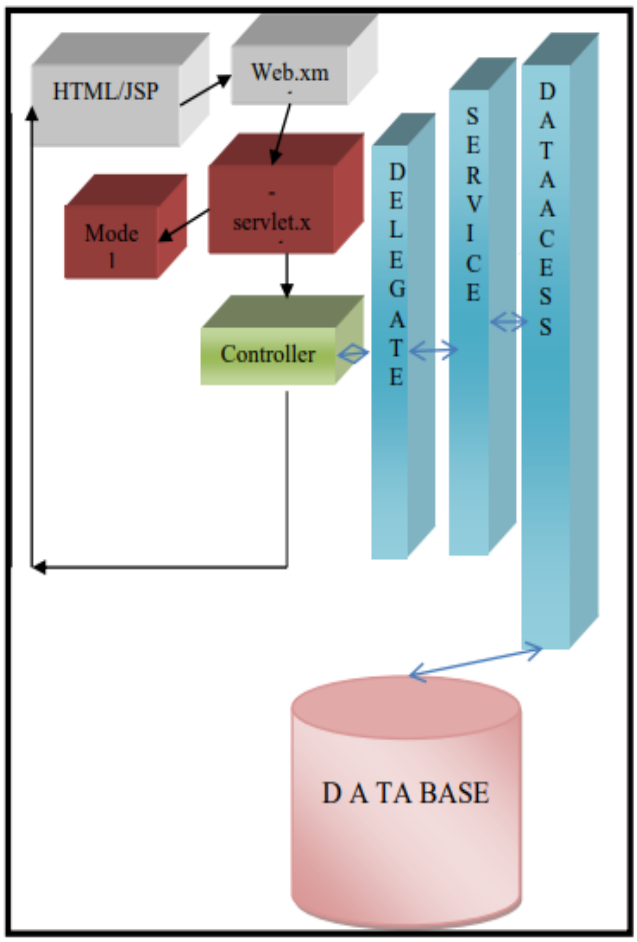


Figure 3: Implementation Architecture

## V. RESULTS

The MPI Classification input, which allows the end user to specify numerous values in order to decide the category in which the following attributes belong.

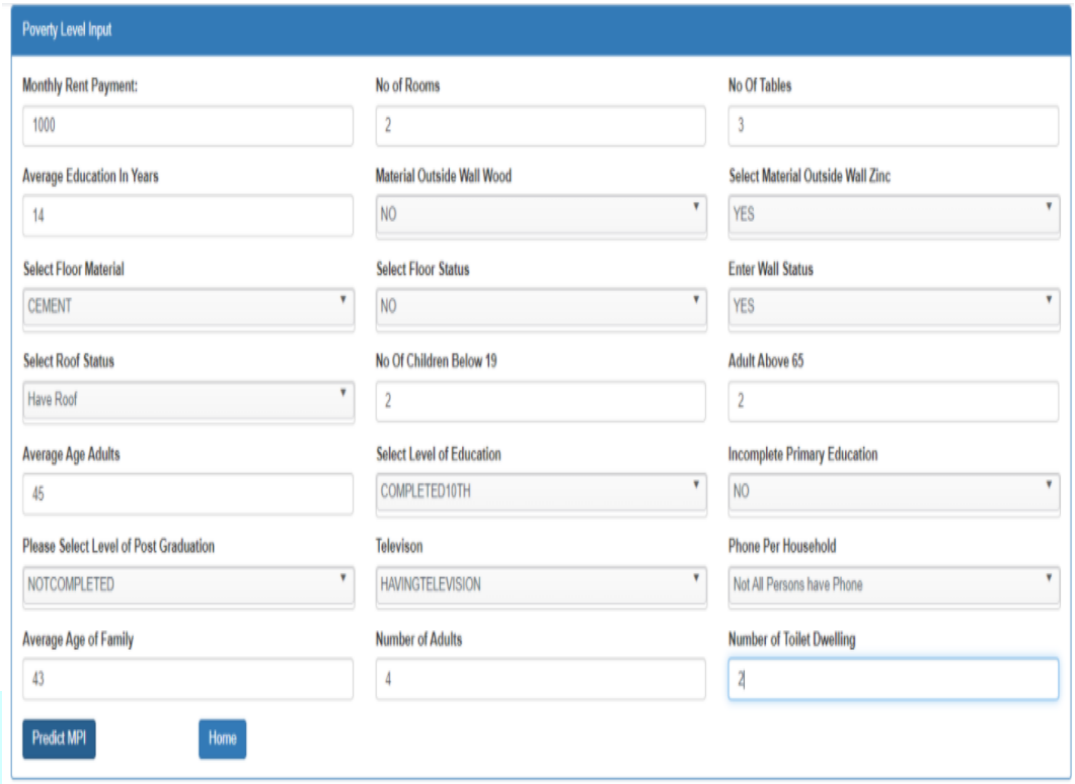


Figure 4: MPI Classification

The MPI algorithm's categorization result. As indicated in the diagram, the expected cluster number is 4 and the class is NONVULNERABLE . The reasons why a class is anticipated to be NONVULERABLE are based on the highest probability.

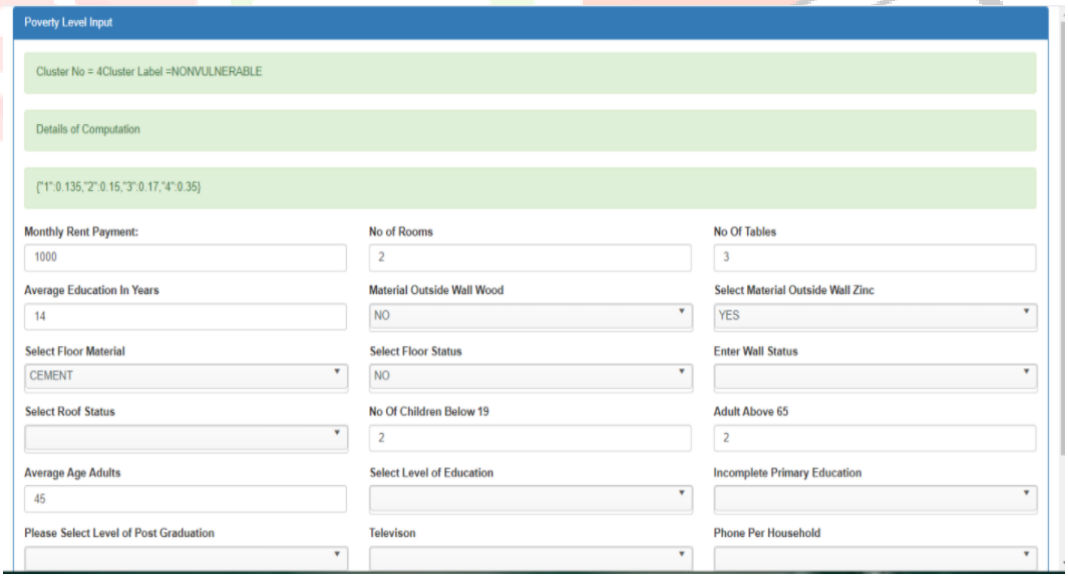


Figure 5: MPI Classification Result

Input to the random forest algorithm. The many properties of the algorithm are covered in this algorithm. The input attributes are taken into account to determine who is to blame to conduct data analysis and then run the random forest algorithm.

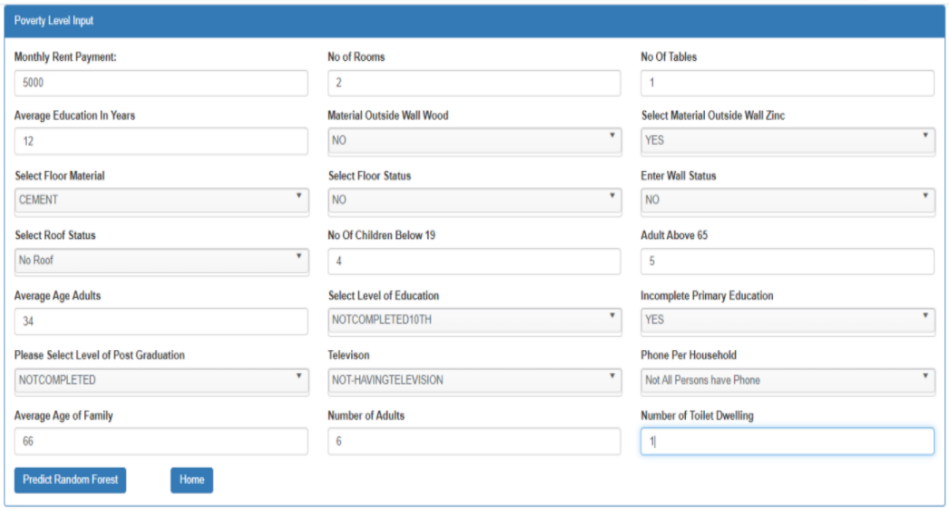


Figure 6: Random Forest Input

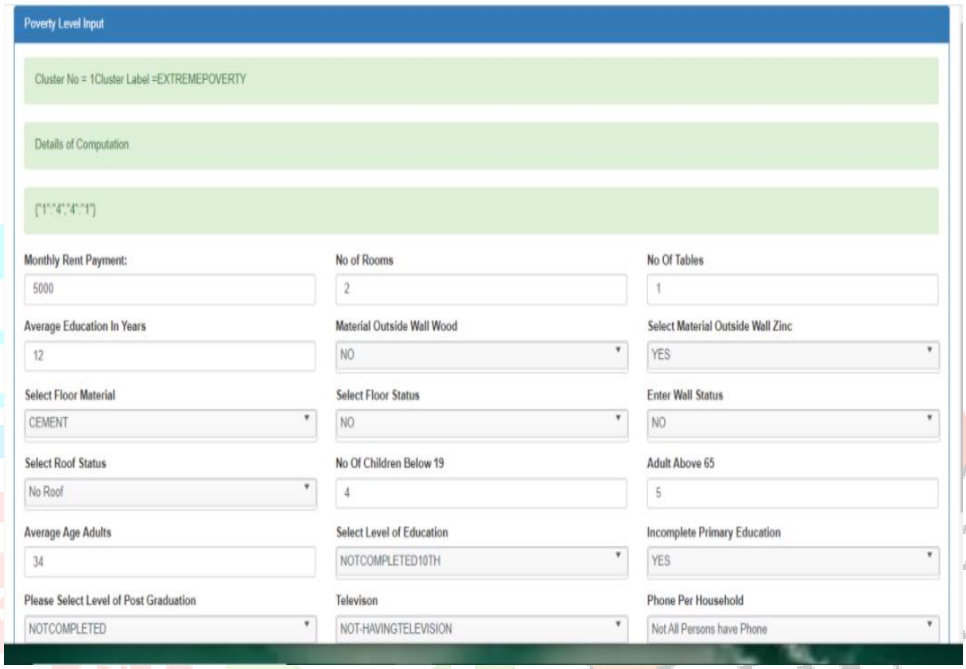


Figure 7:Random Forest output

The amount of time it takes for various algorithms to complete. When compared to the MPI technique, the Random Forest algorithm will take the least amount of time throughout all iterations.

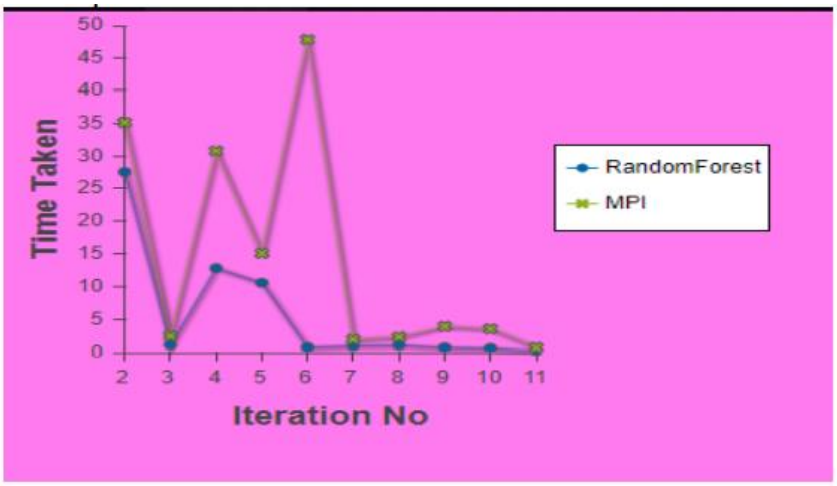


Figure 8: Time Taken by the Algorithm

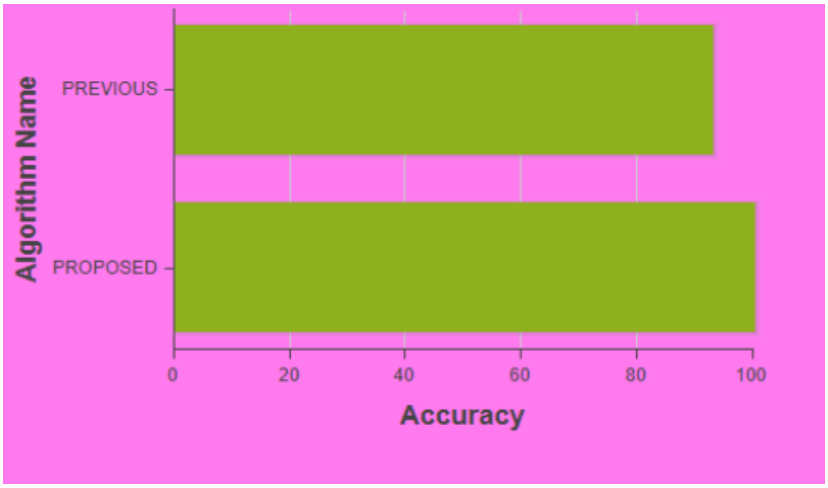


Figure 9: Accuracy

The accuracy of the proposed method is 100 percent, as indicated in the graph, compared to 92.85 percent for the old method.

## VI. CONCLUSION

Multiple separate data sets are created from the data sets. All of the data rows will be used as input for the MPI method, and a prediction of MPI class label will be obtained. The whole data sets will have been partitioned by the random forest technique into a number of separate data sets. The output class label for each decision tree is calculated from each dataset, and the process is repeated for the remaining decision trees. The number of output class labels is considered, as well as the real class is established. The MPI approach and the proposed Random Forest are compared across all iterations, and the suggested Random Forest takes less time. The Random Forest algorithm has a high level of accuracy as compared to MPI The proposed method's accuracy is always aim for the top.

## REFERENCES

[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. MartnezTrinidad, “General framework for class-specific feature 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.