POVERTY ANALYSIS USING MACHINE LEARNING

1.ABSTRACT:

The impact of poverty in all over countries, through a comprehensive analysis of poverty rates, etc. We aim to involve interdisciplinary collaboration, innovation in algorithms and techniques, and a strong emphasis on ethics and responsible use of technology. The analysis begins by defining poverty in its various forms, including income poverty, multidimensional poverty, and relative poverty, to create a comprehensive framework for assessment. An exploration of the underlying factors that contribute to poverty, including economic, social, political, and environmental determinants, is undertaken to uncover the root causes of this complex issue.Regional disparities in poverty rates are examined, shedding light on the geographic variations in poverty and highlighting regions most in need of targeted interventions.The study delves into the profound impacts of poverty on individuals and communities, including health, education, and access to basic services, emphasizing the intergenerational cycle of poverty.An evaluation of existing poverty alleviation policies and programs is conducted to determine their effectiveness and identify areas for improvement.Based on the analysis, a set of evidence-based policy recommendations is provided to address the multifaceted nature of poverty and promote sustainable development, social equity, and economic growth.

To know the poverty ratio and poverty rate across the countries and to thoroughly investigate the multifaceted dimensions of poverty, encompassing income levels, educational opportunities, healthcare accessibility, housing conditions, and the availability of social services, with the ultimate goal of offering a comprehensive understanding of how poverty affects both individuals and communities. By this analysis we can determine the main cause for poverty and we can Analyse the regional and global disparities in poverty. To achieve this, we use a variety of ML algorithms, including random forest, naïve bayes, KNN, K-means, Logistic, Decision tree to analyse and model datasets. The main things to move the analysis include data pre-processing, handling missing values, dimensionality reduction, model selection, data integration, tuning. The project which we are doing is to find out the country's which as more poverty rate.

Going on to our results of this project, we focused on the factors like reporting levels, based on national or urban or rural levels poverty. This predictive model has the potential to check the countries that which part has more poverty based on reporting levels. By leveraging historical data and advanced ML techniques, we aim to provide a valuable tool to improve searching of poverty by different parameters. We are using many different ML algorithms to find out the poverty-related countries.

2.INTRODUCTION:

Precise targeting stands as one of the fundamental pillars of a successful and efficient food security or social safety net intervention. To achieve accurate targeting, project implementers seek to minimize rates of leakage and undercoverage [1]. Being able to exactly identify poor areas is the underlying concept of poverty mitigation. The spatiotemporal analysis of poverty is conducive to the formulation of regional policies for the purpose of poverty-alleviation, Traditionally, the data sources for poverty assessment are the census or household surveys collected by local governments or national organizations [2]. The fact that it does not predict any increases in average local wealth on the country level is concerning given the clear trend in economic development as one of Africa's fastest-growing economies [3]. The absolute poverty line signifies the expenditure required to meet the essential needs for both food and non-food items at a minimum level. Consequently, this metric is rooted in expenditures rather than income. This means that a household is considered poor if it spends less than a certain amount per month on food and non-food products [4]. The incapacity to achieve essential capabilities due to insufficient access to affordable, dependable, and secure energy services is considered, while also considering the presence of viable and reasonable alternative means to attain these capabilities [5].

Poverty is a worldwide issue that acts as a barrier to achieving sustainable development. Eradicating poverty is a global goal and one of the greatest challenges for developing countries. Measuring and monitoring poverty are essential for governments to help prevent poverty traps and promote resource reallocation [6]. The common way in which energy poverty manifests itself is the lack of access to modern energy services in many developing countries, notably in sub-Saharan Africa [7]. Monitoring the advancement of poverty reduction is essential for achieving the goal of eradicating poverty within the planned timeframe and implementing effective policy measures. We focused on village-level poverty identification because it enables an understanding of poverty with fine spatial granularity, and data can also be conveniently aggregated from different sources [8]. In this paper, we employ the term "energy poverty" to denote the sufficient provision of energy services within residential settings, particularly in the context of the ongoing energy transition. Apart from research conducted in the United Kingdom and Ireland, the bulk of energy poverty studies have predominantly concentrated on developing nations. In these regions, a significant portion of the population faces challenges in accessing modern energy services. Additionally, it's worth noting that agricultural inputs play a pivotal role in influencing crop yields. These inputs encompass critical elements like irrigation, fertilization, and effective field management practices. Irrigation and fertilizer are vital agricultural inputs, and many studies have focused on their relationship with poverty [10].

Poverty alleviation remains a daunting challenge for humanity and one of the sustainable development goals. The photovoltaic poverty alleviation project (PPAP) not only subsidizes the energy consumption of residents but also plays a vital role in improving local economic income and reducing carbon emissions [11]. In computer science, there are also challenges related to reproducibility, insufficient specification of the versioning of the libraries or frameworks used, lack of availability of codes, execution errors, discrepancies between GPU floating point numbers, and incompatibility between alleged and presented results [12].Policy efforts and incentives to eradicate energy poverty may leave extreme energy-poor households behind because that requires some additional incentives and effective support programs. Failing to identify individuals who are situated at the lowest rungs of socioeconomic status deprives them of the additional incentives and financial assistance necessary to improve their circumstances. These individuals may possess distinct characteristics when compared to others in higher income brackets to moderate energy-poor and their deprivations may be more chronic deprivations comparatively [13]. There exists a compelling necessity to gauge the well-being of both individuals and nations. This involves the systematic counting, measurement, assessment, and evaluation of the global population. However, beyond these seemingly stark statements lies a more compassionate perspective, wherein the act of measuring and counting individuals serves as a crucial foundation for humanitarian and developmental endeavours. It aids in precise targeting, comprehensive mapping, and ongoing monitoring of people to advance and ensure their welfare risk of food insecurity, famine, poverty and disease [14]. Research that delves into the impact of poverty in middle-income nations has revealed that these countries are susceptible to substantial effects from poverty-related issues. Consequently, it becomes imperative to devise methodologies that emphasize the prediction and analysis of poverty in developing countries. Predicting poverty comes with its set of complications [15].

3.LITERATURE SURVEY

3.1. Survey Details:

McBride, et al. (2015) has done that the current popular methods for creating PMT tools focus on minimizing prediction errors within the sample used for development [1]. Yin et al. (2020). The authors conducted a study focused on poverty identification and analysis in Guizhou Province, China, from 2012 to 2019. They employed remote sensing data, including nighttime lights imagery and geographical data, to extract relevant spatial features and used the random forest machine learning method for poverty identification [2]. Kondmann, et al. (2020). the authors address the crucial issue of obtaining timely and accurate data on local wealth to effectively combat poverty on a global scale. They recognize that in many countries, such data is either unavailable or outdated [3]. Alsharkawi, et al. (2021). The authors address the issue of multidimensional poverty in Jordan and propose an innovative machine learning approach to assess and monitor the poverty status of Jordanian households [4]. Wang, et al. (2021). the authors address the challenge of identifying and targeting energy poverty in developing countries, particularly in rural areas. They propose a novel approach that combines satellite remote sensing data with socioeconomic survey data to predict and address energy poverty [5].

Li, et al. (2021). The author focuses on enriching the Asian perspective on rapidly identifying urban poverty and understanding its implications for housing inequality, with a specific focus on China [6]. Dalla Longa, et al. (2021). The authors address the issue of energy poverty in developed countries, with a specific focus on the Netherlands. They introduce a framework for categorizing energy poverty risk and propose the use of machine learning techniques to predict energy poverty risk based on various socio-economic parameters [7]. Hu, et al. (2022). The authors address the crucial challenge of tracking progress in poverty alleviation and identifying the distribution of poor areas, especially in regions with limited statistical data. They propose an innovative approach that integrates various sources of geospatial data to identify village-level poverty [8]. van Hove, et al. (2022). the authors address the issue of energy poverty in Europe and aim to identify its drivers using machine learning techniques. They emphasize the importance of establishing predictors for energy poverty that are valid across countries, as this could provide a foundation for effective policy measures targeting energy-poor households [9]. Tian, et al. (2022) the authors investigate the relationship between arable land use and poverty reduction, with a focus on how arable land relates to Sustainable Development Goal 1 (SDG1) of eradicating extreme poverty and SDG2 of eliminating hunger. They utilize a variety of indicators related to agricultural inputs, crop intensification, and extensification, and employ non-parametric machine learning methods for their analysis [10].

Yin, et al. (2022). the authors focus on evaluating the performance of the Photovoltaic Poverty Alleviation Project (PPAP) in China, specifically in Jinzhai County, Anhui province. They employ machine learning techniques and satellite imagery to assess the impact and effectiveness of the project [11]. Machicao, et al. 2022). The authors address the challenges of reproducibility and replicability (R & R) in computer science experiments, with a specific focus on experiments involving Deep Learning (DL) techniques. The complexity of DL methods often makes reproducing experiments difficult. The paper conducts a review of the reproducibility of three DL experiments that analyse visual indicators from satellite and street imagery [12]. Abbas, et al. (2022). this research study advances our understanding of energy poverty in developing countries by adopting a multidimensional approach and utilizing machine learning techniques. It provides valuable insights into the extent of energy poverty and the key socioeconomic factors contributing to its severity, offering a foundation for evidence-based policy measures to alleviate extreme energy poverty [13]. Hall, et al. (2023). The authors examine the emerging field of using satellite imagery and machine learning to predict and analyse poverty and welfare indicators. Their analysis of the literature in this rapidly growing field yields several important findings and implications [14]. Satapathy, et al. (2023). The authors address the critical issue of poverty using machine learning techniques and multidimensional poverty index data. They focus on predicting multidimensional poverty before and during the COVID-19 pandemic, leveraging data from the Oxford Poverty and Human Development Initiative [15].

Table 1

AUTHORS & YEAR	MODEL USED	PARAMETERS USED	MERITS	DEMERITS	FUTURE SCOPE
McBride, (2015) [1]	Regression and Quantile Regression Forests	Importance of Accurate Targeting, Full Means Tests, Proxy Means Tests (PMTs), PMT Tool Development,	Random Sampling, Random Forest Models, Out-of-Bag Estimates, Bootstrap Testing, Trade-Off Analysis	Sampling Size and Representativeness, Sampling with Replacement, Bias in Model Evaluation, Limited Data Split Approach, Bootstrap Sample Size	The future scope for proxy means tests in food security and social safety net interventions is likely to involve a combination of technological advancements, data innovation, and policy evolution.
Yin, (2020) [2]	Random Forest machine learning method	Research Objective, Data Source, Feature Extraction, Accuracy Metrics	Comprehensive Approach, Feature Extraction, Machine Learning Expertise, Accuracy Comparison	Limited Generalizability, Data Quality and Availability, Model Overfitting, Algorithmic Bias, Extrapolation Risk	A wide range of potential research directions and areas of development for the study focused on poverty identification and analysis in Guizhou, China, using remote sensing, machine learning, and geographical indicators.
Kondmann, (2020) [3]	Regression Model	Data Collection & Preparation, Feature engineering, Model Selection Hyperparameter tuning	Accuracy, Reduced Misdiagnosis, Public Health, Impact Early Detection Accuracy	Lack of Epidemiological Data – works on only symptoms, Limited Comparison to Image-Based Methods, Model sensibility & scalability	The future study can enhance its ability to accurately capture trends in economic development over time and provide valuable insights for addressing global poverty through data-driven interventions and policies.
Alsharkawi,	Decision Trees	Algorithm Evaluation, Representative Dataset, Data Preparation, Addressing	Multidimensional Perspective, Data Utilization, User-Friendly	Data Availability and Quality, Model Generalization, Changing Sociopolitical	Future work can contribute to a more comprehensive understanding of poverty in Jordan and provide valuable tools and

(2021) [4]		Class Imbalance, Robustness and Changes Over Time	Model, Algorithmic Evaluation, Robustness to Changes	Context, Scalability and Accessibility	insights for policymakers and practitioners working towards poverty reduction and economic development in the region.
Wang, (2021) [5]	Ensemble Classification KNN Statistical Naïve Bayes	Feature Selection Cross-Validation Evaluation Metrics.	Robustness Enhanced Generalization Feature Selection	Cost and Resources Complexity Model Selection Overfitting	Future research work can contribute to a deeper understanding of the complex relationship between environmental factors and energy poverty and provide valuable insights for designing sustainable energy access solutions in vulnerable communities.
Li, (2021) [6]	Random Forest, SVM, gradient boost	Data Set Data Preprocessing Collaboration with Domain Experts Ethical Considerations	Early Detection High Accuracy Resource Allocation Global Reach XAI interpretability	Data Quality Availability of Images Risk of misdiagnosis Model Drift Regulatory hurdles	Future can continue to contribute to the understanding of urban poverty in Chinese cities and serve as a valuable resource for policymakers and practitioners seeking to address housing inequality and promote equitable urban development.
Dalla Longa, (2021) [7]	Decision-tree algorithm	Datasets, Socio-economic Features, Geographical Scale, Statistical Analysis	Data-Driven Insights, Handling Complex Data, Non-linear Dependency, Ease of Exploration	Statistical Limitations, Ethical and Legal Implications, Generalization to Other Countries	Future research work can be contribute to a more comprehensive understanding of energy poverty and provide valuable tools and insights for policymakers, energy providers, and communities to address this important issue effectively.
Hu, (2022) [8]	Random forest	Selection of Explanatory Variables, Generation of Explanatory Variables from Data, Modelling with Random Forest Algorithm,	Comprehensive Data Integration, Contextualized Explanatory Variables, Remote Sensing and Geospatial Techniques, Accessibility Consideration	Data Availability and Quality, Feature Selection, Algorithm Sensitivity, Class Imba	Future it can be continue to contribute to the advancement of poverty assessment methodologies and provide valuable insights for evidence-based policy decisions and targeted interventions to alleviate poverty at the village level.
van Hove, (2022) [9]	Energy Poverty Framework	Data-driven Approach, Scope of Research, Pan-European Predictors, Universal and Contextual Predictors	Complex Relationship Discovery, Efficiency in Data Analysis, Previous Success in the Netherlands, Cross-Country Validity	Absence of Cross-Country Indicator, Challenges in Cross-Country Comparis, Complexity of Energy Poverty Factors, Debate over Universal Predictors	Future research, our work can continue to inform evidence-based policy decisions, contribute to the reduction of energy poverty, and support a just and equitable energy transition in Europe and beyond.
Tian, (2022) [10]	Random forest, KNN algorithm,	Hypothesized linkage with poverty in Crop intensification, Crop extensification	RAPY was low in developing countries and high in developed countries.	High RAPY is associated with a low poverty rate, but high RAPC does not necessarily result in low poverty rates.	To inform evidence-based policies and interventions that promote agricultural productivity, reduce poverty, and contribute to achieving broader sustainability goals on a global scale.
Yin, (2022) [11]	Random Forest SVM Logistic Regression	Choice of Pretrained Model Fine Tuning Strategy Learning rate Deployment	Faster Deployment Scalability Continuous Learning Robustness	Domain Adaption, Data quality Cost, Ethical Concerns Deployment Challenges	In future,our research can contribute to advancing the use of machine learning in renewable energy planning, optimizing PV system performance, and promoting sustainable energy solutions. Additionally, it can provide valuable insights for addressing climate change and achieving renewable energy targets.
Machicao, (2022) [12]	Random Forest Regressor	Reproducibility and Replicability (R & R)	Complex Task Accomplishment, Innovative Approach, Relevance to Research Practices	Complexity and Reproducibility, Data Quality and Generalization, Resource Intensive	Future, the scientific community can make significant strides in ensuring that DL experiments are reproducible, transparent, and valuable contributions to scientific knowledge. This, in turn, will enhance the credibility and impact of DL research across various domains.
Abbas, (2022) [13]	SVM, classification	Multidimensional Energy Poverty Index, Geographical Scope, Socioeconomic Variables	Multidimensional Approach, Data Analysis Across Developing Countries, Identification of Extreme Energy Poverty	Data Availability and Quality, Interpretability, Policy Implementation Challenges	The future scope of this research lies in advancing our understanding of the complex challenges posed by severe energy poverty and in developing evidence-based strategies to alleviate it. By combining multidisciplinary approaches, innovative data analytics, and policy advocacy, researchers can contribute to meaningful progress in achieving energy access for all.
Ola Hall, (2023) [14]	Regression Model	Training, Evaluation, Welfare Indicators, Performance Metrics, Data Integration, Temporal Analysis, Transferability	High Accuracy, Novel Intersection of Technologies, Data-Driven Insights, Systematic Review Approach, Spatial Resolution Insights	Data Quality and Availability, Sampling Bias, Heterogeneous Data, Model Generalization, Data Preprocessing Variability, Limited Temporal Analysis	Future directions, researchers can further advance the field of AI-driven poverty analysis and contribute to evidence-based policymaking aimed at reducing poverty and improving the well-being of vulnerable populations.

Satapathy, (2023) [15]	Multiple Linear Regression, Decision Tree Regressor	Multi-dimensional Poverty Index Data, Poverty Indicators, Data Analysis Techniques	Practical Application, Multidimensional Approach, Temporal Analysis.	Data Quality, Generalization, Assumptions.	Future scope areas, researchers can advance the field of poverty prediction and contribute to evidence-based strategies for poverty reduction and sustainable development, aligning with the United Nations' Sustainable Development.

3.2. PROBLEM STATEMENT:

Our goal is to predict the reporting level on countries. That means we are finding the country poverty based on national or urban or rural level's.

3.3. OBJECTIVE:

- To know the poverty ratio and poverty rate across the countries and to thoroughly investigate the multifaceted dimensions of poverty, encompassing income levels, educational opportunities, healthcare accessibility, housing conditions, and the availability of social services, with the ultimate goal of offering a comprehensive understanding of how poverty affects both individuals and communities.
- By this analysis we can determine the main cause for poverty and we can Analyse the regional and global disparities in poverty.
- This research paper endeavours to enhance our comprehension of poverty while also making a meaningful contribution to the continuous endeavours aimed at mitigating poverty and fostering an inclusive environment for economic and social development.
- Understand poverty ratio and poverty rate across counties. Availability of social services
- Gain a comprehensive understanding of how poverty affects individuals and communities.
- Identify the main causes of poverty. Analyse regional and global disparities in poverty.
- Contribute to efforts aimed at mitigating poverty.
- Foster an inclusive environment for economic and social development.

4.Proposed Work

4.1. Proposed Architecture:

In this, at first data set to be read and after that the null values in the data set will be replaced by 0's. The data must be pre-processed and cleaned. The data must be split into training and testing sections. A suitable model must be selected to evaluate the dataset and to predict values. Thereafter, predicted values must be compared to original values to find accuracy of the model.

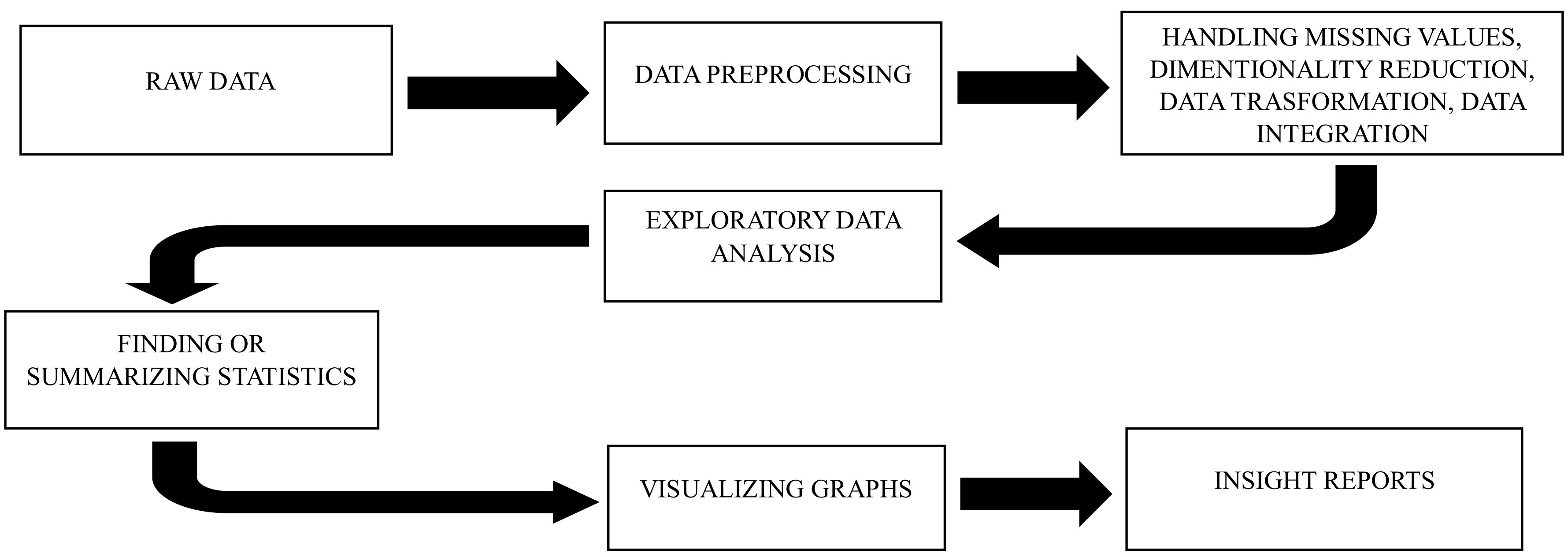

Fig-1: Representing the process of poverty Analysis

Collect data:

The first step in doing analysis using machine learning algorithms is to gather the data related to poverty analysis. The data should be more accurate to do analysis. This includes historical data on poverty outbreaks, economic data of an individual and countries, population data, information on health infrastructure. Sometimes data and its parameters depend on what bases we are doing analysis.

Data Exploration:

We Explore data to find relationships between variables by examining correlations, cross-tabulations, and other relevant associations related to data analysis. This can involve calculating correlation coefficients or creating pivot tables for categorical variables. Finding the relationships between variables enhances the view of data which leads to good analysis and prediction.

Feature Engineering:

Feature engineering is a crucial process in the field of machine learning and data analysis. It entails the careful selection of data and transformation of variables or features from the raw data to create a dataset that is well-suited for training machine learning models. This step is pivotal because the quality of the features directly impacts the model's performance and its ability to extract meaningful patterns from the data. Feature Engineering plays a key role in data analysis as it helps in cleaning the data which leads to a good analysis.

Exploratory data analysis:

EDA serves as a pivotal initial step in the data analysis process. By utilizing these techniques, analysts can gain valuable insights, detect outliers or anomalies, and make informed decisions about how to proceed with more advanced data modelling or in-depth analyses. It's a crucial phase in extracting meaningful information and knowledge from datasets.

Pattern Recognition:

Look for recurring patterns or trends in the data that might provide insights into the underlying processes or phenomena. It helps in data analysis to recognize and find the patterns, amount of relationship between variables in data.

Data Visualization:

Create visual representations of the data using charts, graphs, and plots. Common visualizations include histograms, box plots, scatter plots, bar charts, and line graphs. Visualization helps in understanding data distributions, identifying patterns, and spotting outliers.

4.2. DATASET DESCRIPTION

The dataset for the Poverty Analysis is taken from the Kaggle. This dataset is used to analysis the poverty rate across the countries and to thoroughly investigate the multifaceted dimensions of poverty, encompassing income levels, educational opportunities, healthcare accessibility, housing conditions, and the availability of social services, with the ultimate goal of offering a comprehensive understanding of how poverty affects both individuals and communities.

Table 2

Attribute Name	Count	Mean	Std	Min	25%	50%	75%	Max
year	4877.000000	2005.759893	9.438782	1967.000000	2000.000000	2007.000000	2013.000000	2021.000000
ppp_version	4877.000000	2013.999385	3.000308	2011.000000	2011.000000	2011.000000	2017.000000	2017.000000
survey_year	4411.000000	2005.939896	9.497211	1967.000000	2000.000000	2007.000000	2014.000000	2021.000000
survey_comparability	4411.000000	1.639084	1.402898	0.000000	1.000000	1.000000	2.000000	6.000000
headcount_ratio_international_povline	4877.000000	11.081565	18.190433	0.000000	0.278387	2.041412	13.390963	96.871427
headcount_ratio_lower_mid_income_povline	4877.000000	21.672687	27.104145	0.000000	0.807877	9.186717	34.166712	99.999000
headcount_ratio_upper_mid_income_povline	4877.000000	36.538818	33.674140	0.000000	3.032081	28.582685	64.864719	99.999000
headcount_ratio_100	4877.000000	3.164721	7.568626	0.000000	0.044772	0.371755	2.158480	79.532619
headcount_ratio_1000	4877.000000	50.351520	36.002943	0.000000	10.328433	55.337658	85.133887	99.999000
headcount_ratio_2000…	4877.000000…	69.105964…	34.132391…	0.920154…	44.575897…	84.916259…	97.179433…	100.000000…
decile8_thr	4877.000000	26.273736	24.171141	1.010000	8.680000	16.980000	35.550000	120.100000
decile9_thr	4877.000000	35.110607	30.671393	1.460000	11.960000	24.560000	48.350000	164.700000
Gini	4401.000000	0.375645	0.088840	0.177920	0.308719	0.355622	0.427676	0.657556

mld	4401.000000	0.264371	0.140325	0.053563	0.163090	0.220968	0.320859	0.937047
polarization	4401.000000	0.327385	0.101136	0.146643	0.252448	0.300364	0.380201	0.815704
palma_ratio	4395.000000	1.886528	1.138502	0.596408	1.154071	1.465800	2.152142	8.343586
s80_s20_ratio	4395.000000	8.219764	6.018946	2.430173	4.720943	6.171947	9.043292	72.681737
p90_p10_ratio	4877.000000	7.424451	42.891943	2.191176	3.937500	5.149826	7.458333	2892.000000
p90_p50_ratio	4877.000000	2.532256	0.954387	1.475248	2.017751	2.290076	2.780911	11.491525
p50_p10_ratio	4877.000000	2.660335	11.785204	1.485294	1.942113	2.207904	2.740385	809.000000

PROCEDURE TO SOLVE THE PROBLEM:

Poverty analysis can be done in many different ways using different models. We can use models like Logistic Regression, Naïve Bais, Decision Tree, Random Forest, KNN, XG Boost. Among these models Random Forest and Decision Tree gives the accurate results. Where as XG Boost may overfit the model.

Random Forest:

Random decision trees and random forests are ensemble learning techniques used for various machine learning tasks, including classification and regression. These methods operate by creating numerous decision trees during the training phase and then aggregating their predictions to make a final prediction.

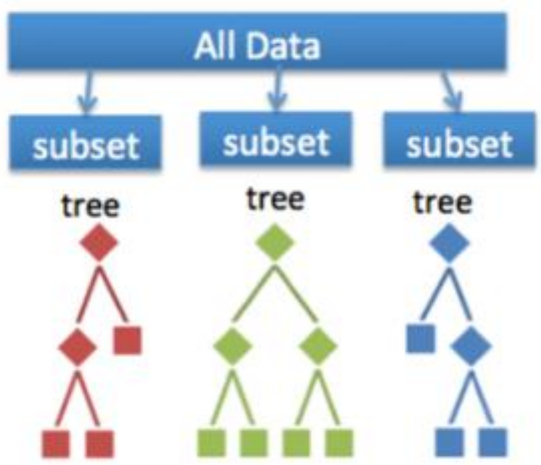

Fig-2

A random forest is a meta-estimator that fits a number of trees on various subsamples of data sets and then uses an average to improve the accuracy in the model's predictive nature. The sub-sample size is always the same as that of the original input size but the samples are often drawn with replacements.

XG BOOST

In this algorithm, a sequential approach is used to create decision trees, and the concept of weights plays a crucial role in XGBoost (Extreme Gradient Boosting). Each independent variable is assigned a weight, and these weighted variables are used as input to the decision trees for making predictions.

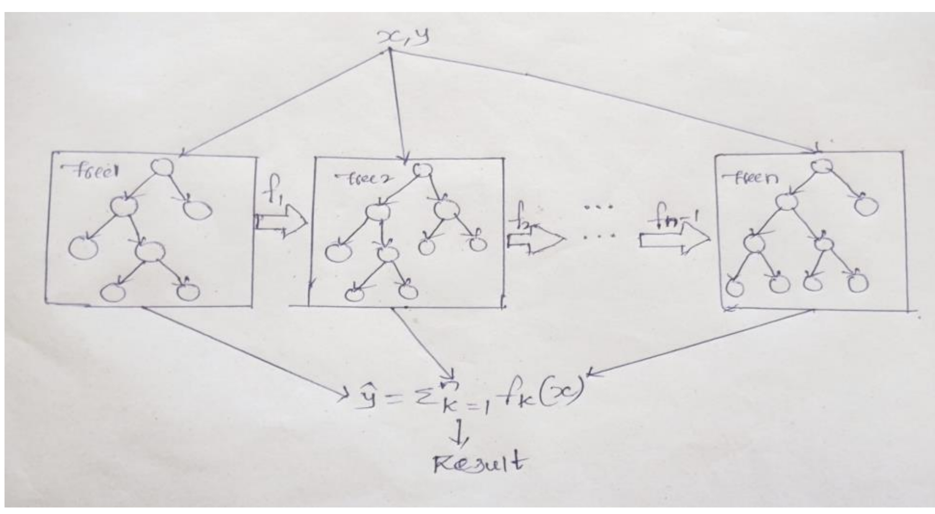

Fig-3

The unique aspect of XG Boost is its ability to adapt and learn from its mistakes. When a decision tree makes incorrect predictions for certain variables, the weights of those variables are adjusted or increased. These adjusted variables are then used as input for a second decision tree. This iterative process continues, where each subsequent decision tree focuses on correcting the mistakes of the previous ones.

The individual decision trees, or classifiers, generated in this manner, come together to form an ensemble model. This ensemble approach combines the predictions of all the individual classifiers to create a more robust and accurate model.

XG Boost is a versatile algorithm that can be applied to various types of machines learning tasks, including regression, classification, ranking, and even user-defined prediction problems. Its adaptability, thanks to the weighting and ensemble techniques, makes it a powerful tool for improving predictive accuracy across a wide range of applications.

5.Experimental Work

In the experimental work of poverty analysis project, we mainly focused on the data collection, dataset normalization, Exploratory data analysis and the basic key parameters used for predicting the countries.it provides essential for advancing scientific fields and making evidence-based decisions in various domains.

5.1. Experimental Setup

Our experimental setup includes the following key elements:

Data Collection

We collected the dataset from the dataset administrator. The dataset consists of many unknown parameters we understand it by codebook.

Data Pre-processing

Basically, Information preprocessing includes the basic assignment of planning crude information to form it reasonable for utilization in machine learning models. Our Real-world information regularly contains different issues, counting commotion, lost values, and groups that are unacceptable for coordinate utilize in machine learning calculations. As a result, information preprocessing could be a vital step that not as it were cleans and organizes the information but moreover improves the precision and proficiency of machine learning models. This beginning and principal organize is indispensable when creating machine learning models.

Exploratory data analysis

EDA serves as a pivotal initial step in the data analysis process. By utilizing these techniques, analysts can gain valuable insights, detect outliers or anomalies, and make informed decisions about how to proceed with more advanced data modelling or in-depth analyses. It's a crucial phase in extracting meaningful information and knowledge from datasets.

Data Visualization

Data visualization is an important part of the machine learning (ML) process. It helps data scientists and analysts discover, understand, and communicate patterns, relationships, and insights in data. Interactive visualization libraries such as Matplotlib, Seaborn, Plotly, and libraries for domain-specific visualizations are commonly used in ML to create visualizations. Visual images are informative and insightful.

5.2. PERFORMANCE PARAMETERS:

To survey the prescient model's execution, we utilized the taking after key execution metrics:

ACCURACY (ACC):

This metric gages the in general rightness of our model's forecasts, evaluating the ratio of precisely anticipated Building Arrangement cases to the entire cases.

ACC = (TP + TN) / (TP + TN + FP + FN)

TP -> Genuine Positive TN -> Genuine Negative FP -> Untrue Positive FN -> Wrong Negative

PRECISION (P):

Exactness may be a significant assessment metric in machine learning, particularly in classification errands like Designing Situation Expectation. It measures the model's capability to form precise positive expectations, particularly, the extent of positive expectations that were redress. Exactness is profitable when minimizing false positives is basic.

P = TP / (TP + FP)

RECALL (R):

Review, moreover known as Affectability or True Positive Rate, could be a principal evaluation metric in machine learning, particularly for classification assignments like Building Situation Expectation. Review surveys a model's capacity to precisely recognize all positive occasions in a dataset, which is especially profitable when maintaining a strategic distance from wrong negatives is basic.

R = TP / (TP + FN)

F1 SCORE:

The F1 Score, too alluded to as the F1 Degree or F1 Score, may be a broadly utilized assessment metric in machine learning, particularly for classification errands like Building Arrangement Expectation. It amalgamates exactness and review into a single metric, giving a adjusted measure of a model's performance, particularly when dealing with imbalanced datasets. The F1 Score is particularly advantageous when there's a have to be strike a adjust between precision and review, as these measurements frequently have a trade-off relationship.

F1 score = 2 * (accuracy * review) / (accuracy + review)

6.RESULTS:

Table 3

DATASET	MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
CSV DATASET	LOGISTIC REGRESSION	80%	92%	0.87	0.89
CSV DATASET	NAVIE BAYES	14%	0.84	0.08	0.15
CSV DATASET	KNN	98%	0.95	0.99	0.97
CSV DATASET	RANDOM FORSET	99%	0.99	1.00	0.88
CSV DATASET	DECISION TREE	98%	0.99	0.99	0.99
CSV DATASET	XG BOOST	1.0	1.0	1.0	1.0

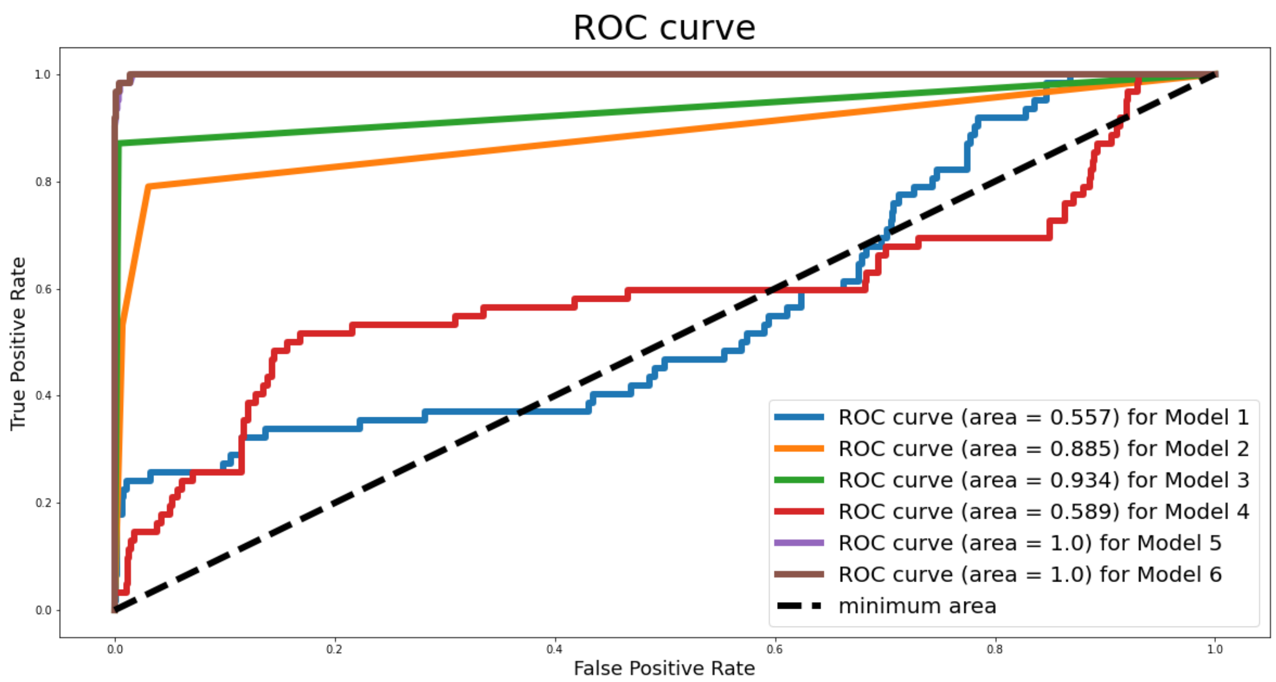

Fig-4 ROC CURVE REPRESENTING MODEL PERFORMANCES

After developing as well as evaluating all the models and their corresponding results through csv file, we can clearly observe that "XG Boost & Random Forest" hold the top position in maintaining the most accurate levels at their level of performance. Hence, XG Boost fits very well to the model with an accuracy of 1.0 which is **1.0.**

7. CONCLUSION:

Based on our comprehensive analysis of various poverty prediction models, the Random Forest model has emerged as the most accurate performer. This finding suggests that Random Forest is a robust and dependable algorithm for addressing poverty-related issues. In our research we used XG Boost model and it get overfit to the data, so the accurate model is random forest.The superior accuracy of the Random Forest model demonstrates its capacity to capture intricate data relationships and provide more precise poverty level predictions. This attribute can be immensely valuable in guiding policy decisions, optimizing resource allocation, and strategizing interventions aimed at mitigating poverty and its associated hardships. Nevertheless, it is crucial to acknowledge that accuracy is just one facet to consider when selecting an appropriate model. Factors like interpretability, computational resources, and model complexity must also be weighed in the context of the specific application. Furthermore, sustaining the model's effectiveness over time necessitates ongoing data collection, rigorous model validation, and refinement efforts. In conclusion, while Random Forest stands out in terms of accuracy, a holistic approach that encompasses various considerations is essential for the successful implementation of poverty alleviation strategies.

FUTURE SCOPE:

Future scope of our research work is we can use DL model's like GNN or other XAI model. We can also use the highly developed model in DL in future.Future poverty analysis will likely involve multidimensional frameworks that consider factors like health, education, access to basic services, and social inclusion. Developing robust multidimensional poverty indices will be a key area of research. With the increasing availability of big data and advances in technology, future poverty analysis can benefit from more granular and real-time data. Machine learning and data analytics techniques can help identify trends and patterns in poverty dynamics. Developing predictive models to forecast poverty trends can be a powerful tool for policymakers. Machine learning algorithms can analyse historical data to predict future poverty rates and identify potential interventions.

LIMITATIONS:

1. We haven't used DL model's like GNN or other XAI model ,Data Availability and Quality: One of the primary limitations of poverty analysis is the dependence on data. In many regions, particularly low-income areas, data availability and quality can be limited. Inaccurate or outdated data can lead to unreliable assessments of poverty levels, hindering the accuracy of poverty analysis.

2. Measurement Issues: Another significant limitation is the choice of poverty measurement methods and thresholds. Different poverty lines and metrics may yield different estimates of poverty rates. This variability can make it challenging to compare poverty levels across regions or over time, potentially leading to inconsistent results.

3. Multidimensional Nature of Poverty: Poverty is a multifaceted issue that encompasses more than just income or consumption levels. It includes aspects like access to education, healthcare, housing, and social inclusion. Traditional poverty measures may not capture all these dimensions, resulting in an incomplete understanding of poverty.

4. Temporal Variability: Poverty is not a static condition; it can change over time due to various factors, including economic fluctuations, policy changes, and external shocks like natural disasters or economic crises. Static poverty assessments may fail to capture these dynamic changes adequately, making it difficult to formulate effective poverty reduction strategies.

8.REFERENCES:

[1] McBride, L., & Nichols, A. (2015). Improved poverty targeting through machine learning: An application to the USAID Poverty Assessment Tools. *Unpublished manuscript. Available at: http://www. econthatmatters. com/wp-content/uploads/2015/01/improvedtargeting_21jan2015. Pdf*

[2] Yin, J., Qiu, Y., & Zhang, B. (2020). Identification of poverty areas by remote sensing and machine learning: a case study in Guizhou, Southwest China. *ISPRS International Journal of Geo-Information*, *10*(1), 11.

[3] Kondmann, L., & Zhu, X. X. (2020). Measuring changes in poverty with deep learning and satellite imagery.

[4] Alsharkawi, A., Al-Fetyan, M., Dawas, M., Saadeh, H., & Alyaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability*, *13*(3), 1412.

[5] Wang, H., Maruejols, L., & Yu, X. (2021). Predicting energy poverty with combinations of remote-sensing and socioeconomic survey data in India: evidence from machine learning. *Energy Economics*, *102*, 105510.

[6] Li, G., Cai, Z., Qian, Y., & Chen, F. (2021). Identifying urban poverty using high-resolution satellite imagery and machine learning approaches: Implications for housing inequality. *Land*, *10*(6), 648.

[7] Dalla Longa, F., Sweerts, B., & van der Zwaan, B. (2021). Exploring the complex origins of energy poverty in The Netherlands with machine learning. *Energy Policy*, *156*, 112373.

[8] Hu, S., Ge, Y., Liu, M., Ren, Z., & Zhang, X. (2022). Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *International Journal of Applied Earth Observation and Geoinformation*, *107*, 102694.

[9] van Hove, W., Dalla Longa, F., & van der Zwaan, B. (2022). Identifying predictors for energy poverty in Europe using machine learning. *Energy and Buildings*, *264*, 112064.

[10] Tian, F., Wu, B., Zeng, H., Watmough, G. R., Zhang, M., & Li, Y. (2022). Detecting the linkage between arable land use and poverty using machine learning methods at global perspective. *Geography and Sustainability*, *3*(1), 7-20.

[11] Yin, H., & Zhou, K. (2022). Performance evaluation of China's photovoltaic poverty alleviation project using machine learning and satellite images. *Utilities Policy*, *76*, 101378.

[12] Machicao, J., Ben Abbes, A., Meneguzzi, L., Corrêa, P. L. P., Specht, A., David, R., ... & Mouillot, D. (2022). Mitigation strategies to improve reproducibility of poverty estimations from remote sensing images using deep learning. *Earth and Space Science*, *9*(8), e2022EA002379.

[13] Abbas, K., Butt, K. M., Xu, D., Ali, M., Baz, K., Kharl, S. H., & Ahmed, M. (2022). Measurements and determinants of extreme multidimensional energy poverty using machine learning. *Energy*, *251*, 123977.

[14] Hall, O., Dompae, F., Wahab, I., & Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*.

[15] Satapathy, S. K., Saravanan, S., Mishra, S., & Mohanty, S. N. (2023). A Comparative Analysis of Multidimensional COVID-19 Poverty Determinants: An Observational Machine Learning Approach. *New Generation Computing*, *41*(1), 1