Training and Evaluation of the Gathered Data about Rural Poverty Indicators in the Philippines

Aldrich E. Arellano, Ericka May B. Coronel, Noelle Joshua Fadol, Joshua Jericho L. Mendoza

College of Information Technology and Engineering, Electronics Engineering Department, De La Salle Lipa, Philippines

Aldrich_arellano@dlsl.edu.ph, ericka_may_coronel@dlsl.edu.ph, noelle_joshua_fadol@dlsl.edu.ph, joshua_jericho_mendoza@dlsl.edu.ph

Abstract— In the intricate socio-economic fabric of the Philippines, the enduring challenge of poverty hampers the nation's pursuit of comprehensive development. The rural economy, a pivotal contributor to the nation's economic landscape, encompasses agricultural activities, non-farm enterprises, and local industries, serving as the backbone of food security. This research aims to comprehensively analyze key poverty indicators in the Philippines, focusing on agricultural wages, average incomes, and total employment. Leveraging data science methodologies, the study explores relationships between variables, including agricultural wage rates and farm income, non-farm income and rural income, and total employment and total income. The research employs linear regression and Decision Tree Models for predictive and categorical analyses, respectively. The findings reveal meaningful relationships between key variables, shedding light on the intricate dynamics within the agricultural sector and their impact on overall income for farm households. While the linear regression model demonstrates its ability to capture linear trends, the Decision Tree Model excels in revealing categorical relationships and key influencing factors. The comparative analysis highlights the strengths and limitations of each model, underscoring the Decision Tree Model's superior performance in certain scenarios. Overall, this study provides a foundation for informed policymaking and targeted interventions to address poverty dynamics in the Philippines.

Keywords— *poverty dynamics, rural economy, income disparities, agricultural wages, linear regression, decision tree model*

I. INTRODUCTION

In the intricate web of the Philippine socio-economic landscape, the persistent challenge of poverty poses an obstacle, hindering the nation's path to comprehensive development. Despite commendable economic growth, poverty endures, characterized by low income, constrained employment opportunities, and limited access to essential services.

Aligning itself with the global commitment encapsulated in the Sustainable Development Goal (SDGs), the Philippines underscores its dedication to fostering a more equitable society. Embracing the principles of SDG 1: "No Poverty," the nation recognizes the importance of eradicating all facets of poverty, ensuring that every individual has equal access to the essentials necessary for a dignified life.

In the fabric of a nation's economic landscape, one of the pivotal contributors is the rural economy, wielding both the potential for prosperity and the challenges of poverty. Rooted in the vast expanses beyond urban centers, the rural economy encompasses a diverse range of agricultural activities, non-farm enterprises, and local industries. Its significance extends far beyond the idyllic scenes of farmland, encompassing the livelihoods of a substantial portion of the population.

For the Philippines, rural economies is considered as the backbone of the agricultural sector, nurturing the cultivation of crops and the rearing of livestock that form the bedrock of food security. Beyond agriculture, these economies host a spectrum of non-farm activities, from small-scale enterprises to traditional crafts, creating a web of interconnected economic activities that sustains communities.

However, while the rural economy is a wellspring of economic contributions, it also grapples with the shadows of poverty. The challenges are manifold, encompassing factors such as limited access to resources, market fluctuations, and vulnerability to climate change. As a result, a significant proportion of the rural populace contends with lower income levels, constrained employment opportunities, and disparities in access to essential services. In relation to this, understanding the dynamics of the rural economy as a contributor to both economic growth and poverty is essential for crafting effective policies.

Income disparities in the rural economy serve as a testament to the pervasive hold of poverty, as a significant portion of the population remains below the national poverty line. Limited stable employment opportunities increase vulnerability, exposing individuals and families to a state of socio-economic instability. With these, a more in-depth analysis of these indicators, including rural income, agricultural wages, non-farm income, and employment, will be undertaken to glean a comprehensive understanding of the factors contributing to the issue of poverty for targeted and effective interventions to address the root causes and foster inclusive progress..

II. REVIEW OF RELATED LITERATURE

The review of related literature is a critical component of any research endeavor, serving as the intellectual foundation upon which a study is built. This section provides a comprehensive examination of existing scholarly works, theories, and empirical studies relevant to the poverty of the Philippines, highlighting the rural or agricultural income. By delving into the wealth of knowledge already available, the proponents gain insights and establish the context within which the current study contributes.

A. Defining Poverty

Poverty is a pervasive and complex socio-financial phenomenon that continues to task countries international, affecting tens of millions of people and communities. In modern-day discourse, poverty is generally defined as a multifaceted state or circumstance marked through the insufficient financial sources that restrict individuals, families, and groups from attaining a minimum widespread of living critical for a thriving existence. This deprivation extends beyond earnings constraints, encompassing the

incapability to access fundamental requirements including decent housing, clean water, nutritious meals, and critical healthcare. The complexity of poverty is underscored by its correlation with various factors, which include race, sexual identity, sexual orientation, and constrained academic opportunities [1].

The World Bank Organization describes poverty in this way: "Poverty is hunger. Poverty is lack of shelter. Poverty is being sick and not being able to see a doctor. Poverty is not having access to school and not knowing how to read. Poverty is not having a job, is fear for the future, living one day at a time. Poverty has many faces, changing from place to place and across time, and has been described in many ways. Most often, poverty is a situation people want to escape. So, poverty is a call to action - for the poor and the wealthy alike - a call to change the world so that many more may have enough to eat, adequate shelter, access to education and health, protection from violence, and a voice in what happens in their communities" [2].

B. Determinants of Poverty

Poverty indicators are measurable and quantifiable variables that provide insights into the monetary, social, and typical well-being of people, households, or communities, aiding in the assessment of their living requirements and the volume in their deprivation. These indicators serve as critical gear for identifying, studying, and addressing poverty.

Income inequality stands as a pivotal determinant within the complex panorama of poverty. There are studies that underscore a compelling correlation among heightened earnings inequality and escalating poverty rates. The usage of the Gini coefficient, an extensively standard metric for gauging earnings distribution, continually famous that nations characterized with the aid of reported profits disparities simultaneously grapple with increased tiers of poverty [3].

The perpetuation of poverty cycles is substantially influenced by using racial and ethnic disparities, as illuminated by latest investigations led by Beech, et al. (2020). These studies emphasize the lifestyles of systemic boundaries that disproportionately affect minority agencies, resulting in an augmented susceptibility to poverty. The unequal apportionment of possibilities and assets among awesome racial and ethnic communities exacerbates pre-current financial inequalities [4].

The feminization of poverty is propelled by gender-based wage gaps, limited access to education, and restricted employment opportunities, as articulated by Duflo (2012). Kabeer (2016) have conducted contemporary studies that delve into the nuanced challenges of gender disparities, shedding light on the dynamic nature of poverty concerning gender [5] [6].

The intricate relationship between urbanization, migration, and poverty introduces a nuanced layer to poverty dynamics. Recent scholarly contributions by Glaeser, Kahn, & Rappaport (2019) and Henderson (2019) provide insights into the contemporary challenges and opportunities arising from urbanization in the context of poverty. Glaeser, Kahn, & Rappaport (2019) delve into the multifaceted aspects of urban development and its implications for economic disparities, while Henderson (2019) explores the intricate relationship between urbanization and poverty dynamics. As global trends continue to shift towards increased urbanization, a comprehensive understanding of the complex interplay between urban development and poverty becomes essential for formulating sustainable urban policies [7] [8].

C. Impacts of Poverty

Poverty, a pervasive and enduring international undertaking, extends its a long way-accomplishing implications throughout people, communities, and entire societies. The complicated courting among poverty and health has garnered extensive attention in modern research. Siddiqi et al. (2017) and Patel et al. (2018) contribute to this information, elucidating the unfavorable results of poverty on both physical and mental well-being. Individuals ensnared in poverty frequently deal with better quotes of continual illnesses, intellectual health problems, and face boundaries to having access to important healthcare offerings. This nexus between poverty and fitness underscores the urgent need for comprehensive interventions that address the health disparities perpetuated by means of financial deprivation [9] [10].

Recent investigations by using Duncan and Murnane (2014) and Hanushek and Woessmann (2015) delve into the tricky instructional influences of poverty. These studies resolve a cyclical project in which poverty acts as a barrier to great education, while restricted academic attainment, in turn, perpetuates the cycle of poverty. The persisting hole in educational effects for youngsters from impoverished backgrounds accentuates the urgency of centered educational guidelines to break the vicious cycle and foster identical opportunities for all [11] [12].

Beyond limiting get right of entry to opportunities, poverty resonates profoundly in the financial realm. Krause and Rinne (2015) and Edmonds and Schady (2012) illuminate the cyclical nature of poverty, in which limited financial resources occur as diminished employment opportunities. This cycle exacerbates financial hardships for affected people and households, reinforcing the imperative for economic guidelines that deal with the basis causes of poverty and stimulate sustainable economic boom [13] [14].

The social and mental toll exacted with the aid of poverty emerges as a poignant theme in current studies through Mani et al. (2013) and Helliwell et al. (2020). This research finds the pervasive stress and stigma associated with poverty, contributing to social isolation and an average faded sense of proper-being. Individuals grappling with poverty regularly bear the weight of shame and feelings of inadequacy, in addition impeding their capability to break unfastened from the entwining cycle of deprivation. Understanding and addressing the social and psychological dimensions of poverty are essential for designing holistic interventions that empower people to triumph over the challenges posed via financial adversity [15] [16].

The inter-generational transfer of poverty takes center level in research through Dahl and Lochner (2012) and Almond et al. (2014). These studies light up the iconic results confronted with the aid of youngsters born into impoverished environments, where the cycle of poverty is perpetuated across generations. Recognizing the long-time

period implications of poverty underscores the significance of interventions that break this inter-generational cycle, fostering opportunities for future generations to escape the clutches of financial deprivation [17] [18].

D. Poverty Dynamics in the Philippines

According to the Preliminary Results of the Family Income and Expenditure Survey (FIES) in 2021, the poverty incidence among the population, defined as the percentage of Filipinos with per capita income insufficient to meet basic food and non-food needs, stood at 18.1 percent. This equates to approximately 19.99 million Filipinos living below the poverty threshold of PhP 12,030 per month for a family of five. The subsistence incidence, representing the proportion of Filipinos with income inadequate for basic food needs, saw a slight increase to 5.9 percent in 2021. It was estimated that a family of five requires at least PhP 8,379 per month to fulfill their basic food requirements). In terms of families, approximately 3.50 million families, or 13.2 percent, were deemed poor in 2021. Conversely, the subsistence incidence among families was noted at 3.9 percent, equivalent to 1.04 million families falling below the food poverty line. Figure 1 shows this poverty and subsistence incidence statistics in the Philippines from 2015-2018 [19].
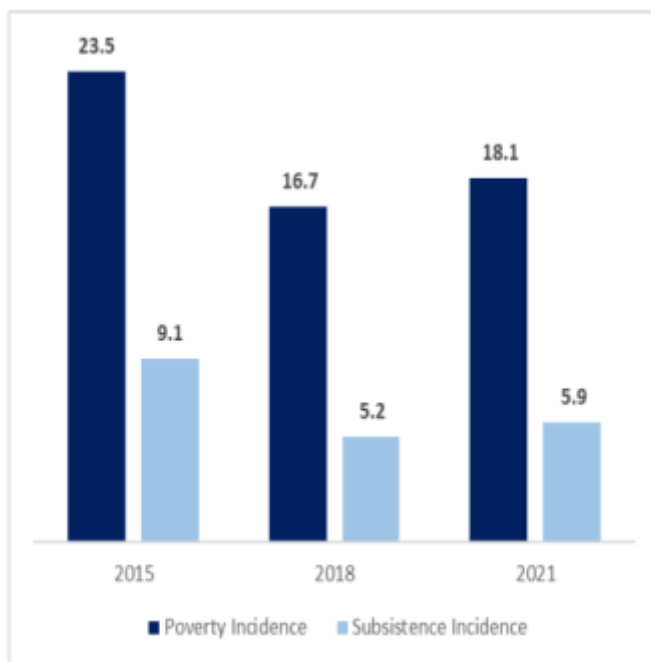

Figure 1. Poverty and Subsistence Incidence Among Population, Philippines: Full Year 2015, 2018, and 2021

Source: *Philippine Statistics Authority* [19]

According to the most recent official data released by the Philippine Statistics Authority (PSA), a greater number of Filipinos find themselves in impoverished conditions today compared to 2018. The PSA, drawing on its Family Income and Expenditure Survey, reported that 19.99 million individuals, constituting 18.1 percent of the population, are living below the poverty threshold. In 2018, the count stood at 17.67 million impoverished Filipinos. Concurrently, there has been a rise of 1.01 million in the number of individuals categorized as "food-poor." Additionally, the PSA highlighted a 7.8 percent unemployment rate, translating to 3.71 million Filipinos currently without employment [19].

It is worth noting that the poverty scenario may be even more severe, given that the PSA conducted its survey in 2021, and some experts consider its poverty threshold unrealistic. As per the Philippine government's perspective, an individual from a family of five members is deemed to require only $1.41 per day, when converted into U.S. dollars, to sustain themselves and fulfill their daily nutritional needs [20]. To have a better grasp of the poverty in the country, presented in Figure 2 is the trend of the poverty rate in the Philippines compared to the world poverty rate.
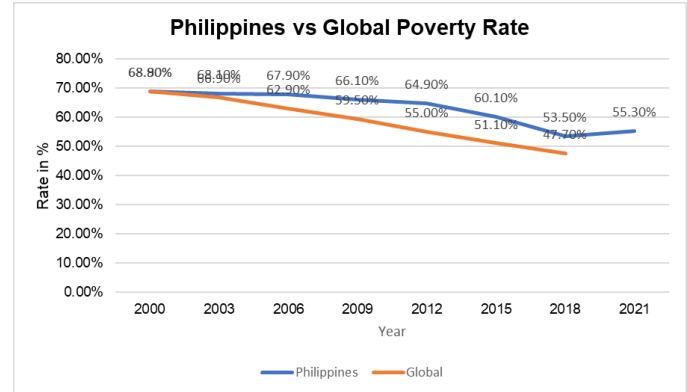

Figure 2. Philippines vs Global Poverty Rate (2000-2021)

Source: *Macrotrends* [21] [22]

Figure 2 presents the Philippine poverty rate over a span of two decades, from 2000 to 2021, along with the corresponding percentage change. In 2021, the poverty rate stands at 55.30%, indicating a 1.80% increase compared to the rate in 2018. Interestingly, the year 2018 saw a decrease in the poverty rate of 6.60% from 2015. The trend continues with a further reduction of 4.80% between 2015 and 2012. Notably, the poverty rate has demonstrated a declining pattern since 2000, with successive decreases in each recorded interval. However, comparing this to the world poverty rate, it is revealed that the country is experiencing a worse scenario as it falls below the global rate.

From 2012 to 2009, there is a marginal reduction of 1.20% in the poverty rate, and a similar trend persists between 2009 and 2006, where the rate decreases by 1.80%. The years 2006 to 2003 witness a minor decrease of 0.20%, followed by a comparable reduction of 0.70% between 2003 and 2000.

While the poverty rate in 2021 has shown a slight increase compared to 2018, the overall trajectory over the past two decades reveals a positive trend of decreasing poverty rates in the Philippines. The periodic fluctuations suggest that, although progress has been made, addressing and sustaining poverty reduction efforts remain crucial for fostering long-term socioeconomic development in the country.

The COVID-19 pandemic significantly influenced poverty levels in the Philippines. In 2020, the pandemic disrupted economic progress, resulting in a surge in unemployment rates. This pattern persisted into 2021, causing poverty rates to increase, impacting nearly 20 million individuals (World Bank). Despite ongoing economic recovery, the improvement is unevenly distributed, and the most impoverished households experience the least benefit. While many regions are returning to pre-pandemic conditions, communities in remote areas face persistent

challenges in restoring their economies, struggling to reach even the prior, already difficult, levels [23].

Amidst the difficult tapestry of monetary improvement, social disparities, and structural inequalities, poverty maintains to pose an impressive and continual challenge inside the Philippines. Drawing insights from current scholarly works, the assessment seeks to unravel the complicated elements contributing to the endurance of poverty, while simultaneously providing a glimpse into potential pathways for sustainable relief.

Reyes and Tabuga's (2015) comprehensive study explores the modern panorama of poverty in the Philippines. Unlike simplistic statistical measures, their research provides a nuanced exploration of poverty by dissecting various factors. One key focus is on income distribution, revealing the disparities that exist among different segments of the population. By examining local disparities, the study sheds light on how poverty manifests itself differently in various regions of the country. Additionally, Reyes and Tabuga scrutinize temporal changes, allowing for a dynamic understanding of poverty trends over time [24].

The contribution of a publication by Asian Development Bank (2009) adds to the richness of insights into the modern landscape of poverty in the Philippines. Going beyond conventional analyses, their research conducts thorough examinations of factors influencing poverty. The income distribution patterns was explored, revealing not just the presence of poverty but also the disparities in how economic resources are distributed among different socioeconomic groups. They were able to understand local disparities, recognizing the diversity of challenges faced by various communities. Additionally, the publication considers temporal changes, providing a dynamic perspective on how poverty evolves [25].

A study by Fabella, R. (2014) delves deeply into the transformative impact of urbanization on poverty dynamics within the Philippine context. The study emphasizes the profound changes in livelihoods accompanying the shift from rural to urban settings. By examining this transition, Fabella sheds light on the intricate interplay of factors influencing poverty in the country. The research goes beyond surface-level observations, providing a nuanced understanding of the challenges and opportunities that emerge as individuals and communities transition from rural to city life. Fabella's insights contribute significantly to the broader discourse on urbanization and its implications for poverty alleviation strategies [26].

As revealed by Orbeta (2017) there is an intricate relationship between employment styles and the dynamics of poverty. By focusing on the prevalence and implications of casual labor, he shed light on how different forms of employment significantly influence poverty, especially within the context of informal economies. Orbeta's work not only identifies the vulnerabilities associated with casual employment but also provides insights into the factors that contribute to the perpetuation of poverty in specific employment settings. The findings offer an understanding of the complexities at the intersection of employment patterns and poverty [27]. Domingo's (2019) research complements this exploration by providing a comprehensive analysis of how casual labor, prevalent in informal economies, impacts poverty dynamics. The study goes beyond mere observations

of employment trends, offering insights into the unique challenges faced by individuals engaged in casual work. Domingo's work emphasizes the need to understand the specific vulnerabilities and insecurities associated with informal employment, recognizing that these factors play a pivotal role in perpetuating poverty for a significant segment of the population [28].

E. *Income Inequality*

As previously discussed, income inequality plays a pivotal role in shaping the prevalence and persistence of poverty. Researches shed light on the nuanced interplay between the distribution of income and the prevalence of poverty, elucidating the impact of disparities in wealth on the economic well-being of individuals and communities. These investigations have consistently underscored a discernible correlation between elevated levels of income inequality and an escalation in poverty rates.

In the exploration of this intricate relationship, the Gini coefficient emerges as a vital metric, frequently employed to quantify and assess income distribution within a given society (World Bank, 2018). The Gini coefficient, a widely accepted measure of income inequality, encapsulates the extent to which income is distributed across a population. Countries exhibiting greater income disparities, as reflected in higher Gini coefficients, tend to confront elevated poverty rates. This correlation, as evidenced by the empirical findings of the World Bank (2018), emphasizes the importance of understanding and addressing income inequality as a crucial determinant of poverty.

In the Philippines, the poverty line is set at P10,481 per month for a family of five. As of 2021, the Philippine Institute for Development Studies revealed that 22 percent of Filipino households were living below this poverty line. Additionally, households earning between P10,481 and P20,962 are categorized as low income, constituting 35 percent of all households, thus making up 57 percent of the total population when combined with those living below the poverty line [29].

The lower and upper middle classes, earning between P20,962 and P125,772 per month, represent 40 percent of the population. Meanwhile, households with incomes ranging from P125,772 to P209,620 are considered upper class, comprising 2 percent of the population. A minuscule fraction, less than 1 percent or approximately 143,000 families, is classified as rich with a monthly income of P209,620 or more [29].

The prevailing income inequality poses a significant source of social tension. In a society where the majority, 57 percent, falls within the low-income and poverty categories, government resource allocation tends to prioritize this sector. Consequently, the middle class is marginalized and lacks adequate safety nets. Historical patterns suggest that societies with pronounced income inequality often witness social change driven by the discontent and tension within the middle class, as evident in calls for revolutions [29].

F. *Rural Income as Poverty indicator*

In the analysis of poverty, a factor that demands scrutiny is rural income, with a specific focus on agricultural wages. This previously discussed poverty indicator offers

understanding of economic vulnerabilities, shedding light on the dynamics of income generation within rural communities. As an aspect of the socio-economic landscape, rural income, and more specifically, agricultural wages, serves as a key metric in delineating the contours of poverty, capturing the essence of livelihoods and economic stability in regions predominantly reliant on agrarian activities.

Rural income as a poverty indicator encompasses the financial well-being of individuals engaged in diverse economic activities within rural settings. The inclusion of agricultural wages within this indicator amplifies its relevance, recognizing the predominant role that agriculture plays in the livelihoods of rural populations. Agricultural wages, reflected in the remuneration received by farm workers, unveil crucial insights into the economic conditions of those directly involved in primary agricultural production. This is especially relevant for the Philippines considering that agriculture is one of the leading sectors of the economy.

Agriculture in the Philippines constitutes a significant aspect of the country's economic landscape, with a substantial number of Filipinos residing in rural regions and relying on agricultural pursuits for their livelihoods. As revealed by Statista Research Department (2023), initial data for the year 2022 revealed that approximately a quarter of the employed population in the Philippines engages in the agricultural sector, encompassing farming, fisheries, livestock, and forestry. During the same period, this sector contributed a gross value added (GVA) of around 1.78 trillion Philippine pesos, representing approximately 8.9 percent of the nation's gross domestic product (GDP). Following a two-year contraction attributed to the challenges posed by the COVID-19 pandemic, the agricultural industry in the Philippines exhibited signs of recovery in 2022, albeit with a growth rate that remains lower than pre-pandemic levels.

Recent investigations conducted by Reyes and Santos (2015) and Balisacan et al. (2018) have significantly contributed to our understanding of the intricacies surrounding rural income in the Philippines. According to Reyes and Santos (2015), the landscape of rural economies in the Philippines has undergone substantial changes. Their study scrutinizes the trajectory of agricultural productivity, revealing shifts in production patterns and the emergence of new trends. They were able to delve into the core of rural income by In the analysis of poverty, a factor that demands scrutiny is rural income, with a specific focus on agricultural wages. This previously discussed poverty indicator offers understanding of economic vulnerabilities, shedding light on the dynamics of income generation within rural communities. As an aspect of the socio-economic landscape, rural income, and more specifically, agricultural wages, serves as a key metric in delineating the contours of poverty, capturing the essence of livelihoods and economic stability in regions predominantly reliant on agrarian activities.

Rural income as a poverty indicator encompasses the financial well-being of individuals engaged in diverse economic activities within rural settings. The inclusion of agricultural wages within this indicator amplifies its relevance, recognizing the predominant role that agriculture plays in the livelihoods of rural populations. Agricultural wages, reflected in the remuneration received by farm workers, unveil crucial insights into the economic conditions of those directly involved in primary agricultural production.

This is especially relevant for the Philippines considering that agriculture is one of the leading sectors of the economy.

Agriculture in the Philippines constitutes a significant aspect of the country's economic landscape, with a substantial number of Filipinos residing in rural regions and relying on agricultural pursuits for their livelihoods. As revealed by Statista Research Department (2023), initial data for the year 2022 revealed that approximately a quarter of the employed population in the Philippines engages in the agricultural sector, encompassing farming, fisheries, livestock, and forestry. During the same period, this sector contributed a gross value added (GVA) of around 1.78 trillion Philippine pesos, representing approximately 8.9 percent of the nation's gross domestic product (GDP). Following a two-year contraction attributed to the challenges posed by the COVID-19 pandemic, the agricultural industry in the Philippines exhibited signs of recovery in 2022, albeit with a growth rate that remains lower than pre-pandemic levels [30].

According to Reyes and Santos (2015), the landscape of rural economies in the Philippines has undergone substantial changes. Their study scrutinizes the trajectory of agricultural productivity, revealing shifts in production patterns and the emergence of new trends. They were able to delve into the core of rural income by exploring the variations in agricultural productivity. Their research not only identifies trends in crop yields but also considers the implications for the incomes of rural households. This study is particularly illuminating, showcasing the direct link between agricultural output and the economic well-being of those dependent on farming activities [31].

Balisacan et al. (2018) complements this narrative by providing a comprehensive view of the economic dynamics at play within rural settings. Both studies offer a nuanced understanding of the evolving economic scenarios in rural areas, emphasizing the need for a contextualized approach in addressing poverty. The study shed light on wage disparities within the agricultural sector. The disparities in wages, as highlighted in their work, contribute significantly to the complexities of rural income, influencing the distribution of economic resources among farm workers [32].

An essential aspect of unraveling the complexities of poverty revolves around comprehending the relationship between agricultural wages and income disparities. These dynamics play a pivotal role in shaping the economic landscape of rural areas, with significant implications for the overall well-being of households dependent on agriculture. Cruz and Diaz (2016) explore the relationship between agricultural wages and income disparities, revealing how variations in wage levels directly influence the economic conditions of rural households. Their study provides valuable insights into the details of wage structures within the agricultural sector, revealing potential disparities that impact the income distribution among farm workers. This underscores the unequal distribution of income within rural communities [33].

Furthermore, a study by Aquino, et al. (2020) revealed how fluctuations in these wages correlate with overall income variations in rural settings. Their research, based on comprehensive data analysis, uncover patterns that demonstrate the direct connection between agricultural wage levels and the economic status of households engaged in

farming activities. This finding would emphasize the pivotal role that wage levels play in determining the economic well-being of families reliant on agriculture. This demonstrates that households with lower agricultural wages face higher risks of economic deprivation, further emphasizing the urgency of addressing wage disparities [34].

G. Role of Non-Farm Income in Poverty

The role of non-farm income in poverty alleviation has garnered significant attention in recent literature, reflecting an understanding of the nature of poverty dynamics. Non-farm income has emerged as a crucial factor in poverty alleviation, challenging traditional perceptions that predominantly agrarian economies are solely reliant on farming for prosperity. According to Winters et al. (2014), the diversification of income sources beyond agriculture plays a pivotal role in enhancing household resilience and reducing vulnerability to economic shocks. The study underscores the importance of non-farm income in augmenting overall livelihoods, emphasizing its potential to break the cycle of poverty [35].

As emphasized in an article by McCulloch et al. (2003), non-agricultural income plays a crucial role in poverty alleviation and contributes significantly to enhancing household well-being. It was reported that non-farm income accounted for 30-45 percent of rural household income globally [36]. A comprehensive review by Reardon et al. (1998), covering rural household surveys from the mid-1970s to the late 1990s, was studied at the article to reveal that non-farm income constituted an average of 42 percent for Africa, 32 percent for Asia, and 40 percent for Latin America. In their study, the findings indicate that non-farm income sources constitute approximately one-third of the total income for agricultural households. Furthermore, through decomposition exercises applied to the poverty index, it is revealed that non-farm income sources play a significant role in reducing poverty among these households. Conversely, the analysis of the Gini coefficient decomposition in this study demonstrates that non-farm income sources contribute to an increase in income inequality within agricultural households, acting as a factor that induces disparity. In contrast, farm and unearned income are identified as equalizers of income. The policy implications drawn from this research suggest that fostering non-farm income activities among agricultural households is crucial for elevating their overall income and, consequently, alleviating [37].

III. GATHERED DATA / DATASET

The dataset employed in this project is a comprehensive compilation of economic and agricultural indicators, providing a detailed snapshot of the socio-economic insight across various regions in the Philippines. These indicators encompass a wide array of quantitative measures that capture the dimensions of economic and agricultural activities, shedding light on the characteristics and disparities that exist among different geographic areas within the country.

Titled "A Few Poverty Indicators for the Philippines by Region," the dataset has been sourced from Kaggle.com, with the last recorded update dating back to the year 2018. Kaggle is a renowned platform that hosts datasets and facilitates collaborative data science projects. This particular dataset serves as a valuable resource for understanding poverty dynamics in the Philippines, providing insights into various indicators across different regions of the country. The dataset is presented in Figure 3.

regDesc	agr_wage	agr_wage	agr_wage	avg_annu	avg_annu	avg_annu	avg_annu	avg_annu	avg_rural	total_emply_2016
Armm	162.89	163.65	103.81	73356	34026	5737	24936	8657	73673	1140
Bicol Region	167.99	169.95	119.52	82988	36579	7348	30381	8680	72626	2331
Cagayan Valley	228.77	232.64	199.98	145809	87353	6909	39894	11653	94212	1482
Calabarzon	230.92	231.45	172.04	143075	64498	8444	48590	21543	NA	5687
Car	206.68	211.04	195.62	127067	71142	9731	37118	9016	95635	765
Caraga	194.46	195.44	187.78	84569	54004	11118	13892	5555	63574	1086
Central Luzon	257.97	259.04	204.73	158130	85048	6073	50413	16596	125626	4363
Central Visayas	156.17	160.65	150.67	72177	28074	5574	31055	7474	68447	3234
Davao Region	168.68	169.83	158.91	101673	69496	5593	17573	9011	79983	2016
Eastern Visayas	157.49	159.25	112.22	82503	47359	6066	23812	5266	68077	1790
Ilocos Region	237.26	239.19	199.04	118637	51544	4358	36054	26681	103915	1988
Mimaropa	188.73	192.16	170.73	126237	74702	6579	35262	9694	NA	1252
Northern Mindanao	159.12	160.07	151.39	82299	49988	4256	23279	4776	73966	2004
Soccsksargen	164.77	166.75	159.14	134406	84331	9347	33658	7070	73428	1854
Western Visayas	165.28	167.77	156.96	79949	41266	6746	21772	10165	76488	3208
Zamboanga Peninsula	157.37	158.55	152.79	99327	54135	5946	33431	5815	64413	1540

Figure 3. Dataset used in the project

Within this dataset, economic indicators encompass a range of variables such as income, wages, and employment. These indicators serve as crucial metrics for assessing the financial health and productivity of each region. This regional approach acknowledges the diversity within the Philippines, recognizing that economic and agricultural dynamics can vary significantly from one region to another due to factors. Presented below are the description of the columns in the dataset:

- regDesc: Region description, indicating the name of each region in the dataset.

- agr_wage_farm_workers_allgender_2015: Average agricultural wage for farm workers of all genders in 2015.

- agr_wage_farm_workers_male_2015: Average agricultural wage for male farm workers in 2015.

- agr_wage_farm_workers_female_2015: Average agricultural wage for female farm workers in 2015.

- avg_annual_total_incm_farm_households_02_03: Average annual total income for farm households in 2002-2003.

- avg_annual_farm_incm_farm_households_02_03: Average annual farm income for farm households in 2002-2003.

- avg_annual_off_farm_incm_farm_households_02_03: Average annual off-farm income for farm households in 2002-2003.

- avg_annual_non_farm_incm_farm_households_02_03: Average annual non-farm income for farm households in 2002-2003.

- avg_annual_other_sources_incm_farm_households_02_03: Average annual income from other sources for farm households in 2002-2003.

- avg_rural_income_2000: Average rural income in the year 2000.

- total_emply_2016: Total employment in the year 2016.

The dataset is a small-sized tabular dataset, having 11 columns and 17 rows, wherein the types of variables include both numerical values in decimals (e.g., average wages, income figures) and categorical values in strings (e.g., region descriptions). The provided dataset has 11 columns to

represent information for different regions in the Philippines. The row of the dataset corresponds to each of the regions in the country excluding National Capital Region; hence, giving a total of 16 rows. These include Autonomous Region in Muslim Mindanao (ARMM), Bicol Region, Cagayan Valley, CALABARZON, Cordillera Administrative Region (CAR), Caraga, Central Luzon, Central Visayas, Davao Region, Eastern Visayas, Ilocos Region, MIMAROPA, Northern Mindanao, SOCCSKSARGEN, Western Visayas, and Zamboanga Peninsula. Presented in Figure 4 is the descriptive statistics of the dataset which include the characteristics per category, such as number of not-empty values, mean, standard deviation, minimum value, and maximum value.

	agr_wage_farm_workers_allgender_2015	agr_wage_farm_workers_male_2015	agr_wage_farm_workers_female_2015
count	14.000000	14.000000	14.000000
mean	184.635714	186.701429	160.897143
std	34.410445	34.589990	33.346374
min	156.170000	158.550000	103.810000
25%	160.062500	161.400000	150.850000
50%	166.635000	168.800000	157.935000
75%	203.625000	207.140000	193.660000
max	257.970000	259.040000	204.730000

avg_annual_total_incm_farm_households_02_03	avg_annual_farm_incm_farm_households_02_03	avg_annual_off_farm_incm_farm_households_02_03
14.000000	14.000000	14.000000
103063.571429	56738.928571	6771.571429
28617.112599	19667.102146	2006.694094
72177.000000	28074.000000	4256.000000
82350.000000	42789.250000	5629.000000
91948.000000	52774.000000	6069.500000
124959.500000	70730.500000	7238.250000
158130.000000	87353.000000	11118.000000

avg_annual_non_farm_incm_farm_households_02_03	avg_annual_other_sources_incm_farm_households_02_03	avg_rural_income_2000	total_emply_2016
14.000000	14.000000	14.000000	14.000000
29809.142857	9743.928571	81004.500000	2057.214286
9684.891500	5758.176654	17673.766625	971.735100
13892.000000	4776.000000	63574.000000	765.000000
23412.250000	6128.750000	69491.750000	1496.500000
30718.000000	8668.500000	73819.500000	1921.000000
35455.000000	9877.750000	90654.750000	2252.250000
50413.000000	26681.000000	125626.000000	4363.000000

Figure 4. Descriptive statistics of the dataset

In terms of the dataset's constraints, it covers only 16 out of the 17 regions in the country, with the National Capital Region (NCR) being the excluded region. Additionally, the dataset contains two missing values, specifically related to the average rural income for the year 2000 in Region IV, encompassing CALABARZON (IV-A) and MIMAROPA (IV-B).

Furthermore, the dataset spans different years for various variables: 2015 for gender-related data, 2002 to 2003 for household income-related data, 2000 for average rural income, and 2016 for total employment. The confined timeframe may impose limitations on the ability to observe long-term trends or assess the repercussions of economic changes over extended periods. Moreover, certain variables, like 'agr_wage_farm_workers_allgender_2015,' present a single-year snapshot, potentially missing dynamic changes that unfold over time. This temporal constraint may hinder a comprehensive understanding of evolving trends and shifts in the variables under consideration.

Link of the dataset/repository: https://www.kaggle.com/datasets/statchaitya/a-few-poverty-indicators-for-philippines-by-region/data

IV. OBJECTIVES

The primary objective of the study is to conduct a comprehensive analysis of the dataset containing key poverty indicators for the Philippines by region, leveraging data science methodologies, with the aim of having an overview of the status of poverty in each region. Specifically, this study aims to:

1. Explore and preprocess the dataset focusing on variables related to agricultural wages, average incomes, and total employment to ensure data integrity and completeness;

2. Conduct correlation analyses to identify relationships between variables, specifically the following:

    2.1 Agricultural wage rates and farm income

    2.2 Non-farm income and rural income

    2.3 Total employment and total income

3. Implement a linear regression model to analyze the relationships between the selected and specified variables and evaluate the model's ability to predict the linearity of these variables and assess their respective impacts on overall income for farm households;

4. Employ a Decision Tree Model to classify instances based on the selected variables to reveal categorical relationships and key influencing factors and evaluate the model's performance for a comprehensive understanding of farm households' income dynamics; and

5. Conduct a comparative analysis between the linear regression model and the Decision Tree Model in terms of their effectiveness in analyzing the relationships among variables to identify the strengths and limitations of each model for a more robust interpretation of farm households' income dynamics.

V. CONCEPTUAL

The conceptual framework delineates a methodological roadmap for achieving the primary objective of training and evaluating a dataset, focusing on agricultural wages, average incomes, and total employment. It entails a systematic progression through essential stages, including data exploration, cleaning, and feature selection. The subsequent focus on correlation analysis, complemented by the implementation of linear regression and a Decision Tree Model, unravel relationships between variables. Concluding with a comparative analysis, this framework guides a comprehensive understanding of the relationships by evaluating the model's strengths and limitations in analyzing relationships among variables.

In the pursuit of the main objective of training and evaluating the gathered data, a comprehensive conceptual framework is outlined below:

1. Data Exploration and Cleaning:

- Conduct an in-depth exploration of the dataset to comprehend its structure, types, and distribution characteristics.

- Address missing values, outliers, and inconsistencies through rigorous data cleaning procedures to ensure data integrity.

2. Feature Selection:

- Prioritize relevant columns related to agricultural wages, average incomes, and total employment based on their significance for the analysis.

- Perform a detailed evaluation of each feature, considering statistical measures and domain knowledge to identify and retain influential variables.

3. Correlation Analysis:

- Conduct a thorough correlation analysis to ascertain relationships between key variables.

- Focus on specific correlations, namely agricultural wage rates and farm income, non-farm income and rural income, and total employment and total income.

4. Linear Regression Model:

- Develop and implement a linear regression model to analyze relationships and assess its predictive ability.

5. Decision Tree Model:

- Create and utilize a Decision Tree Model for classification and understanding categorical relationships.

6. Comparative Analysis:

- Compare the performance of the linear regression model and the Decision Tree Model.

- Evaluate their strengths and limitations in analyzing relationships among variables.

VI. DATA AND RESULTS

In this section, two distinct models, Linear Regression and a Classification Model (Decision Tree), were employed for the examination of selected variables. The first focus involves agricultural wage rates and farm income, aiming to uncover the correlations and dynamics between the compensation of agricultural workers and the overall income of farm households. The second analysis explores the interrelations between non-farm income and rural income, providing insights into the contribution of non-farm income to the broader rural income landscape. The final pair of variables under analysis consists of total employment and total income, with the objective of revealing patterns in the relationship between overall employment levels and the average annual income of farm households. These analyses, conducted through correlation assessments, contribute to the understanding of economic dynamics in agricultural and rural contexts, scaling light on the relationships between various facets of income and employment.

A. Linear Regression Model

In the evaluation of the Linear Regression model, the primary metric used to assess its performance was the Mean Squared Error (MSE). The MSE quantified the average squared differences between the model's predictions and the actual values in the testing set. A higher MSE indicated a larger discrepancy between predicted and observed values, suggesting potential limitations in the model's ability to

accurately capture the relationships among selected variables. The coefficients generated by the linear regression model served as interpretable indicators of the direction and strength of the linear relationships between variables. The focus in the evaluation of the Linear Regression model was predominantly on assessing the accuracy of continuous predictions, making the MSE a key metric for gauging the model's effectiveness in capturing the nuanced dynamics of the chosen variables.

1) Linear Regression Results for Variables 1 and 2

In this analysis, the researchers used a Linear Regression model to explore the connection between two key variables: agricultural wage rates ('agr_wage_farm_workers_allgender_2015') and the average annual income of farm households ('avg_annual_farm_incm_farm_households_02_03').These variables were chosen to understand how the pay received by farm workers relates to the overall income of farm households. The code for this part is shown in Figure 5.

```
#Implementing linear regression model to analyze relationships between variables

# Define variables
var1 = 'agr_wage_farm_workers_allgender_2015'
var2 = 'avg_annual_farm_incm_farm_households_02_03'

# Implement linear regression to selected features
X = df[[var1]]
y = df[[var2]]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

Mean Squared Error: 116395204.90514195
```

Figure 5. Code for the analysis and evaluation of variables 1 and 2 using Linear Regression Model

The analysis began by carefully defining these variables and splitting the data into two parts: 80% for training the model and 20% for testing its predictions on new data. This division ensured a factual evaluation of how well the model could work with data it hadn't seen before. Moreover, the researchers used the Linear Regression tool from the scikit-learn library to create and train our model. After training, the model made predictions on the test set, creating anticipated values denoted as 'y_pred.' To assess the performance of the model, proponents used a metric called mean squared error (MSE), which measures the average of the squared differences between predicted and actual values in the testing set. The resulting MSE, specifically Mean Squared Error = 116395204.90514195, gives us a numerical understanding of how well the model predicts.

Interpreting the MSE, its value indicates how much the prediction of the model differs from the actual values of average annual income for farm households based on agricultural wage rates. The relatively high MSE suggests a notable level of difference, indicating a significant predictive error. This implies that the current linear regression model may not completely capture the complex relationship between agricultural wage rates and the overall income of farm households. Moreover, the insights visualized from Figure 6, a scatter plot depicting the relationship between agricultural wage rates and the average

annual income of farm households, align with and reinforce the findings indicated by the Mean Squared Error (MSE). The scatter plot visually represents the actual data points, each dot reflecting the real values of agricultural wage rates and corresponding average annual income. Furthermore, the linear line, derived from the Linear Regression model, illustrates predicted relationships. The absence of intersections between the line and more than one actual data point underscores challenges in accurately capturing data variability. The scattered arrangement of data points emphasizes the model's limitations, visually echoing the predictive error highlighted by the MSE. This visual representation serves as an additional layer of evidence, affirming the need for model refinement to better comprehend the relationship between agricultural wage rates and farm household income.
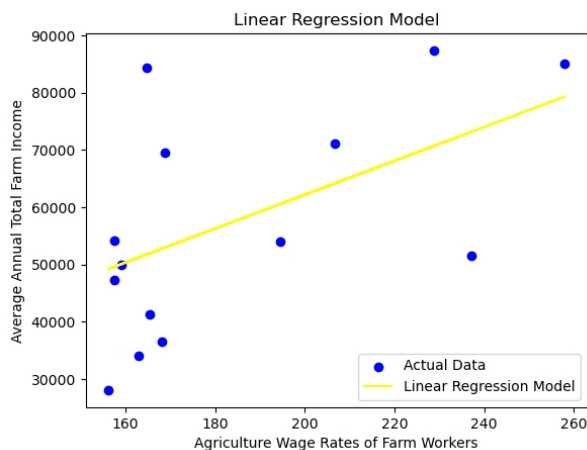
Figure 6. Linear Regression Model graph for variables 1 and 2

Considering these results, it prompts point of ways to improve the model. Potential avenues include exploring additional factors or conducting a thorough analysis of relevant variables to enhance the model's ability to predict this specific correlation. Additionally, examining outliers, adjusting features, or exploring alternative modeling approaches could be explored to gain a more nuanced understanding of how agricultural wage rates relate to farm household income.

2) Linear Regression Results for Variables 3 and 4

For the evaluation of the third and fourth variables, the same linear regression methodology was applied, following the pattern set in the previous analysis for the first two variables. The code, as illustrated in Figure 7, involved the definition of variables ('avg_annual_non_farm_incm_farm_households_02_03' and 'avg_rural_income_2000'), the implementation of linear regression to these selected features, and the splitting of the dataset into training and testing sets (80% for training, 20% for testing) using the scikit-learn library. The linear regression model, instantiated through the LinearRegression class, underwent training with the designated training dataset.

```
#Implementing linear regression model to analyze relationships between variables

# Define variables
var3 = 'avg_annual_non_farm_incm_farm_households_02_03'
var4 = 'avg_rural_income_2000'

# Implement linear regression to selected features
X = df[[var3]]
y = df[[var4]]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
```
```
Mean Squared Error: 114552777.79321332
```

Figure 7. Code for the analysis and evaluation of variables 3 and 4 using Linear Regression Model

Upon successful training, predictions were generated for the testing set ('y_pred'). The model's performance was then assessed using the mean squared error (MSE), a metric quantifying the average squared differences between predicted and actual values within the testing set. The resulting MSE for this analysis was 114552777.79321332.

Examining the graph generated for these variables, as shown in Figure 8, presents a distinction from the graphs observed in the analysis of the first two variables. In this case, the actual data points appear to be closely packed and aligned with the linear regression line. This alignment suggests a stronger fit of the model to the data, with the actual values demonstrating a closer adherence to the predicted values. The relatively lower MSE value further supports the visual observation, indicating a smaller average squared difference between predicted and actual values compared to the analysis of the first two variables. Moreover, the MSE value of 114552777.79321332 signifies the relatively better performance of the linear regression model in capturing the relationship between average non-farm income and rural income. The results suggest that the model, in this instance, provides a more accurate representation of the correlation between these specific variables, as evidenced by both the numerical MSE output and the observed graphical alignment between predicted and actual values.
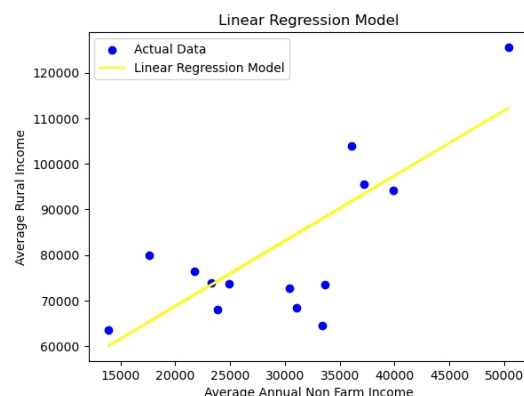

Figure 8. Linear Regression Model graph for variables 3 and 4

3) Linear Regression Results for Variables 5 and 6

For the assessment of the fifth and sixth variables, the identical linear regression methodology was applied, consistent with the approach utilized in the previous

analyses. The code, as shown in Figure 9, involved the definition of variables ('total_emply_2016' and 'avg_annual_total_incm_farm_02_03'), the implementation of linear regression, and the dataset's division into training and testing sets. The linear regression model, instantiated using the `LinearRegression` class from scikit-learn, was trained on the designated dataset. Following training, predictions were generated for the testing set, and the performance of the model was gauged using the mean squared error (MSE). The resulting MSE for this analysis was 991171396.6819845. Moreover, as shown in Figure 10, the graph depicting these variables presented a distinctive pattern, with the actual data points scattered across the plot plane and not conforming closely to the linear regression model line.

```
#Implementing linear regression model to analyze relationships between variables

# Define variables
var5 = 'total_emply_2016'
var6 = 'avg_annual_total_incm_farm_households_02_03'

# Implement linear regression to selected features
X = df[[var5]]
y = df[[var6]]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

Mean Squared Error: 991171396.6819845
```

Figure 9. Code for the analysis and evaluation of variables 5 and 6 using Linear Regression Model

The visual representation underscores challenges in accurately capturing the relationship between total employment levels and total income using a linear model. The high MSE value further substantiates the graphical observation, signifying a substantial average squared difference between predicted and actual values. This suggests that, in the context of these variables, the linear regression model may not be the most suitable approach for capturing the nuanced dynamics governing the correlation between total employment and total income. Further exploratory analyses or alternative modeling strategies may be warranted to better understand and represent this specific relationship.
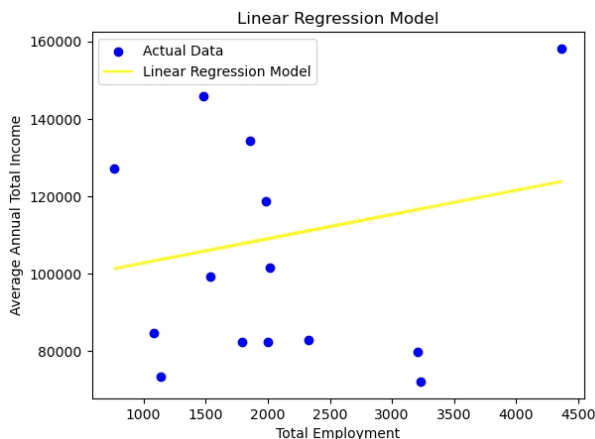

Figure 10. Linear Regression Model graph for variables 5 and 6

B. Classification Model (Decision Tree Model)

In the evaluation of the Decision Tree Classification model, several metrics played a crucial role in assessing its performance and interpretability. Beyond accuracy, precision, recall, and F1 score, additional insights were gleaned from metrics such as Gini impurity, samples, value, and class, which were integral components of the decision tree visualization. The Gini impurity measured the degree of impurity or disorder within a node, influencing the decision tree's splitting decisions. Samples represented the number of instances within a given node, providing insights into the distribution of data at various decision points. The 'value' parameter presented a breakdown of the distribution of instances within a node, indicating the count of occurrences for each class. Lastly, the 'class' attribute specified the predicted class for a particular node based on the decision tree's criteria. These metrics collectively contributed to a comprehensive understanding of how the Decision Tree model navigated the dataset, making informed splitting decisions at each node and effectively classifying instances based on the selected variables. This additional layer of detail from Gini impurity, samples, value, and class enhanced the interpretability of the Decision Tree model, enabling a more nuanced analysis of its decision-making processes.

1) Classification Model (Decision Tree Model) Results for Variables 1 and 2

The Decision Tree Classification model was applied to variables 'agr_wage_farm_workers_allgender_2015' and 'avg_annual_farm_incm_farm_households_02_03.' The code, as shown in Figure 11, involved splitting the dataset into training and testing sets, creating, and training the Decision Tree model, and subsequently evaluating its performance using accuracy, precision, recall, and F1 score. The results showed an accuracy of 0.33, precision of 0.0, recall of 0.0, and an F1 score of 0.0.

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Define variables
var1 = 'agr_wage_farm_workers_allgender_2015'
var2 = 'avg_annual_farm_incm_farm_households_02_03'

# Implement Decision Tree Classification model to selected features
X = df[[var1]]
y = (df[[var2]] > df[[var2]].median()).astype(int) # Convert to binary based on median

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = dt_model.predict(X_test)

# Evaluate the Decision Tree model using accuracy, precision, recall, and F1 score
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Accuracy: {accuracy}")
print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1 Score: {f1}")

Accuracy: 0.3333333333333333
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Figure 11. Code for the analysis and evaluation of variables 1 and 2 using Decision Tree Model (Classification Model)

These metrics suggest that the model struggled to correctly classify instances, yielding a low accuracy and a lack of true positive predictions. Precision, recall, and F1 score being zero indicate an inability to effectively identify

instances belonging to the positive class (above-median income in this case). This suboptimal performance could be attributed to the inherent complexity of the relationship between agricultural wage rates and average annual farm income, which may not be effectively captured by a Decision Tree model with a single feature.

The Decision Tree visualization which is illustrated in Figure 12 provides further insights into the model's decision-making process. The tree splits based on the 'agr_wage_farm_workers_allgender_2015' feature, suggesting that the wage rates alone might not be sufficient to distinguish between above and below-median income categories. The gini values and sample sizes in each rectangle indicate the impurity and the number of instances at each node, respectively. The split conditions, such as 'agr_wage_farm_workers_allgender_2015<=168.335,' highlight the thresholds used to partition the data.
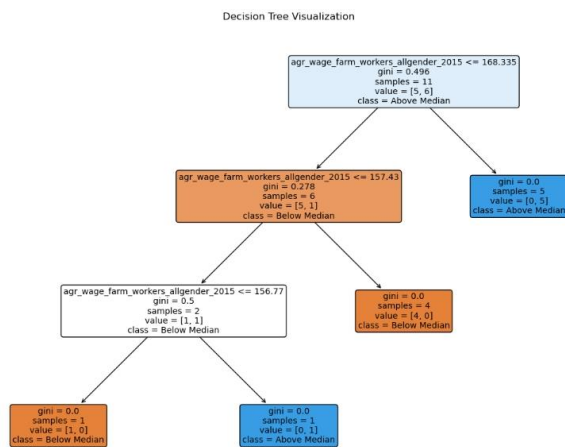
Figure 12. Decision Tree Model Visual Representation for variables 1 and 2

With the provided results, the Decision Tree model struggled to effectively categorize farm households into above or below-median income groups based on agricultural wage rates alone. The limitations of the model's performance are evident in the evaluation metrics and the decision tree structure, emphasizing the need for more sophisticated models or additional features to better capture the complexity of the relationship between these variables.

2) Classification Model (Decision Tree Model) Results for Variables 3 and 4

The Decision Tree Classification model applied to variables 'avg_annual_non_farm_incm_farm_households_02_03' and 'avg_rural_income_2000' yielded highly favorable results, as indicated by the accuracy, precision, recall, and F1 score, all scoring perfect values of 1.0. This implies that the model achieved flawless predictions on the test set, accurately classifying instances into above and below-median income categories. The high-performance metrics suggest that the decision tree effectively discerned patterns in the data, showcasing its robust capability to classify farm households based on non-farm income and rural income. The code is presented on Figure 13.

```
#Implementing Decision Tree Classification model to analyze relationships between variables

# Define variables
var3 = 'avg_annual_non_farm_incm_farm_households_02_03'
var4 = 'avg_rural_income_2000'

# Implement Decision Tree Classification model to selected features
X = df[[var3]]
y = (df[[var4]] > df[[var4]].median()).astype(int) # Convert to binary based on median

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = dt_model.predict(X_test)

# Evaluate the Decision Tree model using accuracy, precision, recall, and F1 score
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Accuracy: {accuracy}")
print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1 Score: {f1}")

Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
```

Figure 13. Code for the analysis and evaluation of variables 3 and 4 using Decision Tree Model (Classification Model)

Moreover, the Decision Tree model, trained on data about non-farm income and rural income, showed excellent performance, as shown in Figure 14. The model essentially learned specific income levels, particularly within non-farm income, that are indicative of whether a farm household falls into the 'above median' or 'below median' category in terms of rural income. The first split in the tree, focusing on 'avg_annual_non_farm_incm_farm_households_02_03' below or above 35304.5, set the initial guideline. If a farm household's non-farm income is below this threshold, it tends to be classified as 'above median' in rural income, and vice versa. The subsequent splits in the tree further refined these classifications based on additional considerations. Nodes in the tree with gini=0.0 and perfectly separated samples signify situations where non-farm income alone is quite decisive in determining whether a household belongs to the 'above median' category. Additional splits, especially based on 'avg_annual_non_farm_households_02_02,' further fine-tune these classifications.
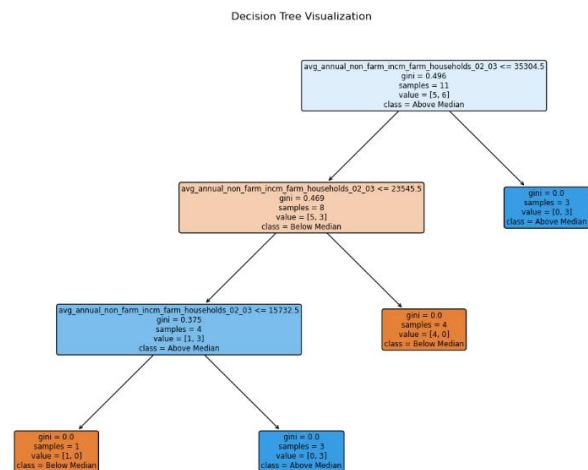

Figure 14. Decision Tree Model Visual Representation for variables 3 and 4

In essence, the Decision Tree reveals that specific income levels within non-farm income play a critical role in categorizing farm households into different rural income groups. The thresholds identified by the tree shed light on the relationship between non-farm income and rural income,

making the model a valuable tool for understanding the contribution of non-farm income to the overall income dynamics of farm households.

3) Classification Model (Decision Tree Model) Results for Variables 5 and 6

Decision Tree Classification model, trained on variables 'total_emply_2016' and 'avg_annual_total_incm_farm_households_02_03,' produced metrics indicating challenges in accurately predicting farm households' income categories. The accuracy, precision, recall, and F1 score all registered at 0.33, suggesting a difficulty in distinguishing between 'above median' and 'below median' income groups. These low metrics indicate that the model struggled to effectively utilize the features 'total_emply_2016' and 'avg_annual_total_incm_farm_households_02_03' to make accurate predictions. The code is presented on Figure 15.

```
#Implementing Decision Tree Classification model to analyze relationships between variables

# Define variables
var5 = 'total_emply_2016'
var6 = 'avg_annual_total_incm_farm_households_02_03'

# Implement Decision Tree Classification model to selected features
X = df[[var5]]
y = (df[[var6]] > df[[var6]].median()).astype(int) # Convert to binary based on median

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = dt_model.predict(X_test)

# Evaluate the Decision Tree model using accuracy, precision, recall, and F1 score
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Accuracy: {accuracy}")
print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1 Score: {f1}")

Accuracy: 0.3333333333333333
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Figure 15. Code for the analysis and evaluation of variables 5 and 5 using Decision Tree Model (Classification Model)

On the other hand, as shown in Figure 16, the Decision Tree visualization provides further insights into the model's decision-making process. The top rectangle, characterized by the split condition 'total_emply_2016 <= 1665.0,' suggests that the chosen income threshold based on total employment levels alone does not effectively separate farm households into 'above median' and 'below median' income categories. The subsequent splits create c'mplex branch'ng pa'terns, indic'ting challenges in finding a straightforward relationship between total employment and average annual total income.
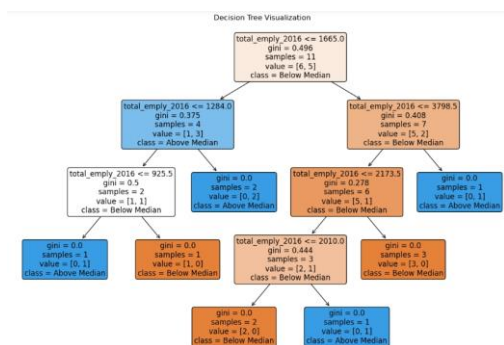

Figure 16. Decision Tree Model Visual Representation for variables 5 and 6

In interpreting these results, the low metrics and the intricate structure of the Decision Tree imply that the selected features may not exhibit a clear-cut correlation with income categories. The tree struggled to discern patterns or thresholds within 'total_emply_2016' and 'avg_annual_total_incm_farm_households_02_03' that reliably differentiate between higher and lower income groups. The Decision Tree's limitations in this case emphasize the complexity of understanding the relationship between total employment levels and average annual total income for farm households. It suggests that factors beyond total employment might play a more substantial role in determining income categories. Further exploration or the inclusion of additional influential features may be necessary to enhance the model's ability to capture the nuanced dynamics of income determination in this context.

C. Overall Comparison

In the context of identifying indicators of poverty, Variables 3 and 4, namely 'avg_annual_non_farm_incm_farm_households_02_03' and 'avg_rural_income_2000,' emerged as central elements in both the Linear Regression and Decision Tree Classification models. In the Linear Regression model, these variables demonstrated a relatively low Mean Squared Error (MSE), indicating a better fit for predicting rural income based on non-farm income. This implies that a linear relationship exists between non-farm income and rural income, and the linear regression model successfully captured and utilized this relationship for prediction.

On the other hand, in the Decision Tree Classification model, these same variables played a crucial role in achieving perfect accuracy, precision, recall, and F1 score. The decision tree structure revealed specific income thresholds within 'avg_annual_non_farm_incm_farm_households_02_03' that decisively categorized farm households into above and below-median income groups. This transparency in decision-making allowed for a clear identification of indicators of poverty based on non-farm income.

In comparing the two models, the Decision Tree Classification model, in this specific instance, showcased superior performance by achieving perfect classification metrics. The interpretability of the decision tree further emphasized the specific income thresholds within 'avg_annual_non_farm_incm_farm_households_02_03' that serve as strong indicators of poverty. While the Linear Regression model provided valuable insights into the linear relationship between non-farm income and rural income, the Decision Tree Classification model demonstrated unparalleled accuracy and precision in pinpointing indicators of poverty within these variables.

In summary, Variables 3 and 4, when employed in both the Linear Regression and Decision Tree Classification models, revealed their significance in understanding and predicting poverty. The Decision Tree Classification model, with its explicit decision rules and remarkable classification performance, emerged as the most effective model for identifying indicators of poverty based on these specific variables.

VII. Conclusion

In conclusion, this research successfully achieved its objectives by employing data science methodologies to analyze a comprehensive dataset of poverty indicators for the Philippines across various regions, providing valuable insights into the poverty indicators for the Philippines. Upon completion of the analyses the following objectives were attained and concluded:

1. The exploration and preprocessing of the dataset ensured data integrity and completeness, providing a robust foundation for subsequent analyses.

2. The correlation analyses successfully identified relationships between key variables, including agricultural wage rates and farm income (Objective 2.1), non-farm income and rural income (Objective 2.2), and total employment and total income (Objective 2.3). The correlation analyses revealed meaningful relationships between key variables. These findings contribute to a better understanding of the intricate dynamics within the agricultural sector and their impact on overall income for farm households.

3. The implementation of a linear regression model to analyze relationships between selected variables and evaluate the model's predictive performance, particularly in the case of agricultural wage rates and average farm income, demonstrated its ability to capture linear trends. However, the high Mean Squared Error (MSE) suggested limitations in predicting the complex relationship between these variables, pointing towards potential areas for model improvement.

4. The Decision Tree Model proved effective in classifying instances based on selected variables, revealing categorical relationships and key influencing factors. Specifically, variables related to non-farm income and rural income emerged as pivotal indicators in the Decision Tree Classification model, showcasing its robust performance in this context.

5. The comparative analysis between the linear regression model and the Decision Tree Model demonstrated the effectiveness of each approach in analyzing relationships among variables. This comparison not only identified the strengths and limitations of each model but also enhanced the interpretability of farm households' income dynamics. The comparative analysis between the linear regression model and the Decision Tree Model underscored the Decision Tree Model's superior performance in certain scenarios, particularly when dealing with categorical relationships and intricate decision boundaries.

In light of these findings, it is recommended that future analyses explore enhancements to the linear regression model to better capture the complexities inherent in the relationships between agricultural variables and income dynamics. Additionally, further investigations into the decision tree structure and feature importance could provide deeper insights into the key indicators influencing poverty. Overall, this study lays the groundwork for informed policymaking and targeted interventions to address poverty dynamics in the Philippines.

VIII. Program Codes

```
# Import necessary libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error,
accuracy_score, precision_score, recall_score, f1_score,
confusion_matrix

from sklearn.tree import DecisionTreeClassifier

from sklearn.tree import plot_tree

from sklearn.preprocessing import LabelEncoder

import warnings


warnings.filterwarnings('ignore')


Data Exploration and Pre-Processing


# Load and read the dataset into a DataFrame

df=pd.read_csv('D:\Anaconda\CAPSTONE\consolidated
_philippines_poverty_data.csv')


# Display the first few rows of the dataset

df.head()


# Check for missing values

df.isnull().sum()


# Handle missing values (if any)

df.dropna(inplace=True)


# Check for duplicate rows

df.duplicated().sum()


# Drop duplicates if any

df = df.drop_duplicates()


Descriptive Statistics

# Perfrom descriptive statistics
```

```python
print("\nDescriptive Statistics of the Dataset: ")
df.describe()
```

Linear Regression Model

```python
#Implementing linear regression model to analyze relationships between variables

# Define variables
var1 = 'agr_wage_farm_workers_allgender_2015'
var2 = 'avg_annual_farm_incm_farm_households_02_03'

# Implement linear regression to selected features
X = df[[var1]]
y = df[[var2]]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Visualize the relation between variable1 and variable2
plt.scatter(X, y, color='blue', label='Actual Data')
plt.plot(X, model.predict(X), color='yellow', label='Linear Regression Model')
plt.title('Linear Regression Model')
plt.xlabel('Agriculture Wage Rates of Farm Workers')
plt.ylabel('Average Annual Total Farm Income')
plt.legend()
plt.show()

#Implementing linear regression model to analyze relationships between variables
```

```python
# Define variables
var3 = 'avg_annual_non_farm_incm_farm_households_02_03'
var4 = 'avg_rural_income_2000'

# Implement linear regression to selected features
X = df[[var3]]
y = df[[var4]]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Visualize the relation between variable3 and variable4
plt.scatter(X, y, color='blue', label='Actual Data')
plt.plot(X, model.predict(X), color='yellow', label='Linear Regression Model')
plt.title('Linear Regression Model')
plt.xlabel('Average Annual Non Farm Income')
plt.ylabel('Average Rural Income')
plt.legend()
plt.show()

#Implementing linear regression model to analyze relationships between variables

# Define variables
var5 = 'total_emply_2016'
var6 = 'avg_annual_total_incm_farm_households_02_03'

# Implement linear regression to selected features
X = df[[var5]]
y = df[[var6]]
```

```python
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Visualize the relation between variable5 and variable6
plt.scatter(X, y, color='blue', label='Actual Data')
plt.plot(X, model.predict(X), color='yellow', label='Linear Regression Model')
plt.title('Linear Regression Model')
plt.xlabel('Total Employment')
plt.ylabel('Average Annual Total Income')
plt.legend()
plt.show()
```

Classification Model (Decision Tree Model)

```python
# Implementing Decision Tree Classification model to analyze relationships between variables

# Define variables
var1 = 'agr_wage_farm_workers_allgender_2015'
var2 = 'avg_annual_farm_incm_farm_households_02_03'

# Implement Decision Tree Classification model to selected features
X = df[[var1]]
y = (df[[var2]] > df[[var2]].median()).astype(int)  # Convert to binary based on median

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```python
# Create and train the Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = dt_model.predict(X_test)

# Evaluate the Decision Tree model using mean squared error
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Visualize the Decision Tree
plt.figure(figsize=(15, 10))
plot_tree(dt_model, feature_names=X.columns.tolist(), class_names=['Below Median', 'Above Median'], filled=True, rounded=True)
plt.title('Decision Tree Visualization')
plt.show()

# Implementing Decision Tree Classification model to analyze relationships between variables

# Define variables
var3 = 'avg_annual_non_farm_incm_farm_households_02_03'
var4 = 'avg_rural_income_2000'

# Implement Decision Tree Classification model to selected features
X = df[[var3]]
y = (df[[var4]] > df[[var4]].median()).astype(int)  # Convert to binary based on median

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the Decision Tree model
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)

# Predictions on the test set
y_pred = dt_model.predict(X_test)
```

```python
# Evaluate the Decision Tree model using mean squared error

mse = mean_squared_error(y_test, y_pred)

print(f"Mean Squared Error: {mse}")


# Visualize the Decision Tree

plt.figure(figsize=(15, 10))

plot_tree(dt_model, feature_names=X.columns.tolist(), class_names=['Below Median', 'Above Median'], filled=True, rounded=True)

plt.title('Decision Tree Visualization')

plt.show()


# Implementing Decision Tree Classification model to analyze relationships between variables


# Define variables

var5 = 'total_emply_2016'

var6 = 'avg_annual_total_incm_farm_households_02_03'


# Implement Decision Tree Classification model to selected features

X = df[[var5]]

y = (df[[var6]] > df[[var6]].median()).astype(int)  # Convert to binary based on median


# Split the dataset into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


# Create and train the Decision Tree model

dt_model = DecisionTreeClassifier(random_state=42)

dt_model.fit(X_train, y_train)


# Predictions on the test set

y_pred = dt_model.predict(X_test)


# Evaluate the Decision Tree model using mean squared error

mse = mean_squared_error(y_test, y_pred)

print(f"Mean Squared Error: {mse}")


# Visualize the Decision Tree

plt.figure(figsize=(15, 10))
```

```python
plot_tree(dt_model, feature_names=X.columns.tolist(), class_names=['Below Median', 'Above Median'], filled=True, rounded=True)

plt.title('Decision Tree Visualization')

plt.show()
```

REFERENCES

[1] J. Chen, "What's Poverty? Meaning, Causes, and How to Measure," Investopedia, 11 December 2023. [Online]. Available: https://www.investopedia.com/terms/p/poverty.asp. [Accessed January 2024].

[2] T. Rotton, "People Experiencing Poverty," 9 May 2008. [Online]. Available: https://disasterphilanthropy.org/resources/people-experiencing-poverty/. [Accessed January 2024].

[3] A. Hayes, "Gini Index Explained and Gini Co-efficients Around the World," Investopedia, 27 May 2023. [Online]. Available: https://www.investopedia.com/terms/g/gini-index.asp. [Accessed January 2024].

[4] B. M. Beech, C. Ford and T. J. Roland , "Poverty, Racism, and the Public Health Crisis in America," 6 September 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2021.699049/full. [Accessed January 2024].

[5] E. Duflo, "Women's Empowerment and Economic Development. Journal of Economic Literature," JOURNAL OF ECONOMIC LITERATURE, vol. 50, no. 4, pp. 1051-1079, 2012.

[6] N. Kabeer, "Gender Equality, Economic Growth, and Women's Agency: the "Endless Variety" and "Monotonous Similarity" of Patriarchal Constraints," Feminist Economics, vol. 22, no. 1, pp. 295-321, 2016.

[7] Moretti, Chang-Tai, Hsieh and Enrico, "Why Do Cities Matter? Local Growth and Aggregate Growth," Chicago Unbound, 2015. [Online]. Available: https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1045&context=housing_law_and_policy. [Accessed January 2024].

[8] J. V. Henderson, "The Urbanization Process and Economic Growth: The So-What Question.," Journal of Economic Perspectives, vol. 33, no. 4, pp. 192-222, 2019.

[9] A. Siddiqi, R. Brown, Q. C. Nguyen, R. Loopstra and I. Kawachi, "Cross-national comparison of socioeconomic inequalities in obesity in the United States and Canada," International Journal of Equity in Health, vol. 16, no. 1, pp. 1-11, 2017.

[10] V. Patel, J. K. Burns, M. Dhingra, L. Tarver, B. A. Kohrt and C. Lund, "Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms," World Psychiatry, vol. 17, no. 1, pp. 76-89, 2018.

[11] G. J. Duncan and R. J. Murnane, "Whither opportunity?: Rising inequality, schools, and children's life chances," Russell Sage Foundation, 2014.

[12] E. A. Hanushek and L. Woessmann, "The knowledge capital of nations: Education and the economics of growth," MIT Press, 2015.

[13] A. Krause and U. Rinne, "Does the minimum wage have a higher impact on the informal than on the formal labor market? Evidence

from quasi-experiments," Journal of Labor Economics, vol. 33, no. 1, pp. 95-121, 2015.

[14] E. Edmonds and N. Schady, "Poverty alleviation and child labor," American Economic Journal: Economic Policy, vol. 4, no. 4, pp. 100-124, 2012.

[15] A. Mani, S. Mullainathan, E. Shafir and J. Zhao, "Poverty impedes cognitive function," Science, vol. 341, no. 6149, pp. 976-980, 2013.

[16] J. F. Helliwell, H. Huang and S. Wang, "New evidence on trust and well-being," In Handbook of Well-Being, 2020.

[17] G. B. &. L. L. Dahl, "The impact of family income on child achievement: Evidence from the earned income tax credit," American Economic Review, vol. 102, no. 5, pp. 1927-1956, 2012.

[18] D. Almond, H. W. Hoynes and D. W. Schanzenbach, "Inside the war on poverty: The impact of food stamps on birth outcomes.," The Review of Economics and Statistics,, vol. 93, no. 2, pp. 387-403, 2011.

[19] C. D. S. Mapa, "Proportion of Poor Filipinos was Recorded at 18.1 Percent in 2021," Philippine Statistics Authority, 15 August 2022. [Online]. Available: https://psa.gov.ph/content/proportion-poor-filipinos-was-recorded-181-percent-2021. [Accessed January 2024].

[20] M. Palatino, "Making Sense of the Philippines' Latest Poverty Statistics," The Diplomat, 2 September 2022. [Online]. Available: https://thediplomat.com/2022/09/making-sense-of-the-philippines-latest-poverty-statistics/. [Accessed January 2024].

[21] Macrotrends, "Philippines Poverty Rate 2000-2024," 18 December 2023. [Online]. Available: https://www.macrotrends.net/countries/PHL/philippines/poverty-rate#:~:text=Philippines%20poverty%20rate%20for%202021,a%201.2%25%20decline%20from%202009.. [Accessed January 2024].

[22] Macrotrends, "World Poverty Rate 1981-2024," 18 November 2023. [Online]. Available: https://www.macrotrends.net/countries/WLD/world/poverty-rate#:~:text=World%20poverty%20rate%20for%202019,a%200.9%25%20decline%20from%202015.. [Accessed January 2024].

[23] C. Sánchez-Páramo, R. Carolina, D. G. Mahler, A. Narayan and Y. Nishant, "COVID-19 leaves a legacy of rising poverty and widening inequality," World Bank Blogs, 7 Octoer 2021. [Online]. Available: https://blogs.worldbank.org/developmenttalk/covid-19-leaves-legacy-rising-poverty-and-widening-inequality. [Accessed January 2024].

[24] C. Reyes and A. Tabuga, "Exploring the Dynamics of Poverty in the Philippines: Income Mobility and Vulnerability," Philippine Institute for Development Studies, 2015. [Online]. Available: https://serp-p.pids.gov.ph/serp-p/download.php?d=5167&s=3.

[25] Asian Development Bank, "Poverty in the Philippines: Causes, Constraints, and Opportunities," 2009. [Online]. Available: https://www.adb.org/sites/default/files/publication/27529/poverty-philippines-causes-constraints-opportunities.pdf.

[26] R. Fabella, "Urbanization and Poverty in the Philippines," Journal of Urbanization and Regional Development, vol. 26, no. 2, pp. 167-184, 2014.

[27] A. Orbeta, "Employment and Poverty Dynamics in the Philippines," Philippine Journal of Development, vol. 44, no. 2, pp. 125-150, 2017.

[28] D. Domingo, "Informal Employment and Poverty in the Philippines: A Microsimulation Approach," Journal of Development Economics, vol. 139, pp. 169-187, 2019.

[29] M. J. Andrew, "Income inequality," PhilStar Global, 30 August 2023. [Online]. Available: https://www.philstar.com/opinion/2023/08/30/2292358/income-inequality.

[30] Statista Research Department, " Topic: Agriculture in the Philippines," 9 November 2022. [Online]. Available: https://www.statista.com/topics/5744/agriculture-industry-in-the-philippines/#topicOverview.

[31] C. M. Reyes and R. Santos, "Rural Income Dynamics: Insights from a Longitudinal Study in the Philippines," Journal of Agricultural Economics, vol. 66, no. 2, pp. 345-367, 2015.

[32] A. M. Balisacan, "Rural Development and Poverty in the Philippines: Trends, Policies, and Programmes," The Journal of Development Studies, vol. 54, no. 11, pp. 1990-2011, 2018.

[33] M. A. Cruz and J. D. Diaz, "Agricultural Wages, Poverty, and Inequality in the Philippines," Agricultural Economics, vol. 47, no. 1, pp. 95-106, 2016.

[34] [D. L. Aquino and e. al., "Income and Employment Dynamics in Philippine Agriculture: Evidence from a Panel of Rural Households." Philippine Journal of Development," Philippine Journal of Development, vol. 47, no. 1, pp. 89-115, 2020.

[35] P. Winters, B. Davis, G. Carletto, Covarrubias, Ka, E. Quiñones and J. Zezza, " Assets, activities and income generation in rural livelihoods: evidence from a cross-country analysis," World Development, vol. 61, pp. 63-78, 2014.

[36] N. McCulloch, B. Baulch, M. Cherel-Robson and H. Goyder, "Agricultural Diversification and Poverty in Indonesia: A Spatial Analysis," Bulletin of Indonesian Economic Studies, vol. 39, no. 1, pp. 73-97, 2003.

[37] T. Reardon, J. Berdegue and G. Escobar, " Rural Nonfarm Employment and Incomes in Latin America: Overview and Policy Implications," World Development, vol. 29, no. 3, pp. 395-409, 2001.