Combining Unsupervised and Supervised Learning to Predict Poverty Households in Sakon Nakhon, Thailand

Sutisa Songleknok, Suthasinee Kuptabut*
Department of Computer, Sakon Nakhon Rajabhat University, Sakon Nakhon, Thailand

***Abstract*—Poverty is a problem that various government agencies are attempting to address accurately and precisely. This solution relies on data and analysis of features affecting poverty. Machine Learning is a technique to analyze and focus on poverty features encompassing five livelihood capitals: human, physical, economic, natural, and social capital to understand the household context and environment. The dataset contains 1,598 poverty households from Kut Bak district, Sakon Nakhon, Thailand. K-prototype was used to group categorical and numerical dataset into four clusters and labelled as Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor. The performances of the Decision tree classifier with feature selection algorithms, including MI, ReliefF, RFE, and SFS, are compared. The best performance is SFS with F-measure, precision, and recall at 74.6%, 74.8%, and 74.7%, respectively. The result is the decision tree rules to predict the poverty level of households, enabling the establishment of guidelines for resolving household issues, and addressing broader problems within the areas.**

Keywords*—K-prototype; decision tree; feature selection; Sakon Nakhon poverty households; unsupervised learning; supervised learning

I. Introduction

Poverty alleviation policy holds a significant role in every nation, with each country tailoring its poverty criteria to evaluate household poverty accordingly. Specifically, many of these countries utilize the poverty line measurement, a criterion established by the United Nations [1], to gauge household poverty levels. Subsequent poverty-related issues arise from various factors, extending beyond just income. Pandemic mitigating investments have largely not rectified issues such as poor health, lack of quality agricultural resources, and production quantity. These factors have further contributed to household poverty, making them multidimensional poor.

Thailand has established an information system dedicated to poor households. Responsible organizations, namely the Community Development Department [2] under the Health Board of Quality of Public Life Development (HBL), oversee data collection, database management, and information display through official online platforms. This organization collects basic minimum needs data at a household level in Thailand. This data demonstrates household members' fundamental status across various aspects of life quality, adhering to minimum standards.

In collaboration with the Office of the National Economics and Social Development Council (NESDC) and the National Electronics and Computer Technology Center (NECTEC), the Thai People Map and Analytics Platform (TPMAP) [3], was developed. This platform is designed to identify individuals in impoverished households who meet the criteria for basic minimum needs across five dimensions of poverty: health, living conditions, education, income, and access to public services. In addition, the Program Management Unit on Area-Based Development (PMUA) has developed a system called Practical Poverty Provincial Connext (PPPConnext) [4] to collect data on impoverished households within 20 provinces, which rank lowest in the nation's Human Achievement Index (HAI) regarding income. This data collection focuses on five dimensions of livelihood assets: human capital, economic capital, natural capital, physical capital, and social capital. These dimensions are utilized to develop appropriate solutions that match the specific needs of these households.

Most studies on poverty primarily focus on the target group at a household level, employing various research methodologies, including qualitative, quantitative, and Machine learning techniques. The poverty data were analyzed using two techniques: 1) Supervised Learning: This approach involves formulating the poverty level or target class. For instance, the research in [5], a study on factors affecting poverty, in [6], the prediction of households exhibiting characteristics at risk of poverty, and in [7] the development of prediction models for depression levels among the elderly in low-income households. This model utilizes techniques, such as Decision tree, Logistic regression, Neural networks, and Random forest. 2) Unsupervised Learning: This approach involves processing data without predetermined poverty level formulations, such as the research in [8, 9] that used clustering techniques to categorize impoverished households. Poverty alleviation programs should be prioritized in household clusters based on the poor conditions identified within each cluster. The research in [10] analyzes the factors affecting the sustainable livelihoods of poverty households. Both techniques contribute to predicting the relationship between factors and poverty, shaping policy guidelines and effective solutions to alleviate poverty.

The objectives of this paper are 1) to cluster the poverty households, 2) to compare the performance of feature selection techniques using a Decision tree classifier, and 3) to create rules to predict poverty status households. We utilized data from impoverished households collected through PPPConnext, and applied data mining techniques to comprehend the characteristics associated with poverty. The identified features are analyzed using both unsupervised learning and supervised learning models. Unsupervised learning, specifically the K-prototype algorithm, is employed to cluster the households based on their living capital aspects. Supervised learning, including Mutual information (MI), ReliefF, Recursive feature elimination (RFE), and Sequential forward selection (SFS) are utilized to reduce the information sizes. A Decision tree is utilized to construct a comparative model, select appropriate features, explain household characteristics, and predict the poverty level of households effectively.

The paper is structured as follows: Section II provides a comprehensive review of the related works; Section III outlines the proposed framework, detailing both clustering and classification steps, and Section IV presents the obtained results. Section V and Section VI wrap up the paper with a discussion and conclusion, respectively.

II. Related Work

A. Multidimensional Poverty

Poverty can be defined from different perspectives. Organizations, such as NESDC use the "poverty line" to establish standards for basic minimum food needs and essential goods, quantified in Baht per person per month. The poverty line has changed accordingly over the years. Individuals earning less than this threshold are classified as "poor." This categorization is determined by comparing monthly income against the poverty line [11, 12]. Poverty has a relationship with the households' economic status. The measurement of poverty in general uses the poverty line criteria as the condition for classifying poverty and non-poverty of households, which is widely used at provincial, national, and global levels. Nevertheless, extensive global research revealed that household poverty can be attributed to either insufficient income or various other contributing factors.

The UNDP and the Oxford Poverty and Human Development Initiative (OPHI) [11] have jointly created the Multidimensional Poverty Index (MPI) comprising 10 dimensions: nutrition, child mortality, and years in schooling, school attendance, cooking fuel, sanitation, drinking water, electricity, housing, and asset ownership. A lower MPI value for a country signifies reduced poverty, whereas a high MPI suggests significant multidimensional challenges, related to inequalities in areas like gender, ethnicity, and infrastructure.

B. Feature Selection

Feature selection (FS) involves the process of selecting, removing, and reducing duplication of the relevant features. There are three approaches to feature selection: Filter, Wrapper, and Embedded [13 - 15].

1) Filter approach. Filters evaluate the relevance of features based on intrinsic characteristics. The popular filter approaches include MI and ReliefF.

2) Wrapper approach. The wrapper approach constructs prediction models considering the feature interactions. The well-known wrapper approaches are REF and SFS.

3) Embedded approach. In feature selection, three main approaches are employed: Filter, Wrapper, and Embedded. Embedded approaches like LASSO (L1-Regularization) and RIDGE (L2-Regularization) combine aspects of both filters and wrappers.

In this paper, feature selection was implemented using the Filter approach (MI and ReliefF) and the Wrapper approach (REF and SFS).

C. Unsupervised Learning

The clustering technique, an unsupervised learning approach, is used for categorizing data based on similarities in their characteristics. In the analysis of poverty, this technique segregates impoverished households according to dataset features. The parameter k (number of clusters) can be determined either through predefined business rules, or varied techniques to determine an appropriate value for "k".

When employing the clustering technique with the numeric dataset, the utilization of K-means clustering is required. For instance, the study in [8] applied the K-means algorithm to assess poverty status and categorize it into three levels: low, medium, and high poverty. This method was implemented in households in Hulu Sungai Tengah Regency, Indonesia. The study's findings would be utilized to develop policies tailored to individual households to achieve specific goals. The study in [9] analyzes poverty conditions within a community in the Philippines, and groups households into three clusters: stable, critical, and at-risk. Each cluster offers valuable insights into poverty conditions, guiding the community's planning and program implementation.

In study [16], the utilization of the clustering technique to group poverty households of Lagangilang, Abra, Philippines was divided into three clusters: non-poor, near poor, and poor. Each group describes different characteristics of households following health and nutrition, education, income, and livelihood to formulate appropriate poverty reduction programs. In cases [10] involving both numerical and categorical data, the K-prototype algorithm was employed. For example, this approach was utilized to analyze factors influencing the sustainable livelihoods of impoverished households, using data mining from poor households in Kut Bak district in Sakon Nakhon province, Thailand. The dataset classified poor households into four clusters: extreme, high, moderate, and low poverty levels.

D. Supervised Learning

Classification, a supervised machine learning method, involves constructing models to predict data labels in a given dataset. For example [17], poverty prediction was carried out using three techniques: Softmax classification, Random forest classification, and Multi-layer perception classifier. In predicting impoverished household data from the Cambodia DHS dataset, two types of predictive outcomes are considered: three-class classification (poor, middle, rich), and five-class classification (poorest, poorer, middle, richer, richest). The study revealed that the three-class classification achieved a higher accuracy of 87%, compared to the five-class classification.

Another research in [18] focused on identifying the causes of old-age poverty in South Korea, the decision tree algorithm was applied using 13 variables, with old-aged poverty as the target. The study revealed that earned income was the most significant factor influencing elderly poverty. In another research [19] conducted in Malaysia, Naive Bayes, Decision tree, and K-nearest neighbors' classifiers were employed to predict the bottom 40 percent of poverty households. Among these models, the decision tree model achieved the highest performance. Additionally, in [20] the decision tree model was utilized to predict household poverty based on health status using the Cuatro Santos health and demographic surveillance databases in Nicaragua. The key indicators of poverty, such as the presence of piped water with a meter, the highest education level in households, and ownership of a refrigerator.

This paper employed the Decision tree algorithm to generate a tree-like structure to represent classification rules. In this structure, internal nodes represent dataset features, branches represent decision rules, and each leaf node represents a specific class label.

III. METHODOLOGY

We acknowledge the importance of outlining the reasons behind our selection of the "proposed framework" in addressing the specific problems at hand. Here are the reasons that make the proposed framework appropriate for addressing such problems. Collectively, these facets enable comprehensive analysis, feature emphasis, and effective model evaluation, making the framework apt for addressing the problem. The proposed framework follows a systematic four-phase structure, ensuring methodical handling from data preprocessing to model evaluation. It sources the poverty dataset from a reliable source (PPPConnext) and employs K-prototype clustering for grouping similar data elements, aiding focused analysis. Automatic labelling of poverty households within clusters streamlines interpretation. Multiple feature selection techniques (MI, ReliefF, REF, and SFS) enhance model efficiency by emphasizing impactful attributes. Evaluation via Decision tree classifier comparing feature sets based on F-measure, precision, and recall ensures a robust model selection.

The studies as mentioned earlier used either unsupervised or supervised learning techniques. However, our framework utilized both unsupervised and supervised learning. Unsupervised learning was used to determine the poverty status of households, while supervised learning selected suitable feature datasets to generate rules using a decision tree model for predicting poverty status.

The limitations of the existing framework that may hinder its suitability for the current problems include the dependency on specific datasets restricting adaptability to new data structures, the clustering method may struggle to effectively group elements in different data types, limited feature selection techniques hindering the identification of crucial attributes, rigid evaluation metrics might not align with the problem domain, and inflexibility in model selection might limit adaptability to diverse data demands.

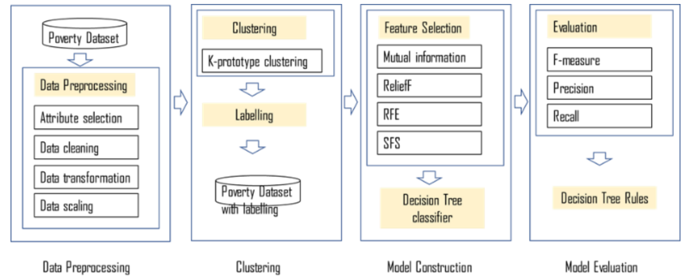

Fig. 1. Conceptual framework.

The conceptual framework shown in Fig. 1 outlines a four-phase process: Data preprocessing, Clustering, Model construction, and Model evaluation. In the data preprocessing phase, the poverty dataset was sourced from the PPPConnext database and underwent preparation for analysis. Utilizing K-prototype clustering, data objects with similar characteristics were grouped into clusters; ensuring data with differing characteristics were placed in distinct clusters. The poverty households cluster was then automatically labelled. Following this, feature selection techniques, MI, ReliefF, REF, and SFS, were applied to identify relevant features. The study then evaluated the performance of the Decision tree classifier by comparing various feature sets, selecting the one with the best performance based on F-measure, precision, and recall.

A. Data Set

The dataset utilized in this study originates from the PPPConnext database, focusing on poor households situated in Kut Bak district, Sakon Nakhon province, Thailand. Sakon Nakhon province ranked 71st out of 78 provinces in terms of income index in 2019. Within this province, Kut Bak district had a poverty rate reaching TPMAP in 2019 at the highest level. This district also held the distinction of having the lowest income level in Sakon Nakhon province. The dataset consists of five types of livelihood assets (human capital, physical capital, economic capital, natural capital, and social capital) to explain the asset limitations of poverty and power within the households. In total, the dataset contains responses from 1,598 households with 76 features (58 categorical features and 18 numerical features). These features are described in detail in Table I.

TABLE I. DETAILED DESCRIPTION OF FEATURES

capitals	**Attributes**	**Explanation**	**Value**
Economic capital (24 features)	income_remitted (N)	monthly remittances	Mean:909.57
	income_farming (N)	monthly income from farming	Mean:3,074.45
	income_non_farming (N)	monthly non-farming income	Mean:6,074.45
	income_welfare (N)	monthly income from state welfare	Mean:893.93
	expenses (N)	monthly household expenses	Mean: 6,951.47
	rice_farming (C)	rice farming households	0=No (14.02%) 1=Yes (85.98%)
	livestock (C)	livestock raising households	0=No (97.43%) 1=Yes (2.57%)

capitals	Attributes	Explanation	Value
	fishing (C)	freshwater fishery households	0=No (79.91%) 1=Yes (20.09%)
	industrial_crop (C)	industrial crop farming households	0=No (86.92%) 1=Yes (13.08%)
	poultry (C)	poultry farming household	0=No (94.62%) 1=Yes (5.38%)
	pig (C)	pig farming household	0=No (99.12%) 1=Yes (0.88%)
	cattle (C)	cattle farming households	0=No (72.03%) 1=Yes (27.97%)
	loan_cousin (C)	a loan from cousins with no collection of interest	0=No (98.56%) 1=Yes (1.44%)
	loan_cousin_ interest (C)	a loan from cousins with a collection of interest	0=No (95.24%) 1=Yes (4.76%)
	loan_community (C)	a loan from the financial savings community-based organizations	0=No (78.85%) 1=Yes (21.15%)
	loan_state (C)	a loan from state-support financial savings	0=No (78.29%) 1=Yes (21.71%)
	loan_BAAC (C)	a loan from the Bank for Agriculture and Agricultural Cooperatives	0=No (82.67%) 1=Yes (17.33%)
	loan_savings_ bank (C)	a loan from the Government Savings Bank	0=No (97.62%) 1=Yes (2.38%)
	loan_commercial _bank (C)	a loan from the Thai Commercial Bank	0=No (98.81%) 1=Yes (1.19%)
	loan_private (C)	a loan from Private AMC	0=No (98.87%) 1=Yes (1.13%)
	loan_creadit_shop (C)	households with credit accessibility from consumer goods shops and production factors	0=No (99.50%) 1=Yes (0.50%)
	loan_informal_ debt (C)	a loan from informal debt	0=No (99.50%) 1=Yes (0.50%)
	student_loan (C)	a loan from a student loan fund	0=No (98.62%) 1=Yes (1.38%)
	savings (C)	households with savings	0=No (45.74%) 1=Yes (54.26%)
Physical capital (31 features)	water_resources (C)	farming using water from water resources, such as rivers, brooks, and ditches	0=No (65.71%) 1=Yes (34.29%)
	reservoirs (C)	farming using water from reservoirs, and village ponds	0=No (93.43%) 1=Yes (6.57%)
	groundwater (C)	farming using water from under groundwater, surface water, artesian aquifer	0=No (85.61%) 1=Yes (14.39%)
	rainwater (C)	farming using rainwater	0=No (40.05%) 1=Yes (59.95%)
	irrigation_canals (C)	farming using irrigation canals	0=No (97.93%) 1=Yes (2.07%)
	ownership_rights (C)	households with issuing legal land rights documents, such as Title Deed	0=No (66.83%) 1=Yes (33.17%)
	ownership_rights _rai (N)	number of arable lands issuing legal land rights documents of ownership (Rai)	
	government_lease (C)	households with land title documents, such as ALRO, S.K.1	0=No (36.55%) 1=Yes (63.45%)
	government_lease _rai (N)	number of arable issuing land title documents (Rai)	
	no_title_arable_ land (C)	households with no land tile documents for arable lands in forest-protected areas or national parks	0=No (90.68%) 1=Yes (9.32%)
	no_title_arable_ land _rai (N)	number of arable lands within forest-protected areas or national parks (Rai)	
	others_rent_free (C)	households farm on others' arable lands rent-free	0=No (84.29%) 1=Yes (15.71%)
	others_rent_ free _rai (N)	number of households farm on others' arable lands rent-free	

capitals	Attributes	Explanation	Value
	others_rent (C)	households farm on others' arable lands with rent payment	0=No (99.50%) 1=Yes (0.50%)
	others_rent_rai (N)	number of arable lands from others with payment (Rai)	
	no_access_water (C)	arable lands with water resources inaccessible for cultivation	0=No (60.76%) 1=Yes (39.24%)
	fertile_soil (C)	number of arable lands with fertile agricultural land	0=No (86.55%) 1=Yes (13.45%)
	risk_area (C)	arable lands in areas with natural disaster risks, such as floods	0=No (98.62%) 1=Yes (1.38%)
	home_ownership (C)	households owning their own homes	1 = Staying with others (0.63%) 2 = Renting a house (0%) 3 = Building houses on others' lands (6.51%) 4 = Owning their houses (92.87%)
	house_condition (C)	house conditions	1 = Need urgent repair (3.32%) 2 = Need remedial action. (38.05%) 3 = No repair needed. (58.64%)
	house_cleanlines s (C)	cleanliness and organizing belongings of households	0= Messy (7.57%) 1= Not messy (92.43%)
	indoor_sewage (C)	households with an indoor sewage system	0=No (16.90%) 1=Yes (83.10%)
	toilet_sanitation (C)	toilets in households with healthy and sanitation conditions	0=No (5.82%) 1=Yes (94.18%)
	waste_separation (C)	households with waste separation	0=No (9.89%) 1=Yes (90.11%)
	electricity_house (C)	households with electricity	0=No (0.75%) 1=Electricity supplied from another house. (1.56%) 2=Yes (97.68%)
	tap_water (C)	households with water supply	0=No (72.09%) 1=Yes (27.91%)
	drinking_water (C)	households buying drinking water	0=No (21.53%) 1=Yes (78.47%)
	mobile_phone (C)	households having mobile phones	0=No (13.33%) 1=Yes (86.67%)
	computer_ ownership (C)	households with computers	0=No (90.30%) 1=Yes (9.70%)
	IT_welfare (C)	households utilizing technology to access state welfare services	0=No (37.23%) 1=Yes (62.77%)
	IT_income (C)	utilizing technology to increase household incomes	0=No (57.81%) 1=Yes (48.19%)
Human capital (18 features)	member_below_ 15 (N)	households with members' ages ranging over 15 years old	Max:7 Min:0 Mean:0.75
	skills_number (N)	number of household members having diverse skills in professions	Max:2 Min:0 Mean:1.16
	education_level (C)	households having individuals with the highest level of education	0 = No schooling education/non-completion of primary level (4.51%) 1 = Elementary school level (37.30%) 2 = Lower secondary school level (22.47%) 3= Secondary school

capitals	Attributes	Explanation	Value
			level or vocational certificate (23.59%) 4 = Diploma or higher vocational degree (3.82%) 5= Bachelor's degree (7.13%) 6 = Higher than a bachelor's degree (1.19%)
	employed_number (N)	working age in the household (15-59 years)	Max:7 Min:0 Mean:1.99
	farming (C)	farming households	0=No (26.47%) 1=Yes (73.53%)
	general_hired (C)	households with general hired occupation	0=No (77.60%) 1=Yes (22.40%)
	agriculture_employ (C)	households' employment in the agriculture sector	0=No (91.36%) 1=Yes (8.64%)
	self_employed (C)	households with self-employed	0=No (93.55%) 1=Yes (6.45%)
	fishery (C)	households with fishery	0=No (98.87%) 1=Yes (1.13%)
	civil_services (C)	households with civil services	0=No (94.24%) 1=Yes (5.76%)
	contract_employee_in_government_sector (C)	household members working as government employees	0=No (96.75%) 1=Yes (3.25%)
	private_employee (C)	household members working in private sectors	0=No (86.11%) 1=Yes (13.89%)
	disabled_number (N)	householders with disabled members with self-reliance	Max:2 Min:0 Mean:0.06
	bedridden_number (N)	household members with bedridden old adults and disabled adults with no self-reliance	Max:2 Min:0 Mean:0.02
	chronic_number (N)	household members with chronic illnesses	Max:4 Min:0 Mean:0.22
	healthy_number (N)	households with healthy members	Max:10 Min:0 Mean:2.96
	welfare_card (C)	households' members possessing public welfare card	0=No (10.95%) 1=Yes (89.05%)
	elderly_number (N)	elderly household members aged over 60 years old	Max:3 Min:0 Mean:0.64
Natural capital (2 features)	natural_living (C)	households using natural resources for livelihood, such as mushrooms, firewood, forest plants, edible insects	0=No (16.08%) 1=Yes (83.92%)
	natural_income (C)	households using natural resources to earn income such as honey, herbs, mushrooms	0=No (32.17%) 1=Yes (67.83%)
Social capital	join_community_group (C)	households joining a community, such as occupational groups, finance groups, social welfare groups	0=No (2.25%) 1=Yes (97.75%)

Data type: N = numerical; C = categorical

B. Data Preprocessing

Data preprocessing is a crucial phase, aimed at simplifying data complexity and enhancing data quality before applying data mining algorithms. This phase encompasses four activities: attribute selection, data cleaning, data transformation, and data scaling.

1) Attribute selection: The dataset contains numerous attributes, some of which are irrelevant. This phase focuses on reducing the dataset size by eliminating irrelevant attributes. For example, features related to social capital that describe community characteristics and non-significant attributes, such as those indicating households without relevant information, were removed.

2) Data cleaning: Data cleaning involves filling in missing values and enhancing the data process of cleaning by filling in missing values. Numeric features are imputed with the average value from the same group, while categorical features are replaced with constant values. For example, if the land size of ownership rights is null, the null values are filled with the average value of the land size of all ownership rights or adjusted welfare allowance for the elderly on age brackets: individuals aged 60-69 received 600 Baht/month, those aged 70-79 received 700 Baht/month, those aged 80-89 received 800 Baht/month, and individuals aged 90 and above received 1,000 Baht/month.

3) Data transformation: Data transformation is used to convert textual information into numerical values for analysis. For example, "farmer" is represented as 1, while "non-farmer" represents 0, the highest education level of all household members, the count of elderly individuals aged 60 and above, as well as the count of working-age individuals between 15 and 59, are calculated.

4) Data scaling: Min-max normalization [21] is a scaling technique in which value rescaled data in a range of 0 to 1 using the formula in (1). The technique is applied to specific numerical attributes, such as the number of households, income, and expenses.

$$X' = (X - X_{min})/(X_{max} - X_{min}) \tag{1}$$

where, X_{max} and X_{min} are the maximum and the minimum values of a feature, respectively.

C. Clustering

Clustering is the process of grouping data with similar characteristics, where clusters exhibit higher similarity within and differ from other clusters. In this paper, the K-prototype is applied to dealing with both numerical and categorical data. This algorithm combines numerical and categorical data to form clusters. The clusters represent 4 categories of poverty, ranging from the most impoverished to the less poor: Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor, according to the PMUA classification. Poverty household status is labelled on each record and defined as a target attribute for creating a model. The resulting cluster offers valuable insights for community planning and implementation.

D. Model Construction

The feature selection process aims to reduce data dimensionality by removing irrelevant features. The algorithm, MI, ReliefF, REF, and SFS are used to select features from the dataset for classification. Relevant features or predictive features are selected by removing the irrelevant features. After selecting the features, the Decision tree classifier builds the model based on several predictive features.

1) Feature selection techniques: The goal of feature selection is to find the optimal feature subset. By eliminating irrelevant features, the number of features can be reduced, the accuracy of the model can be improved, and the running time can be reduced [22].

a) MI: MI is used as a measure of the relationship between a feature and the target output. The higher the value, the more strongly relevant between a feature and the target, which suggests that the feature is selected. If the score is 0 or very low, then a feature and the target are weakly relevant [23].

b) ReliefF: ReliefF is a filter method of the feature selection algorithm. It finds the weights of features in the case where y is a multiclass categorical variable. According to the correlations between features and targets, different weights are assigned to each feature, and the feature with a slighter weight greater than a certain threshold will be removed [24].

c) RFE: RFE is an algorithm to select features in a training set that are more relevant in predicting the target output and removing weak features. The RFE works by searching for a set of features by starting with all features in the training dataset and selecting the most significant features by finding a high correlation between features and target output. Such recursion works until the number of remaining features reaches the desired number [25].

d) SFS: SFS is an algorithm that selects features from the set of features and evaluates them for a model iterate number between the different sets by reducing and improving the number of features so that the model can find the optimal performance and results. The SFS starts with one feature and adds more iteratively [26].

2) Classification: A Decision tree is a technique to build a predictive model through two phases: the training phase builds a model from a training set with a labelled target output, and the testing phase finds the quality of the trained model from the testing set without labelled target output. The model is like a tree structure. The nodes represent the features, the branches represent the decision rules, and the leaf node represents a poverty household status.

E. Model Evaluation

The performance is evaluated based on the calculation of F-measure, precision, and recall using the confusion matrix. It consists of four elements: true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [27].

TP is a condition when the observations coming from positive classes are predicted to be positive. TN is a condition when observations from negative classes are predicted to be negative. FP is a condition when the actual observation comes from negative classes but is predicted to be positive. FN is a condition when the actual observation comes from a positive but in a positive-negative predicted class.

The performance of the experiments is represented using precision, recall, and F-measure that are evaluated using Eq. (2) to Eq. (4), respectively.

$$Precision = TP/(TP+FP) \quad (2)$$

$$Recall = TP/(TP+FN) \quad (3)$$

F-measure = (2*precision*recall) / (precision+recall) (4).

IV. Results

A. Clustering

The clustering algorithm divided the households into 4 clusters. Table II summarizes the cluster based on the important attributes. Cluster 4 constitutes the largest group containing 31.9%. While, the smallest cluster is Cluster 3, making up 19.1% of the entire cluster. The distinct characteristics of each group (four groups) according to the features to formulate the label were addressed below.

TABLE II. THE CLUSTER OF POVERTY HOUSEHOLDS

Cluster	Instance	Percentage	Cluster name
1	397	24.8	Vulnerable non-poor
2	387	24.2	Extreme poor
3	305	19.1	Moderate poor
4	509	31.9	Destitute
Total	**1,598**	**100.0**	

The PMUA labels are assigned post-clustering, and their validity hinges on the methodology employed for the assignment. We used a procedure considering inherent cluster characteristics and, when available, external domain knowledge for labeling. To validate these labels, we cross-referenced them with established poverty classification criteria and assessed alignment with expected poverty attributes. Additionally, we ensured label consistency within clusters by employing quality assessment metrics, aiming to uphold the accuracy and reliability of the assigned PMUA labels.

Group classification using a clustering algorithm resulted in four clusters as shown in Fig. 2.

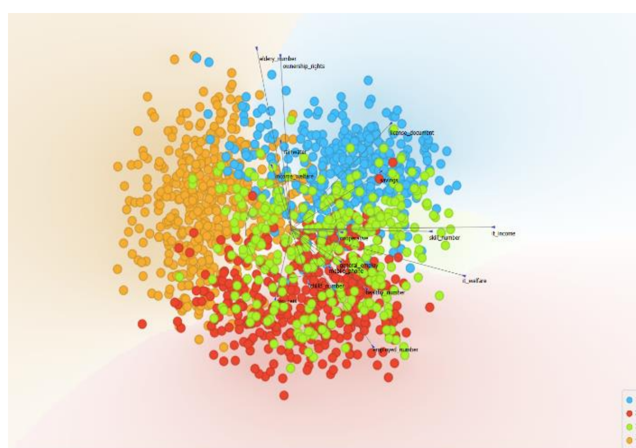

Fig. 2. Freeviz visualization.

Fig. 2 from orange software is used to show the characteristics of four different classes in different colors. Each class describes different characteristics of households. Each axis corresponds to different features; the length of each axis corresponds to the importance of the feature. Cluster 1 is grouped within a blue cluster that is characterized by government lease, savings, and rainwater. Cluster 2 is grouped within a red cluster that is characterized by employed_number, others_rent_free, member_below_15, mobile_phone, general_hired, and healthy_number. Cluster 3 is grouped within a green cluster that is IT_welfare, IT_income, and skills_number, while Cluster 4 is grouped within an orange cluster that is the elderly_number, income_welfare, and ownership_rights. The distribution of instances in Fig. 2 revealed that the instance of Cluster 3 (green area) had fewer colored areas, compared to other clusters. Some instances were mixed up in other clusters and many more were mixed up in Cluster 2 and Cluster 1.

The cluster of the data group divided into four clusters in TABLE II revealed that Cluster 4 had instances more than other clusters. Cluster 1 and 2 had similar cluster stances. Cluster 3 had the least instances. The remaining clusters exhibited similarity in the instance of all clusters, indicating that all householders included members of labor force age, school-age children, elderly, patients with chronic illness, bedridden patients, and disabled members with no self-reliance, disabled members. Most households owned mobile phones. The specific characteristics of each cluster are shown in Fig. 3 as follows:

Cluster 1 included 397 households that were the groups obtaining income from remittances and state welfare programs at the highest mean score, which was more than other clusters. These households also had an average monthly income at the second ranking. The income shows no difference, compared to the cluster with the highest income. The number of elderly and disabled members was at the highest level. The sizes of arable lands with legal land rights documents for ownership, and the legal land rights documents were more than other clusters. Most households completed their education at the elementary school level. The households exhibit the highest savings and use rainwater the most. However, those with features closely aligned to Cluster 3 are referred to as the “Vulnerable non-poor group”.

Cluster 2 included 387 households with an average income from off-farming sectors at the highest level. The average household income from farming sectors and state welfare programs was at the lowest level. The average household expenses ranked second; the labor force age members were at the second-ranking. The arable land sizes were at the lowest rank. Most arable lands had legal rights documents. Most household members completed lower secondary education or vocational certificate education. The households have the highest number of mobile phones, with a quantity similar to Cluster 3. The households have a high number of members under 15 years old, which is also close to that of Cluster 3. This group is referred to as the “Extreme poor group”.

Cluster 3 included 305 households with the average income at the highest level. The average income earned from farming sectors was at the highest level. The income from off-farming sectors was ranked second. Diverse skills in professions were at the highest level. The average number of labor force age and patients with chronic illness were at the highest level. The arable lands of all types were ranked second. Most household members completed their lower secondary education. The households use technology for state welfare applications and income generation, with proximity similar to Cluster 2. This group is referred to as the “Moderate poor group”.

Cluster 4 included 509 households with the average income at the lowest level. The income from off-farming sectors was at the lowest level. The healthy members and the labor force age members were both at the lowest level. The number of elderly was higher than young age members. The diverse skills for professions were at the lowest level. Most household members completed the elementary education level. The average expenses were at the lowest level. The average number of bedridden patients was more than in other clusters but equal to Cluster 1. The average number of chronically ill patients was at the second rank. The average number of disabled members was ranked second to Cluster 1, and the arable lands had the legal land rights documents for ownership. This group is referred to as the “Destitute group”.

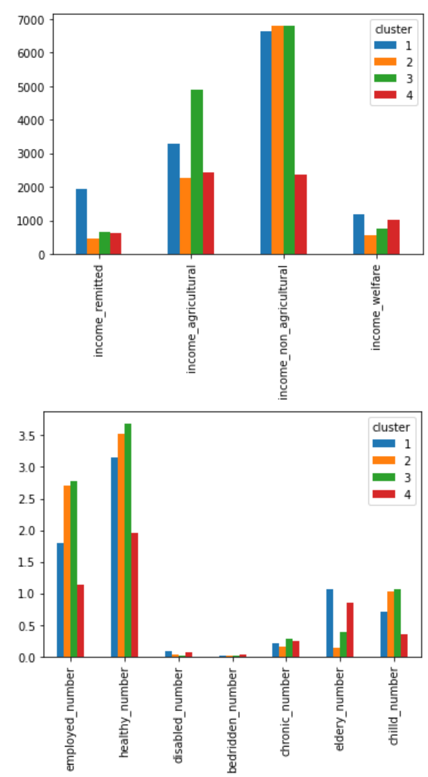

Fig. 3. Data comparison among clusters based on specific characteristics.

B. Feature Selection

The poverty household dataset was classified into four poverty statuses, namely Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor. The values were stored in a column named “target”. Feature selection has been implemented using MI, ReliefF, RFE, and SFS algorithms. These algorithms were implemented in the Scikit-learn library for the selection of optimal features. The number of employed features by the MI algorithm was a high value. The top 20 features were selected to predict the model, shown in Table III. ReliefF calculates a feature score for each feature which can then be applied to rank and select the top 20 scoring features for feature selection, as shown in Table III.

TABLE III. THE SELECTED FEATURES BY EACH ALGORITHM

Features	MI	ReliefF	RFE	SFS	Common
employed_number	✓	✓	✓		✓
IT_welfare	✓	✓	✓	✓	✓
IT_income	✓	✓	✓		✓
ownership_rights_rai	✓	✓			✓
ownership_rights	✓	✓	✓	✓	✓
elderly_number	✓	✓	✓	✓	✓
education_level	✓	✓	✓	✓	✓
skills_number	✓	✓	✓	✓	✓
government_lease_rai	✓	✓	✓		✓
healthy_number	✓	✓	✓		✓
savings	✓	✓	✓	✓	✓
rainwater	✓	✓			✓
no_access_water	✓	✓	✓	✓	✓
expenses	✓	✓	✓		✓
government_lease	✓	✓	✓	✓	✓
cattle	✓	✓	✓		✓
fertile_soil	✓				
income_welfare	✓	✓	✓		✓
loan_state	✓				
rice_farming	✓				
natural_income		✓	✓	✓	✓
home_conditions		✓	✓		✓
child_number		✓	✓		✓
income_farming			✓		
self_employed				✓	
reservoirs				✓	
loan_community				✓	
loan_saving_bank				✓	
loan_commercial_bank				✓	
loan_private				✓	
loan_informal_debt				✓	
student_loan				✓	
welfare_card				✓	
irrigation_canals				✓	
mobile_phone				✓	
bedridden_number				✓	
Total	**20**	**20**	**19**	**21**	**20**

The implementation of the RFE algorithm using the Decision tree classifier on the training set and five cross-validations was performed. The algorithm determined the optimal number of features, and 19 features were selected, indicating the importance of these features on the poverty dataset. The implementation of the SFS algorithm using the Decision tree classifier on the training set and five cross-validations was performed. The algorithm found the best score of 21 features.

Table III shows the selected features that have the most effect on the poverty level. The selected features were considered relevant by at least two selection algorithms. All algorithms selected eight features, and seven features were selected by three algorithms. The five features were selected by 2 algorithms and the other features were selected by only one algorithm. The set of features consists of 20 features, called common features.

C. Model

In this paper, the experiments were divided into 6 experiments that employed different feature selection techniques on a dataset of 1,598 impoverished households.

The selected features gained from the feature selection phase are used to train the Decision tree classifier. The number of features selected by the several types of feature selection algorithms is presented in Table IV. In the processing model, 10-fold cross-validation is applied. Nine folds were used for training and the remaining fold was used for testing. The process was repeated 10 times. The performance metrics such as F-measure, precision, and recall are measured to demonstrate the results and comparative analysis of the feature selection algorithms. The metrics consider the entire features, common features, and the feature set obtained by applying the feature selection techniques. Performance evaluation is given in Table IV.

From Table IV, the performance results varied significantly based on the size of selected features and characteristics attributes. The particular data characteristics associated with SFS using 12 features differed from those of other techniques, thus yielding diverse performance outcomes.

The SFS algorithm has achieved the best performance with F-measure, precision, and recall at 74.6%, 74.8%, and 74.7%, respectively. Thus, the SFS algorithm performed better than other feature techniques. Besides in Table V, a confusion matrix presented the performance of the Decision tree classifier with selected features by SFS algorithm.

TABLE IV. PERFORMANCE OF DECISION TREE CLASSIFIER

Feature selection techniques	No.of features	F-measure	Precision	Recall
All	76	72.0	72.0	72.2
MI	20	71.4	71.5	71.6
ReliefF	20	71.4	71.5	71.6
REF	19	73.6	73.7	73.6
SFS	21	74.6	74.8	74.7
Common features	20	58.9	59.2	58.9

TABLE V. CONFUSION MATRIX OF THE RESULTS OBTAINED BY THE DECISION TREE CLASSIFIER WITH SFS ALGORITHM

Actual	***Predicted***			
	Vulnerable non-poor	***Extreme poor***	***Moderate poor***	***Destitute***
Vulnerable non-poor	74.2%	5.2%	7.8%	6.6%
Extreme poor	6.7%	73.1%	9.8%	7.5%
Moderate poor	8.5%	8.5%	69.5%	6.2%
Destitute	10.6%	13.2%	12.9%	79.7%

From the decision tree model, the total number of trees is 587 nodes, which interprets 293 rules, and the depth is 16. The example tree model showed depths of trees, represented in Fig. 4. The most important feature, is whether using technology to request state welfare benefits is or not (IT_welfare), classified into nodes for with (IT_welfare=1) and without (IT_welfare=0) using technology. The left node shows the high education level of members in the household (education_level), the next level is arable lands with the legal land rights documents (ownership_rights), and the number of older people (elderly_number). The model from the Decision tree classifier can be interpreted and understood in the decision tree rules format. Table VI shows the example of decision tree rules. The relationship rules were able to describe the rule and class.

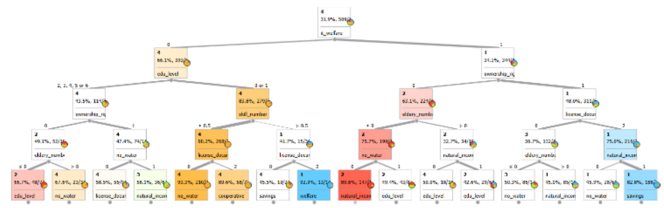

Fig. 4. Decision tree model with 5-level depth.

TABLE VI. DECISION TREE RULES MODEL APPLYING RESULTS

Rules	**Condition (If)**	**Class**	**Description**
1	IT_welfare = 1 and ownership_rights = 1 and government_lease = 0 and elderly_number > 0 and natural_income = 1 and skills_number > 0 and no_access_water = 1 and loan_savings_bank = 0 and education_level in (1,3,4,6) and reservoir = 0	Vulnerable non-poor	IF households utilize technology for requesting the state welfare programs AND have aroma lands with legal land rights documents of ownership AND have aroma lands without legal land rights documents AND have elderly members AND utilize natural resources to earn income AND have skills for professions AND have arable lands with no access to water supply AND no loan from the Savings Banks AND have their highest level of education lower than bachelor's degree AND have access to water supply from reservoirs for farming
2	IT_welfare = 1 and ownership_rights = 0 and elderly_number <= 0 and no_access_water = 0 and natural_income = 1 and education_level in (0,3,4,5,6)	Extremely poor	IF households utilize technology for requesting state welfare programs AND have aroma lands with legal land rights documents of ownership AND no elderly members AND no access to water supply for farming AND utilize natural resources to earn income AND have their highest level of education lower than bachelor's degree
3	IT_welfare = 0 and education_level in (2,3,4,5,6) and ownership_rights = 1 and no_access_water = 1 and natural_income = 0 and savings = 1 and elderly_number <= 0.33	Moderate poor	IF households utilize no technology for requesting the state welfare programs AND have their highest level of education lower than a bachelor's degree AND have aroma lands with legal land rights documents of ownership AND no access to water supply for farming AND no utilizing natural resources to earn income AND have savings AND have at least one elderly member
4	IT_welfare = 0 and ownership_rights = 1 and no_access_water = 0 and government_lease = 0 and savings = 0 and self_employed =0 and education_level = 2	Destitute	IF households utilize no technology for requesting the state welfare programs AND have aroma lands with legal land rights documents of ownership AND no access to water supply for farming AND having no aroma lands with legal land rights documents AND having no savings AND no self-employed business AND have their highest level of education at lower secondary education

V. DISCUSSION

Upon analyzing the factors affecting the classification of poverty levels based on livelihood capital, aiming to construct poverty indicators for classifying poverty levels, it was found that factors derived from the SFS technique and Decision tree classifier yielded the highest F-measure. These factors comprised five factors within human capital, seven factors within physical capital, six factors within economic capital, and 1 factor within natural capital. Nevertheless, there were no factors identified within the social capital category.

Features affecting the predicting of the poverty level, such as education_level, and skills_number, which were features related to the competencies of household members. The study in [28] suggests that the factors influencing poverty in Thailand are linked to the rise in per capita income and education, both showing statistical significance at the 0.10 level. The household members who completed at a basic education level or higher than the basic education level had abilities to earn more income and were able to scaffold knowledge to increasingly improve skills. These could reduce poverty in households.

The feature called IT_welfare for receiving welfare resources from government sectors of Thailand. Some welfare allowances were able to be accessed through the online register, receiving state welfare resources through applications, household members having technological skills and mobile phones. Thus, the household members were able to access state welfare resources and received assistance from the government sectors following the established conditions. The study referenced in [29] shows that the Internet has a significant impact on alleviating the vulnerability to poverty among rural households.

The features, ownership_rights, and government_lease, were related to arable lands of poor households because the land is a critical asset, the primary for generating a livelihood, and a main vehicle for investing, especially for the poor as it provides a means of livelihood through the production and sale of crops and other products. The studies in [30, 31] suggest that land can serve as collateral for credit to invest in the land or be exchanged for capital to start up another income-generating activity. The study in [32] suggests that the absence of land ownership may contribute to a high fertility rate, low capital investment, and consequently lower living standards.

The feature called no_access_water was related to water supply for farming was limited. Most farmers use rainwater as a major waste resource for agriculture. Some areas have water resources close to their arable lands, such as canals, and ponds. Thus, the increasing amount of water for farming areas is important for farmers to be able to farm yearly.

The elderly, children aged from 0-14 years of age, chronically ill, and bedridden patients living in households are vulnerable groups, who are key features affecting poverty because many households, while not currently in poverty, recognize that they are vulnerable to events - a bad harvest, a lost job, an illness, and unexpected expenses, an economic downturn - that could easily push them into poverty [33]. A vulnerability group is a group with no income or few incomes that rely on family members or households with members in the labor force, but rely on the state welfare system, such as the subsidy support budget program for newborn babies to six years old, the subsistence allowance provision for disabled persons. These support household expenses and alleviate poverty for vulnerable groups [34] or households would afford health care if the households had members with inpatients, households with members aged over 65 years, and households with disabilities [35].

VI. Conclusion

The conclusion concisely outlined our key findings, insights, and results, demonstrating the importance and implications of our work in addressing the issue.

The research employed both supervised and unsupervised learning techniques to analyze data on poor households. Initially, unsupervised learning divided the data into four clusters representing different poverty levels: Destitute, Extreme poor, Moderate poor, and Vulnerable non-poor groups. Using supervised learning, the study identified key features influencing poverty levels, employing algorithms like MI, ReliefF, RFE, and SFS. The decision tree model using the SFS algorithm proved most effective, achieving 74.6% F-measure, 74.8% precision, and 74.7% recall in predicting poverty levels. The study then summarized the characteristics of each poverty level in the Kut Bak district:

Destitute Households: These households earn an average monthly income of 6,426 Baht; heavily rely on state welfare due to limited development opportunities, with elderly members, minimal education levels, poor health, and reliance on government services.

Extreme Poor Households: With monthly income averaging around 10,088 Baht, these households mostly derive earnings from non-agricultural sources due to limited arable lands, facing excessive expenses with minimal savings, but possessing skills to access welfare resources.

Moderate Poor Households: These households earn an average income of 13,100 Baht monthly from both farming and off-farming sectors, facing substantial expenses on healthcare, but managing some savings. Lack of water access impacts household development.

Vulnerable Non-Poor Households: This group earns an average of 13,030 Baht monthly, primarily from agricultural sources and remittances, owning considerable arable lands, providing opportunities for an improved quality of life compared to other clusters.

In future work, we will create applications to predict households of poverty in four levels with the conditions drawn from the decision tree and using the class of poor households in the area of Kut Bak district. The leader of the community was a person who selected households following the characteristics of four clusters. After entering the data through the applications for predicting poor households to determine the accuracy between humans and machines and create the acceptance of the results from the predicting with the relevant organization in the areas to apply it for actual practices.

Acknowledgment

This research is part of the “Precision Poverty Alleviation in Sakon Nakhon Province” awarded as a research scholarship from the Program Management Unit on Area Based Development (PMUA) in 2022. Thank you to all households in Kut Bak district, Sakon Nakhon for collecting data.

Ethical Consideration

The data was collected from households in Kut Bak district in Sakon Nakhon located in the northeastern region of Thailand in 2022. Ethics approval for the fieldwork was obtained through the Human Research Ethics Committee at Sakon Nakhon Rajabhat University. Guarantee No. HE 65-095.

References

[1] United Nations, The Sustainable Development Goals Report 2022, New York: United Nations, 2022.

[2] Community Development Department, “Program for Collected, Storage, and Process in 2023 - 2027,” 2023. https://smartbmn.cdd.go.th (accessed Jun. 15, 2023).

[3] National Science and Technology Development Agency, “TPMAP Thai People Map and Analytics Platform,” 2018. https://www.tpmap.in.th/ (accessed Jun. 15, 2023).

[4] Program Management Unit on Area Based Development, “Practical Poverty Provincial Connext: PPPConnext,” 2019. http://www.ppaos.com/ppaos47/frontend/web/ (accessed Aug. 1, 2023).

[5] B. Reddy, K. Srikanya, M. Varshini, S. Srijanya, G. Reddy and C. Shirisha, “The Application of Machine Learning to the Task of Poverty Classification,” 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-4, doi: 10.1109/ACCAI58221.2023.10201169.

[6] H. A. Silva Marchan, O. J. M. Peña Cáceres, D. M. Ricalde Moran, T. Samaniego-Cobo, and C. M. Perez-Espinoza, “A Machine Learning Study About the Vulnerability Level of Poverty in Perú,” Communications in Computer and Information Science, Springer, Cham, vol. 1658, 2022, doi: https://doi.org/10.1007/978-3-031-19961-5_1.

[7] K. M. Kim, J. H. Kim, H. S. Rhee, and B. Y. Youn, “Development of a Prediction Model for the Depression Level of the Elderly in Low-income Households: using Decision Trees, Logistic Regression, Neural Networks, and Random Forest,” Sci Rep, vol. 13, July 2023, doi: https://doi.org/10.1038/s41598-023-38742-1.

[8] N. Istiqamah, O. Soesanto, and D. Anggraini, “Application of the K-Means Algorithm to Determine Poverty Status in Hulu Sungai Tengah,” J. Phys, vol. 2106, 2021, doi: 10.1088/1742-6596/2106/1/012027.

[9] P. R. Michelle and C. R. Rex Aurelius, “Apply Clustering Algorithm on Poverty Analysis in a Community in the Phillippines,” Proceeding of the international conference in industrial engineering and operations management Monterrey, Mexico, 3-5 November 2021, pp. 1511-1521.

[10] S. Kuptabut and S. Songleknok, “The Analysis of Factors Affecting the Sustainable Livelihoods of Poverty Households using Data Mining Technique,” 2022 International Conference on Digital Government Technology and Innovation (DGTi-CON), Bangkok, Thailand, 2022, pp. 20-23, doi: 10.1109/DGTi-CON53875.2022.9849207.

[11] United Nations Development Programme and Oxford Poverty and Human Development Initiative, 2023 Global Multidimensional Poverty Index (MPI): Unstacking global poverty: Data for high impact action, New York: UNDP, 2023.

[12] World Bank, “Poverty & Equity Brief Thailand East Asia & Pacific April 2023,” 2023. https://databankfiles.worldbank.org/ (accessed Jun. 20, 2023).

[13] Y. Bouchlaghem, Y. Akhiat, and S. Amjad, “Feature Selection: a review and comparative study,” E3S Web of Conferences, vol. 351, 2022, doi: https://doi.org/10.1051/e3sconf/202235101046.

[14] A. Alsharkawi, M. Al-Fetyani, M. Dawas, H. Saadeh, and M. Alyaman, “Improved Poverty Tracking and Targeting in Jordan Using Feature Selection and Machine Learning,” IEEE Access, vol. 10, pp. 86483-86497, 2022, doi: 10.1109/ACCESS.2022.3198951.

[15] P. Kambuya, “Better Model Selection for Poverty Targeting through Machine Learning: A Case Study in Thailand,” Thailand and the World Economy, vol. 38, pp. 91-116, 2020.

[16] J. A. Talingdan, “Data Mining using Clustering Algorithm as Tool for Poverty Analysis,” Proceedings of the 2019 8th International Conference on Software and Computer Applications (ICSCA), 2019, pp. 56-59.

[17] G. Wong, “Poverty Prediction and the Identification of Discriminative Features on Household Data from Cambodia,” TechRxiv, pp. 1511-1521, doi: https://doi.org/10.36227/techrxiv.21266226.v1.

[18] L. Soochang and K. Daechan, “Decision Tree Analysis for Prediction Model of Poverty of the Older Population in South Korea,” International Journal of Advanced Culture Technology, vol. 10, no. 2, pp. 28-33, 2022.

[19] N. S. Sani, M. A. Rahman, A. A. Bakar, S. Sahran, and H. M. Sarim., “Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification,” International Journal Advanced Science Engineering Information Technology, vol. 8, no. 4-2, pp. 1698-1705, 2018.

[20] C. Kallestal, E. Blandon Zelaya, R. Pena, W. Perez, M. Contreras, L. A. Persson, O. Sysoev, and K. Ekholm Selling, “Predicting poverty Data mining approaches to the health and demographic surveillance system in Cuatro Santos, Nicaragua,”. Int J.Equity health, vol. 18, no. 165, 2019.

[21] J. Han, M. Kamber, and J. Pei, “Data transformation and data discretization. In Data Mining-Concepts and Techniques; Kaufmann,” M., Ed.; Elsevier: Amsterdam, The Netherlands, 2011, pp. 111–112.

[22] K. Kira and L. A. Rendell, “A practical approach to feature selection,” 9th International Conference on machine learning, 1992, pp 249-256.

[23] J. R., Vergara and P. A. Estévez, “A review of feature selection methods based on mutual information,” Neural Comput & Applic, vol. 24, pp. 175–186, 2014.

[24] M. Robnik-Sikonja and I. Kononenko, “Theoretical and empirical analysis of ReliefF and RReliefF,” Machine Learning, vol. 53, no. 1–2, pp. 23–69, 2003.

[25] E. M. Senan, M. H. Al-Adhaileh, F. W. Alsaade, T. H. H. Aldhyani, A. A. Alqarni, N. Alsharif, M. I. Uddin, A. H. Alahmadi, M. E. Jadhav, and M. Y. Alzahrani, “Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques,” Journal of Healthcare Engineering, vol. 2021, June 2021, doi: https://doi.org/10.1155/2021/1004767.

[26] R. Gutierrez-Osuna, “Pattern Analysis for Machine Olfaction: a review,” IEEE Sensors Journal, vol. 2, no. 3, pp. 189-202, June 2002, doi: 10.1109/JSEN.2002.800688.

[27] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, “Selecting Critical Features for Data Classification based on Machine Learning Methods,” J Big Data, vol. 7, no. 52, July 2020, doi: https://doi.org/10.1186/s40537-020-00327-4.

[28] S. Chotikhamjorn, “The Study of Poverty and Factors Affecting Poverty in Thailand,” Management Science Review, vol. 23, pp. 63-71, July – December 2021.

[29] S. Zhang, Q. Liu, X. Zheng, and J. Sun, “Internet Use and the Poverty Vulnerability of Rural Households: From the Perspective of Risk Response,” Sustainability, vol. 15, no. 2, 2023.

[30] R. Meinzen-Dick, DESA Working Paper No. 91: Property Rights for Poverty Reduction?, New York: United Nations, 2009.

[31] K. Deininger, Land and Policies for Growth and Poverty Reduction: a World Bank policy research report, Washington: Oxford University, 2003.

[32] L. Cellarier, “Is land ownership a ladder out of poverty?” World Development, vol. 146, 2021.

[33] P. Lant, S. Asep, and S. Sudarno, Quantifying Vulnerability to Poverty: A Proposed Measure, Applied to Indonesia. Policy Research Working Paper; No. 2437, World Bank: Washington, DC. 2000.

[34] T. Jiratitikulchai, S. Lilehhannot, and P. Fungpiriya, Discussion Paper No.67: Economic Status of Vulnerable Households and Considerations for Social Protection. Bangkok: Thammasat university, 2022.

[35] N. Wang, J. Xu, M. Ma, L. Shan, M. Jiao, Q. Xia, W. Tian, X. Zhang, L. Liu, Y. Hao, L. Gao, Q. Wu, and Y. Li, “Targeting Vulnerable Groups of Health Poverty Alleviation in Rural China— What is the Role of the New Rural Cooperative Medical Scheme for the Middle Age and Elderly Population?,” Int J Equity Health, vol. 19, pp. 1-13, 2020.