Predicting the Poverty Alleviation in the Province of Eastern Samar using Data Mining Techniques

Jared Harem Q. Celis, Andres C. Pagatpatan, Jr.

Abstract*: *Poverty has been a main concern for century in any part of the world. The abrupt increase of population in the country and the inevitable rise of the inflation rate due to the economic challenges and other factors, it is clearly manifested that poverty is a problem that needs to be addressed seriously. With the available various advanced-technology nowadays, this problem on poverty maybe reduced with the aide of Data Mining which is a part of Data Science. This paper focused on predicting the poverty alleviation using Data Mining techniques based from all available data from the Philippine Statistic's Authority, National Economic Development Authority, and Department of Social Welfare and Development. The application of supervised learning in Data Mining specifically, NaiveBayes Algorithm, Decision Tree J48 Algorithm, and K- Nearest Neighbour Algorithm has been utilized for the prediction of poverty alleviation in the province of Eastern Samar. The results of this study unveil that among the core indicators in identifying poverty, it is the "Economic Sector" with the attribute "Income" is the most significant factor that affects poverty alleviation in the province.

Keywords : Classification Model, Data Mining, Poverty Alleviation, Poverty Incidence, Prediction.

I. INTRODUCTION

Poverty is mainly due to the very low earning capacity of the least fortunate individuals. There are two interrelated root cause behind these things, one because on no access or low education, and the insufficiency of productive opportunities of job. The labor sector is segmented into "bad" and "good" occupations, with the poor working in the latter. They have jobs that are blue-collar, temporary or casual, and minimum or low paid. Prevalent informality means that the least fortunate or poor neither benefit from the minimum wage policy nor from the employment protection legislation [1].

The comparatively low performance in growth and increasingly poverty incidence in the Philippines has raised many serious concerns. The slow reduction of poverty in the country is simply due to the dawdling increase of income, or it is because of the weak strategies of poverty reduction that actually gives a growth rate in aggregate income. Furthermore, the Philippines has been known for a long time for its inequality in the distribution of income and wealth. This disparity in the wealth and income distribution has been identified as the reason for the slow growth of the economy and hindrance to poverty reduction. There is a literature about politics in the Philippines which mentioned that the ''oligarchic'' or incompetent political system in the country has been a primary hindrance for implementing progress oriented policy reforms and thus for poverty reduction [2].

The Philippine Statistics Authority (PSA) published its latest report about the official poverty statistics in the country for the first semester of 2015. The reports of the PSA shows the estimates of poverty incidence using income data taken from the Family Income and Expenditure Survey (FIES) which was conducted on July 2015. Poverty incidence in the Philippines was estimated at 27.9 percent in the year 2012 and 26.3 percent in the year 2012 just having a slight decrease of 1.6 percent in over three years [3].

The continuous increase of poverty incidence or extreme poverty among Filipino families, was estimated at 10 percent during the first semester of 2012. The proportion of families in extreme poverty in the same period of 2015, was recorded at 9.2 percent [3].

Likewise, continuous incidence among Filipinos whose any form of incomes fall below the food threshold, was estimated at 13.4 percent in the first half of the year 2012 while in the year 2015, it was estimated at 12.1 percent during the first semester. This continuous incidence in the Philippines is frequently referred to as the proportion of Filipinos in extreme or existence poverty [3].

During the first semester of the year 2018, the average income of the poor families were shorter by 6.1 percent against the poverty threshold. This means that based on the average estimates, an amount of Php 10,481 monthly income is needed by a least fortunate or poor family with five household members in order to move out of the poverty line during the first semester of the year 2018 [4].

Data mining is the process of extracting knowledge from huge volume of data. The primary reason of using data mining algorithms is to fetch significant and relevant information that can provide us better outcomes. Information mining tools are utilized to discover knowledge, patterns and relations between these data.

Manuscript published on 30 September 2019
***** Correspondence Author

Jared Harem Q. Celis, Information Technology Department, Eastern Samar State University – Guiuan Campus, Guiuan, Philippines. jaredharemcelis@gmail.com

Dr. Andres C. Pagatpatan, Jr., Administration, Eastern Samar State University – Guiuan Campus, Guiuan, Philippines. andrespagatpatan@yahoo.com

Retrieval Number: C6082098319/2019©BEIESP
DOI:10.35940/ijrte.C6082.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This technique integrate additionally a measurable and numerical model. The process of Data mining [5] is performed on raw data which is characterized in different forms like text, web image processing and visuals.

The very significant step in the process is to find the discovered patterns or knowledge from the gathered data by using Knowledge Discovery process and this includes various steps for the extraction of meaningful data [6].

The process of Data Mining will be employed to predict the poverty alleviation in the province of Eastern Samar whether the poverty incidence will increase or not in the next few years to come.

II. AIM OF PAPER

A. Objectives of the Study

This paper aims to analyze and provide a prediction of the poverty alleviation using Data mining techniques based on the available data from the Philippine Statistics Authority. This would also help the concerned government agency to prepare strategies or plans that would resolve issues on poverty based on the result of this study.

B. Literature Review

Many researchers have used exclusive statistics and mining techniques for predicting student information in educational institutions. They have explored some of techniques and carried out various algorithms, specifically decision making trees, various regression methods and finally they have also given a few ideas and suggestions.

A study [7] used a group of classifiers to the pre-processed dataset and the researchers have made attempts to find out the best classifiers among the rest. The best or optimal solution is analysed through comparative analysis and a data cube that contains name, verbal ability and scores in maths have been considered as attributes. The ratings of the student are then saved independently in another cell. The classification analysis is then performed using the 3 attributes on the 2000 records of the students. A software called WEKA is used to perform the pre-processing of data and classification analysis. The software handles the missing value conveniently through imputation technique. Classification algorithm is under supervised learning which actually means that class labels are present.

Another Data mining model used in a study [8], which is clustering with the use of k-means algorithm. This is a very efficient tool to be used in monitoring student's performance in higher education. The researchers have implemented the traditional k-means clustering algorithm which actually uses the Euclidean distance formula in the analysis of the student's scores. They have also applied the Fuzzy clustering technique and then explained how these works when applied on data mining. They have used this method on the dataset taken from a University in Nigeria. The performance indexed that consist of five categories of students was framed. This means that there was five cluster that have been formed and the overall achievement of the students were then calculated.

Furthermore, a research [9] describes the different approaches in Data mining algorithms such as Neural network, K- Nearest Neighbour, Bayesian Classifier, Fuzzy Logic and decision tree classification algorithms for the implementation of Intrusion Detection System (IDS). With the help of the aforementioned different supervised learning algorithms in data mining, predictive analysis has been performed in Intrusion Detection System to detect malicious threats and unauthorized access in a computer system.

III. RESEARCH METHODS

A. Framework of the Study

The framework of this study is anchored on the natural Data Mining methods.

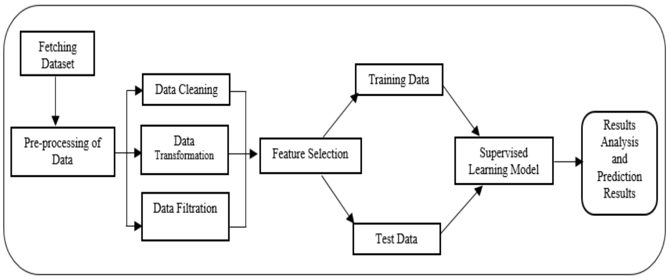


Fig. 1. Research Framework

The approach start with the collection of data or dataset of all households in the province of Eastern Samar from the databases of the PSA, NEDA, and DSWD, then later pre-processing procedures are done to remove and clean unnecessary data. The pre-processing of data is the very tedious part of the whole process and includes but not limited to the following: patching up fields with missing or null values using filtration, correction of attribute or column names, and fixing incorrect file format. After the raw data has been cleaned, it is transformed into a useful dataset. Then the pre-processed data are divided into two groups, namely testing dataset and training dataset. This is very common in Data Mining in order to gain knowledge. Usually, the seventy (70) percent of the total datasets are used as training dataset and thirty (30) percent are used as testing dataset [10].

Next, we have to choose among the supervised learning models such as J48 Algorithm, K- Nearest Neighbour, and NaiveBayes Standard Classification algorithms for better classification. The training and test datasets are feed into the Waikato Environment for Knowledge Analysis (WEKA) and analyzed in a form of decision tree and classification model. The results of these multiple decision trees are analyzed and processed until it reaches to its final decision tree. The visualization of decision tree generated is the attributed results of forecasting or predicting the poverty alleviation. With this, we can analyze whether the poverty alleviation will be reduced or elevated considering the historical data.

B. Classification Algorithms

The NaiveBayes algorithm is based on Bayes' Theorem that finds the likelihood of an event to happen given the probability of another event that had already occurred.

$$P\ (A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

The A is class variable and B is a dependent feature vector (size n) where $A = (x_1, x_2, x_3 \ldots\ldots, x_n)$
Next algorithm is the Decision Tree J48 classifier that has a structure that represents a tree. The potential information is calculated for every attribute, given by a test on the attribute. Then the gain in information is calculated that would result from a test on the attribute. The entropy of $\vec{y}$ is calculated by:

$$\boldsymbol{Entropy}(\vec{y}) = -\sum_{j=1}^{n} \frac{|y_i|}{|\vec{y}|} \log\left(\frac{|y_i|}{|\vec{y}|}\right) \tag{2}$$

$$\boldsymbol{Entropy}(j|\vec{y}) = \frac{|y_j|}{|\vec{y}|} \log\left(\frac{|y_j|}{|\vec{y}|}\right) \tag{3}$$

And then the Gain is calculated using the formula:

$$\boldsymbol{Gain}(\vec{y}, j) = \boldsymbol{Entropy}(\vec{y} - \boldsymbol{Entropy}(j|\vec{y})) \tag{4}$$

The objective is to maximize the Gain, then dividing by the total entropy due to split $\vec{y}$ by the value j.

The last algorithm used in this study is the KNN which is a non-parametric classification method which classify objects based on closest training examples in the feature space. This uses the Euclidean Distance formula to measure the distances between objects.

$$D(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (p_1 - q_1)^2} \tag{5}$$

C. Composition of Datasets

The data extracted from the database of the Philippine Statistic's Authority are cleaned and transformed into a file format readable by WEKA and then the pre-processed data is now known as dataset. The dataset has columns or fields in the table that corresponds to a variable or attributes and each row or instance of the table corresponds to each household's data of the dataset in query.

The dataset consists of four (4) poverty core indicator with its attributes. These includes the following:

1. **Health Sector**
   - Water Source
   - Sanitation
   - Malnutrition
   - Child Mortality
   - Maternal Mortality
   - Disability
2. **Education Sector**
   - School Dropouts
   - Illiteracy
3. **Economic Sector**
   - Income
   - Meal
   - Food Thresholds
   - Unemployment
4. **Social Sector**
   - Housing
   - Land Tenure Status
   - Crime Incidence

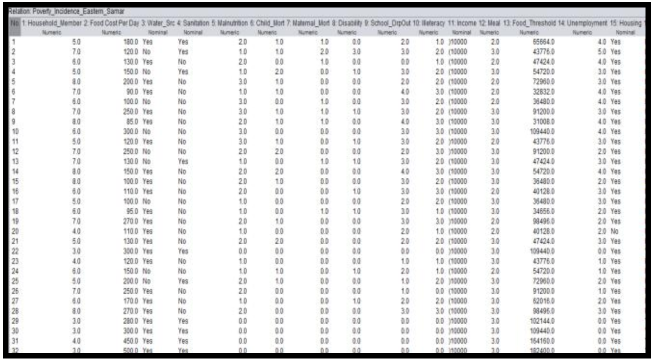
Table- I: Factors Affecting the Poverty

Attributes	Description	Possible Values
Household_Member	The number of household member	{1,2,3,4,5,6,7,8,9,10}
Water_Src	Water source in the household	{Yes, No}
Sanitation	Sanitation or cleanliness in the environment	{Yes, No}
Malnutrition	The number of child with malnutrition	{1,2,3,4,5,6,7,8,9,10}
Child_Mort	Mortality of children in the household	{1,2,3,4,5}
Maternal_Mort	Maternal mortality in the family	{1,2,3,4,5}
Disability	The number of disabled household member	{1,2,3,4,5}
School_DrpOut	The number of dropout household member	{1,2,3,4,5}
Illiteracy	The number of illiterate household member	{1,2,3,4,5}
Income	The monthly income of the family	{<10000, >10000}
Meal	The number of times they have meal	{1,2,3}
Food_Cost_Per_Day	Cost of food per day	{50,100,200,300,400,500}
Food_Threshold	Food threshold of the household per annum	{20000,30000,40000,50000,60000,70000,80000,90000}
Unemployment	The number of unemployed household member	{1,2,3,4}
Housing	Own/Rent a house	{Yes, No}
Land_Tenure_Stat	Has title/rights with the land	{Yes, No}
Crime_Incedence	Number of crime incidence within the family	{Yes, No}
Poor	Considered as Poor	{Yes, No}

The table above shows the eighteen (18) attributes which are divided into four (4) sectors namely Health Sector, Education Sector, Economic Sector, and Social Sector that affects the poverty in the province of Eastern Samar. It is also reflected the descriptions of each variables or attributes together with each corresponding possible values.

IV. RESULT AND DISCUSSION

After the dataset has been fed into the WEKA application, a couple of analysis has been done using different algorithms in supervised learning. In the classification model, the Decision Tree J48 algorithm was utilized to test the performance of the prediction. After the results have been obtained, two more classification models namely the Naivebayes Standard Classification algorithm and K- Nearest Neighbour algorithm has been used to compare which has the better performance and accuracy among the models.

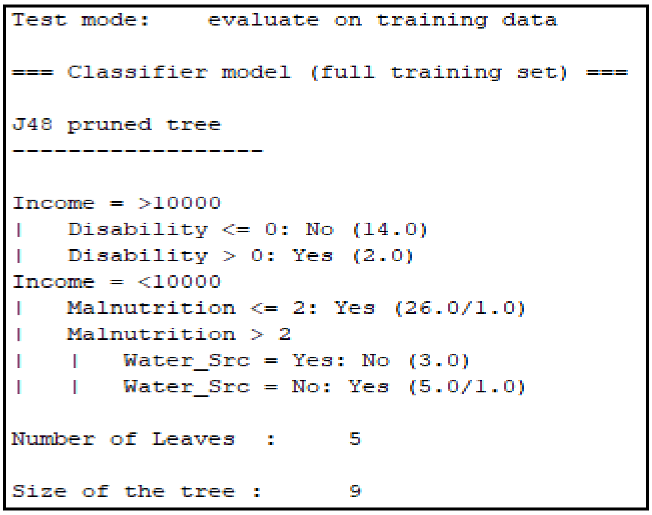


Relation: Poverty_Incidence_Eastern_Samar

No.	1: Household_Member	2: Food Cost Per Day	3: Water_Src	4: Sanitation	5: Malnutrition	6: Child_Mort	7: Maternal_Mort	8: Disability	9: School_DrpOut	10: Illiteracy	11: Income	12: Meal	13: Food_Threshold	14: Unemployment	15: Housing
	Numeric	Numeric	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal	Numeric	Numeric	Numeric	Nominal
1	5.0	180.0	Yes	Yes	2.0	1.0	1.0	0.0	2.0	1.0	>10000	2.0	65664.0	4.0	Yes
2	7.0	120.0	No	Yes	1.0	1.0	2.0	3.0	3.0	2.0	<10000	3.0	43776.0	5.0	Yes
3	6.0	130.0	Yes	No	2.0	0.0	1.0	0.0	0.0	1.0	<10000	2.0	47424.0	4.0	Yes
4	5.0	150.0	No	Yes	1.0	2.0	0.0	1.0	3.0	2.0	<10000	3.0	54720.0	3.0	Yes
5	8.0	200.0	Yes	No	3.0	1.0	0.0	0.0	2.0	2.0	<10000	2.0	72960.0	3.0	Yes
6	7.0	90.0	Yes	No	1.0	1.0	0.0	0.0	4.0	3.0	<10000	2.0	32832.0	4.0	Yes
7	6.0	100.0	No	No	3.0	0.0	1.0	0.0	3.0	2.0	<10000	2.0	36480.0	4.0	Yes
8	7.0	250.0	Yes	No	3.0	1.0	1.0	1.0	3.0	2.0	<10000	3.0	91200.0	3.0	Yes
9	8.0	85.0	Yes	No	2.0	1.0	1.0	0.0	4.0	3.0	<10000	3.0	31008.0	4.0	Yes
10	6.0	300.0	No	No	3.0	0.0	0.0	0.0	3.0	3.0	<10000	3.0	109440.0	4.0	Yes
11	5.0	120.0	Yes	No	3.0	1.0	0.0	1.0	3.0	2.0	>10000	2.0	43776.0	3.0	Yes
12	7.0	250.0	No	No	2.0	2.0	0.0	0.0	2.0	3.0	>10000	3.0	91200.0	2.0	Yes
13	7.0	130.0	No	Yes	1.0	0.0	1.0	1.0	3.0	2.0	<10000	3.0	47424.0	3.0	Yes
14	8.0	150.0	Yes	No	2.0	2.0	0.0	0.0	4.0	3.0	<10000	3.0	54720.0	4.0	Yes
15	8.0	100.0	Yes	No	2.0	1.0	0.0	0.0	3.0	2.0	<10000	3.0	36480.0	2.0	Yes
16	6.0	110.0	Yes	No	2.0	0.0	0.0	1.0	3.0	3.0	<10000	2.0	40128.0	3.0	Yes
17	5.0	100.0	No	No	1.0	0.0	0.0	0.0	2.0	2.0	<10000	3.0	36480.0	3.0	Yes
18	6.0	95.0	Yes	No	1.0	0.0	1.0	1.0	3.0	1.0	<10000	3.0	34656.0	2.0	Yes
19	7.0	270.0	Yes	No	2.0	1.0	0.0	0.0	3.0	3.0	>10000	2.0	98496.0	2.0	Yes
20	4.0	110.0	Yes	No	1.0	0.0	0.0	0.0	2.0	1.0	<10000	2.0	40128.0	2.0	No
21	5.0	130.0	Yes	No	2.0	2.0	0.0	0.0	2.0	2.0	<10000	3.0	47424.0	3.0	Yes
22	3.0	300.0	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0	>10000	3.0	109440.0	0.0	Yes
23	4.0	120.0	Yes	No	1.0	0.0	0.0	0.0	1.0	1.0	<10000	3.0	43776.0	1.0	Yes
24	6.0	150.0	No	No	1.0	1.0	0.0	1.0	2.0	1.0	<10000	2.0	54720.0	1.0	Yes
25	5.0	200.0	No	Yes	2.0	1.0	0.0	0.0	1.0	2.0	>10000	3.0	72960.0	2.0	Yes
26	7.0	250.0	Yes	No	2.0	0.0	0.0	0.0	1.0	0.0	<10000	3.0	91200.0	1.0	Yes
27	6.0	170.0	Yes	No	1.0	0.0	0.0	1.0	2.0	2.0	<10000	3.0	62016.0	2.0	Yes
28	8.0	270.0	Yes	No	2.0	0.0	0.0	0.0	3.0	3.0	<10000	3.0	98496.0	3.0	Yes
29	3.0	280.0	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0	>10000	3.0	102144.0	0.0	Yes
30	3.0	300.0	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0	>10000	3.0	109440.0	0.0	Yes
31	4.0	450.0	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0	>10000	3.0	164160.0	0.0	Yes
32	3.0	500.0	Yes	Yes	0.0	0.0	0.0	0.0	0.0	0.0	>10000	3.0	182400.0	0.0	Yes

Fig. 2. Snippet on the pre-processed dataset.

The figure above reflects some of the instances of the pre-processed dataset. It can be clearly seen that there are no missing or null values in all fields. The consistency of data in each instances can also be observed.



```
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Income = >10000
|   Disability <= 0: No (14.0)
|   Disability > 0: Yes (2.0)
Income = <10000
|   Malnutrition <= 2: Yes (26.0/1.0)
|   Malnutrition > 2
|   |   Water_Src = Yes: No (3.0)
|   |   Water_Src = No: Yes (5.0/1.0)

Number of Leaves  :     5

Size of the tree :      9
```

Fig. 3. Decision tree J48 results.

The results of the classification model created nine (9) trees with five (5) number of leaves.

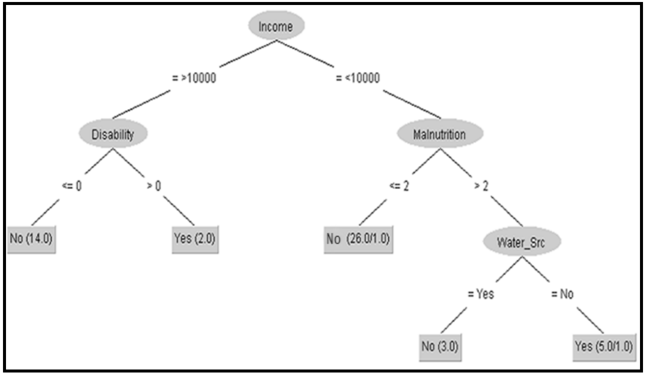
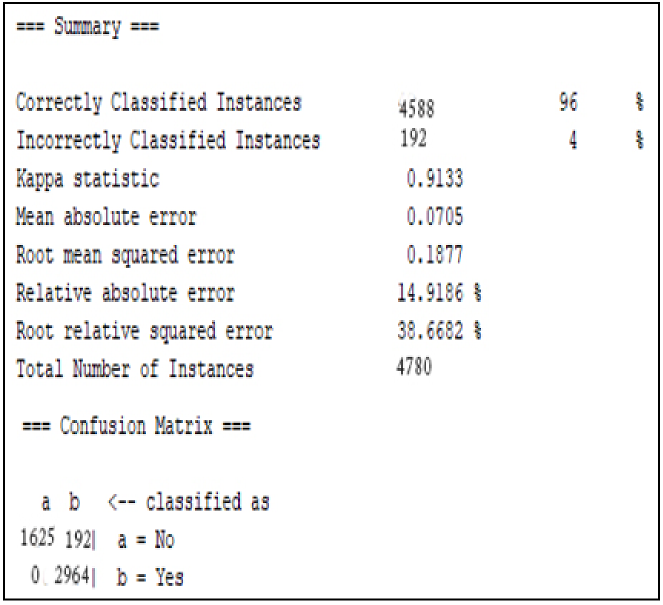


Fig. 4. Visualization of the Decision Tree

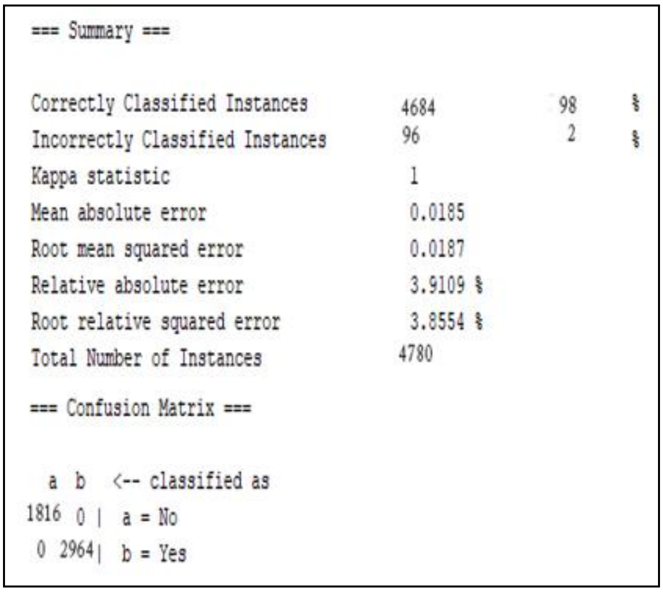
In the figure 4 above, it can clearly manifest that based on the prediction, the attribute Income which is under the Economic Sector has the most affecting factor of poverty in every household.



```
=== Summary ===

Correctly Classified Instances        4588          96    %
Incorrectly Classified Instances       192           4    %
Kappa statistic                          0.9133
Mean absolute error                      0.0705
Root mean squared error                  0.1877
Relative absolute error                 14.9186 %
Root relative squared error             38.6682 %
Total Number of Instances             4780

=== Confusion Matrix ===

  a   b   <-- classified as
1625 192|  a = No
 0  2964|  b = Yes
```

Fig. 4. Summary of Decision Tree J48 algorithm.

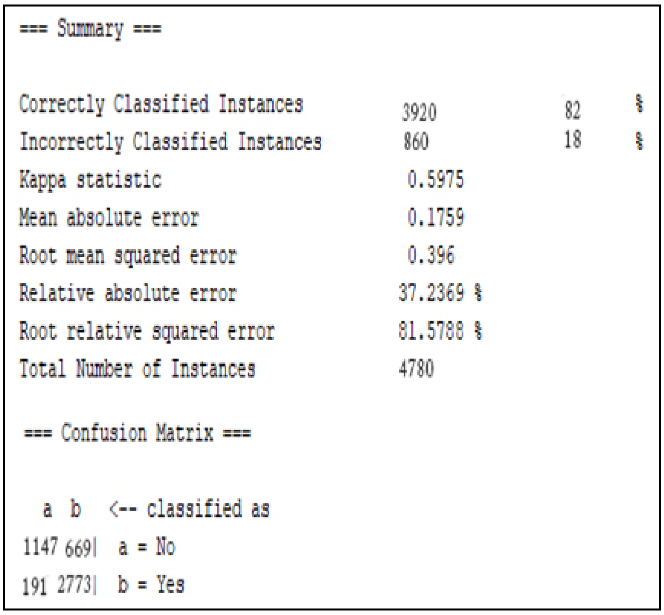
The figure above unveil the summery of the prediction results using the Decision Tree J48 Algorithm having a .9133 of Kappa Statistics with a correct classified instances of 96% and incorrectly classified instances of 4%.



```
=== Summary ===

Correctly Classified Instances        4684          98    %
Incorrectly Classified Instances        96           2    %
Kappa statistic                          1
Mean absolute error                      0.0185
Root mean squared error                  0.0187
Relative absolute error                  3.9109 %
Root relative squared error              3.8554 %
Total Number of Instances             4780

=== Confusion Matrix ===

  a   b   <-- classified as
1816  0 |  a = No
 0  2964|  b = Yes
```

Fig. 5. Summary of the KNN algorithm.

The figure 5 shows that the summary of the prediction results using the K- Nearest Neighbour algorithm garnered a 1 Kappa Statistics with a correct classified instances of 98% and incorrectly classified instances of 2%.



```
=== Summary ===

Correctly Classified Instances        3920          82    %
Incorrectly Classified Instances       860          18    %
Kappa statistic                          0.5975
Mean absolute error                      0.1759
Root mean squared error                  0.396
Relative absolute error                 37.2369 %
Root relative squared error             81.5788 %
Total Number of Instances             4780

=== Confusion Matrix ===

  a   b   <-- classified as
1147 669|  a = No
191 2773|  b = Yes
```

Fig. 6. Summary of the NaiveBayes algorithm.

The figure above depicts the summary of the prediction results using the NaiveBayes Standard Classification algorithm having a .5975 Kappa Statistics with a correct classified instances of 82% and 12% incorrectly classified instances.

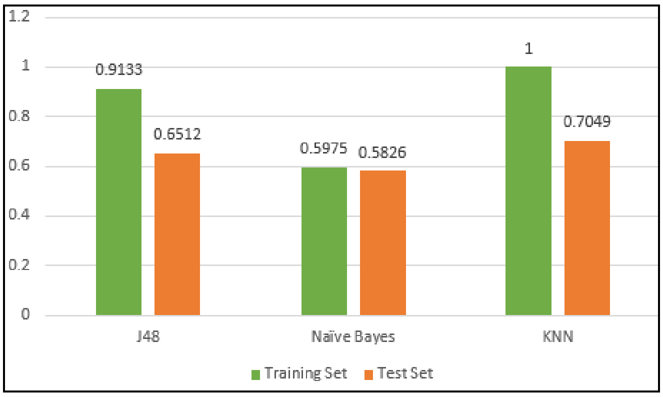


Fig. 7. Comparison of Kappa Statistics.

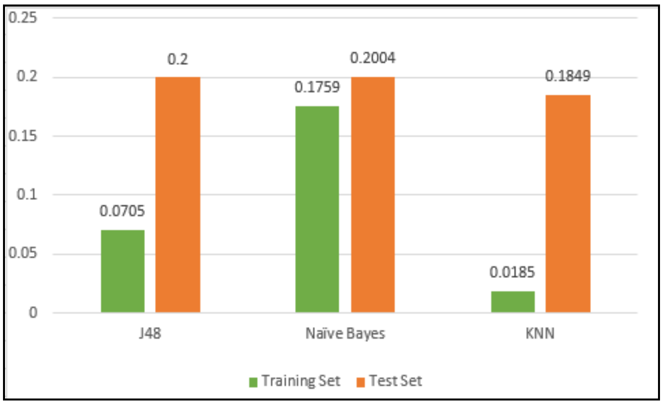


Fig. 8. Comparison of Mean Absolute Error.

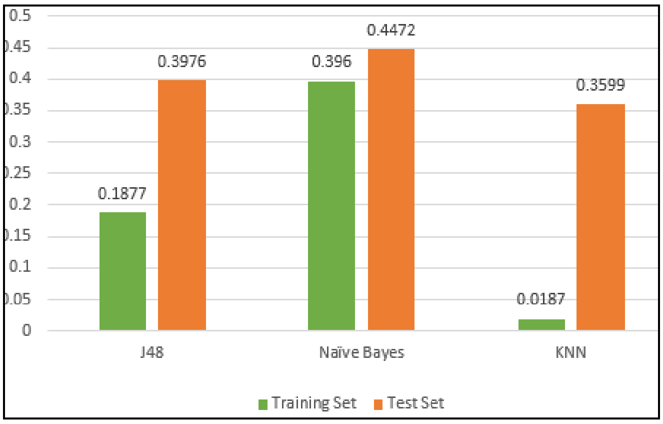


Fig. 9. Comparison of Root Mean Squared Error.

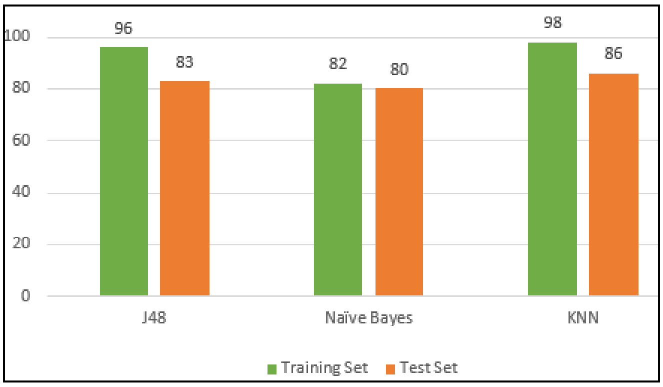


Fig. 10. Comparison of Accuracy.

The figures 7, 8, 9, and 10 are the results of the comparisons of the three algorithms both in training and test dataset.

Table- II: Accuracy and Performance of Different Classification Models in Data Mining.

Classification Model	Accuracy	True Positive Rate	False Positive Rate	True Negative Rate	False Negative Rate	Precision	Recall	F-Measure	KAPPA Stat
Decision Tree J48	96 %	0.895	0	1	0.105	0.962	0.960	0.960	0.9133
KNN	98 %	0.983	0.025	1	0.101	0.979	0.983	1	1
NaiveBayes	82 %	0632	0.065	0.935	0.368	0.825	0.820	0.813	0.5975

The table above depicts the comparison between the performances and accuracy of three different classification models used in prediction of the poverty alleviation in the province of Eastern Samar. The results clearly manifested that the KNN algorithm has the highest accuracy among the rest.

V. CONCLUSIONS

Data Mining Techniques can be used in a variety of ways in the different problems existing in our daily life. This study plays a significant role in predicting the poverty alleviation with regards to what indicators that may affect the most in reducing the poverty to a large extent. Based from the analysis of the three different classification model used to run test on the available dataset, the KNN algorithm is undeniably the best fit algorithm to predict the problem stated above. Likewise, among all the four (4) poverty core indicators with eighteen (18) attributes, the “Economic Sector” indicator with the “Income” attribute is the most significant factor that affects the poverty alleviation in the province of Eastern Samar.

The conclusion drawn in this study was reinforced through one of the study results [11] which indicated that poverty decreases as the overall output of economy increases which are positively affected by the higher **INCOME** factor of every household.

It is with a strong recommendation of the researchers that the results of this study may be used as one of the bases of the provincial and local government of the province to create programs or find ways that focus on what factors are predicted that affects the poverty incidence in the province of Eastern Samar. We hope that with this, poverty incidence if not eradicated, at least be reduced to a large extent, thereby uplifting the way of life of all the constituents of the province.

ACKNOWLEDGMENT

We are grateful to the authors who have shared their insightful and very inspiring studies through the publication of their papers digitally in different journals. We would also like to extend our warmest gratitude to our colleagues and mentors for pushing us to our limits to be able to come up with a scientific study. Above all, thanks be to God for the strength and wisdom bestowed upon us along this journey.

REFERENCES

1. J. Rutkowski, “Employment and poverty in the philipines”, The Philipine Social Protection Note, Report No. 9, December 2015.
2. A. M. Balisacan and N. Fuwa., “Going beyond crosscountry averages: growth, inequality and poverty reduction n the philippines”, World Development, Vol. 32, No. 11, pp 1891-1907, Great Bretain, 2004.

3. L. G. S. Bersales, “Poverty incidence among filipinos registered at 26.3% as of the first semester of 2015”, Philippines Stattistics Authority Press Release, PSA-2016-PHDSD-018, March 18, 2016.
4. L. G. S. Bersales, “Proportion of poor Filipino registered at 21.0 percent in the first semester of 2018”, Philippines Stattistics Authority Press Release, PSA-2019-053, April 10, 2019.
5. C. Marquez, C. R. Morales, and S. V. Soto, “Predicting school failure and dropout by using data mining techniques”, IEEE Journal of Latin-American Learning Technologies, Vo. 8, No. 1, pp. 7-14, Febraury, 2013.
6. A. Kaur, N. Umesh, and B. Singh, “Machine learning approach to predict student academic performance”, International Journal for Research in Applied Science & Engineering Technology, Vol. 6, Issue IV, pp. 734-742, April 2018.
7. S. Agarwal, G. N. Pandey and M. D. Tiwari, “Data mining in education: data classification and decision tree approach”, International Journal of e-Education, e-Business, e-Management and e-Learning, Vo. 2, No. 2, April 2012.
8. A. Kunjir, P. Pandershi and K. Naik, “Recommendation of data mining techniques in higher education”, International Journal of Computational Engineering and Research, Vol. 5, Issue 3, March 2015.
9. G. F. Tzortzis and A. C. Likas, “The global kernel k-means algorithm for clustering in feature space”, IEEE Transactions on Neaural Networks, Vol. 20, No. 7, pp. 1181-1194, July 2009.
10. C. M. H. Sai baba, A. Govindu, M. K. S. Raavi, V. P. Somisetty, “Student performance analysis using classification techniques”, International Journal of Pure and Applied Mathematics, Vol. 115, No. 6, pp. 1-7, 2017.
11. C. B. Cororaton, A. B. Inocencio, M. M. Tiongco, and A. B. S. Manalang, “Assessing the potential economic and poverty effects of national greening program”, De La Salle University Business & Economics Review, Vol. 26, Issue 1, pp. 136-157, 2016.

AUTHORS PROFILE

Mr. Jared Harem Q. Celis is a faculty member and the Director of Planning, Research, Extension and External Affairs of Eastern Samar State University – Guiuan Campus. He earned his degree in Master in Information Technology and also pursuing Doctor of Information Technology at AMA University, Quezon City. His research interest in on Web and Database Technologies, Data Mining, Systems Analysis and Designs.

Dr. Andres C. Pagatpatan, Jr. is a professor I and the Campus Administrator at Eastern Samar State University – Guiuan Campus. Hea earned his degree in Doctor of Philosophy in Educaional Management major in Educational Programs Management at Eastern Visayas State Unviersity, Tacloban City. His research interests is on social studies, industrial product development and information management.