International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 10, Issue 1 (2024)
https://doi.org/10.31436/ijpcc.v10i1.301

Identifying and Predicting Muslim's Community Funeral Funding Protocols

Izzul Ismail Bin Ibrahim, Muhammad Hassan Bin M Ashhuri, Raini Hassan*
Department of Computer Science, Kuliyyah of ICT, International Islamic University Malaysia, Kuala Lumpur, Malaysia

*Corresponding author: hrai@iium.edu.com.my
(Received: 10th January 2022; Accepted: 27th June 2023; Published on-line: 28th January 2024)

Abstract— This research aims to understand funeral poverty among the Muslim community in Malaysia by using Machine Learning algorithms. Generally, the funeral poverty in Malaysia can be described as funeral in Malaysia are exorbitantly priced, inflicting a disproportionate amount of financial hardship on the poor. As the death rate continues to climb year after year, we may conclude that funeral fees are an unavoidable risk. This research problem was inspired by the Sustainable Development Goal number 1 and 3 where the goal of this initiative is to increase awareness about the need of having funeral expenditures and how to prevent funeral poverty. Previous works have shown promising results. However, they are not conclusive enough as they lack predicting capability that Machine Learning can offer. Selected Machine Learning algorithms such as Decision Tree, Random Forest, and Naïve Bayes were used to classify the people that will go through funeral poverty based on a selected dataset and a survey conducted. The research methodology is as follows: 1. Collecting the data, 2. Pre-Process the data, 3. Exploratory data analysis, 4. Selecting the feature, 5. Modelling, 6. Evaluation. The results showed that the accuracy value for the selected algorithm is 0.95, 0.95 and 0.95 for the awareness and 0.8, 0.8, and 0.75 for B40 Status in survey datasets. With a larger taxonomy and more extensive, diverse sets of data, these figures are expected to improve. Linear Regression has been applied to help understand the features of death rate. Both outcomes serve distinct reasons. Because the individuals who would experience funeral poverty are in the lower income neighbourhood, it was necessary to classify impoverished communities in order to anticipate funeral poverty. Following that, the findings of what causes death to rise aid in understanding how the B40 community has been impacted by funeral poverty. In future, this study will aid future research on funeral poverty on a larger scale. When the two datasets are integrated rather than separated, better results can be obtained. The accuracy of the findings may also be improved by employing a wider range of Machine Learning methods and deep learning implementation. A more coherent and detailed model for forecasting funeral poverty can be developed.

Keywords— SDG, Poverty, Funeral, Naive Bayes, Decision Tree, Random Forest, Linear Regression, Death, Gender, Age.

I. Introduction

As poverty levels rise year after year, we must consider the causes of the rise in order to meet one of the Sustainable Development Goals (SDG). The goal of this initiative is to raise awareness about the necessity of having funeral costs and how to avoid funeral poverty. Funeral costs have been increasing at a pace of 3.69 percent per year, which is greater than the rate of inflation [1]. Funerals are exorbitantly expensive, putting a disproportionate financial strain on the poor. As the death rate rises year after year, we may conclude that funeral costs are an inescapable danger. To counter this, we need statistics to show that funeral expenditures have a significant impact on the spread of poverty. Until now, this issue has not been visualized perfectly to help the community understand how dire this issue is. A type of classification method to classify the poor community must be done to understand why the rate of funeral poverty is increasing rapidly. Funeral poverty may be eradicated by focusing on establishing a budget for costs rather than offering contributions and cash with no set price. This research will assist in identifying the primary issue of funeral expenditures, and using the data studied, we will be able to forecast future poverty and mortality rates, which we can then investigate further. The two main focuses of this research are classifying people who will be affected with funeral poverty by using machine learning algorithms and predicting the death factor in Malaysia. Our data is based on the survey that we had to make for the public and a dataset of deaths by state in Malaysia from the year 2001 until 2018. The analysis aids us in gaining an understanding and likelihood of future events based on previous data. The goal for the analysis is to provide the best evaluation of what will

https://doi.org/10.31436/ijpcc.v10i1.301

happen in the future, rather than just knowing what has happened.

Khairat kematian is an informal agreement or group formed with the express goal of assisting the participants' families throughout the burial process. Similar to takaful, which is founded on mutual agreement to assist one another, and contributions based on the 'tabaru' (gift) concept. However, as compared to the latter, the former has some limits in terms of functions and advantages. Funerals are exorbitantly priced, disproportionately hurting the poor. This cost is making it even more difficult to manage money during a time of grief and loss at a critical and sensitive moment. In the Malay Muslim community, when one of the family members passes away, the other community members will aid the family with burial arrangements [2]. This procedure is inexpensive, yet it may have a negative impact on the community's impoverished. For the current khairat fund practice of shifting mosque-managed khairat funds to takaful companies like Syarikat Takaful Malaysia, with premiums ranging from RM15 to RM24 per year [3]. The advantages are in accordance with the contract, which covers 24 hours a day, seven days a week, and anywhere in the globe. The word donation (sadaqa) means "truth, acceptance, and concurrence" in Arabic [4]. However, the word has a deeper connotation in the Quran and hadith, implying extreme compassion and generosity, as well as the occasional suggestion of charity towards others, or merely refraining from harmful activity.

The concept of poverty and the amount of poverty among Malaysians have been a source of contention. Malaysia falls under the classification of an upper middle-income economy. This country has traditionally looked at poverty through the prism of absolute poverty. The third of the five fundamental foundations that support the Islamic rule of behaviour is zakat. In alleviating the problem of poverty, Islam promotes a solution that is very precise and effective, that is making it compulsory for qualified Muslims to give alms or zakat. However, as stated in electronic media and newspapers, there are still numerous difficulties linked with poverty that zakat organizations must solve. This links to funeral poverty in the Malaysian Islamic community. Funeral expenses are extremely expensive.

Despite the fact that most Muslims are buried in public cemeteries, the expense of a Muslim burial has risen over time. To date, very limited academic research has been done on Islamic pre-need funeral plans. One of the solutions which were non machine learning related for the modern khairat fund practice, by transferring the khairat fund which has been managed by mosques to the takaful company such as Syarikat Takaful Malaysia [3]. Some methods were by acquiring qualitative analysis by reviewing all the literature proof to give a higher cause to the issue. Questionnaires and surveys have been done whereby the participants are the mosques, regulators from Islamic councils, those individuals and groups that are involved in funeral management, managers of the funeral scheme and takaful operators. The result, it seems that there are few challenges of managing khairat funds recorded in the analysis.

There are several systems and machine learning algorithms being used in previous works as a new and better way to eradicate poverty. The use of Naive Bayes, Decision Tree and k-Nearest Neighbors algorithm implemented is to classify the B40 population in Malaysia [5]. In order to show the accuracy of the algorithm's performance on classifying the B40 population using a cross-validation method after the feature selections. With results, the Decision Tree model is statistically significant and outperformed other classifiers with an accuracy of 99.27 percent. Another approach by is to use a variety of classification algorithms to understand the poverty in Jordan. a total of 16 classification algorithms were being used to get the best results [6]. As surveys are hard to attend to, machine learning algorithms help the best in this issue. Using the dataset of household features ranging from the year 2002 until 2017, the article shows proof of capability to classify which household is poor and can predict future poverty. As a result, the use of LightGBM proved to be the best and most accurate algorithm to classify. This work sheds a light on the multi-dimensional phenomenon of poverty in Jordan.

Existing websites for “online khairat” pose as a good indicator of awareness to cultivate online capabilities in handling funeral expenses in a faster and centralized pace. However, the variety of websites offering the same service will also create confusion to the community as to which website should they venture into. With that, the main problem that we are facing is the decentralized process of funeral expense service held by many accountable mosques around Malaysia. As some provide good features and service for the people to experience better, some also provide a difficult time for people to deal with the website. For example, the names for all websites available are eerily similar, making consumers confused which one is the legitimate one. This type of confusion will make consumers feel more uneasy to utilize web services in funeral expense management thus they will not be entertaining the use of such online service. Additionally, there might exist spam websites making the available legitimate websites more unnecessary and untrustworthy. On the plus side, the amount of people reaching to use website services on funeral expenses grows larger by day making sure that the future of funeral expenses can be made easy with such platforms. Given below are the virtual aspects of the existing websites available in Malaysia.

https://doi.org/10.31436/ijpcc.v10i1.301

II. Methodology

In the following section, we will talk about the many strategies that we had implemented during our study in order to address the issues that were already there.

A. *Data Collection*

The process of data collection was initiated by the collection of data from our survey form that asked several questions regarding our objective on knowing the awareness level and also the data that can help us predict funeral poverty in the Muslim Community. The questions that we asked in the survey form are age, gender, job status, income per month, strata, and other important questions that can help us to reach our objective. Our next dataset is about Death by state sex and age group in Malaysia from 2001 to 2018. This dataset contains 11041 data about the rate of death from different age groups, gender, and state. On both datasets, we have divided the dataset into training and testing by 70:30 for our selected algorithm implementation.

1) *Tools:* For our experimental setup and analysis, we mostly employed Google collab as python IDE and Tableau for visualization purposes.

- Google Colab: Google Colab is a free cloud management tool that allows us to conduct the coding. One of the most appealing aspects of the collaboration is the ease with which we may adjust the runtime. We use GPU runtime for our study because one of our datasets is big and large.
- Python: Python is a programming language that we employed extensively in our study. It's an object-oriented programming language with dynamic semantics at a high level. When compared to other programming languages, it offers a syntax that helps developers to build programs in fewer lines. It has a wide range of applications in the fields of graphical user interface design, web development, and software and system management
- Tableau: Tableau is a visual analytics platform that is revolutionizing the way we use data to solve issues by enabling individuals and companies to get the most out of their data. Tableau analytics platform, which is the market-leading solution for contemporary business intelligence, makes it simpler for people to explore and manage data, as well as find and share insights that may transform organizations and the world.

B. *Data Preprocessing*

We only have 2 datasets to be analysed. We began the data pre-processing with our survey datasets. The datasets have 10 columns and 103 rows which means that the survey has 10 questions and 103 respondents. First of all, we had dropped the first column which is about the timestamp that is auto generated by the google form. Since it is not related to our objectives, we just drop the column. Next, we also have changed the column name to a simple term for easing us in doing the coding and reading. Since the column name initially is named the same as our survey questions that have very long names, we just changed it to a shorter name. Furthermore, we added one more new column "ID" for making the data have uniqueness. The data that we have several null values on the optional answered questions. So, we used the NumPy library to replace the null value with the value of "0". After that, we also changed certain categorical column data types into integer data types for making the value of the column can be calculated using our selected algorithm (see Figure 1). For example, the 'Job Status column is a categorical data type with yes and no values. So, we create a new column called 'Jobstatus_cat' representing the 'Job status' column with an integer value of 1 and 2. After that, we also created another new column about declaring B40 status based on the Salary per month column. If the salary per month is below Rm 3000, it will be declared as B40 people.

Figure 1 shows us that the datasets are not really correlated with each other but still, the data can be used for predictive analysis. The datasets only have below 0.6 correlation value as the highest most correlated in the datasets which are not very good or very bad.

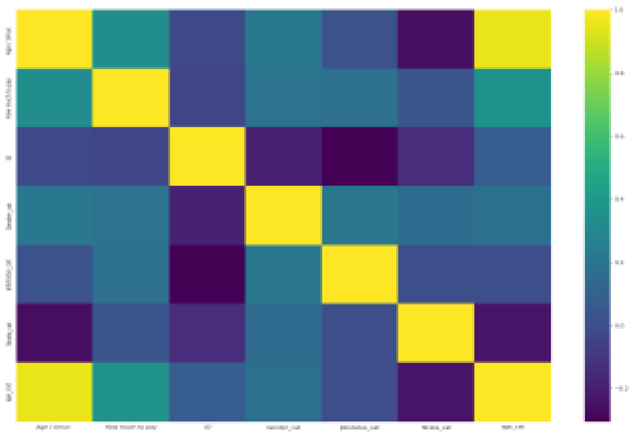

Fig. 1 Correlation Matrix for survey dataset

As for the next dataset which is the death dataset, it contains 11041 rows in a span of 4 columns. To help understand it more, we reduce the data by only including a 5-year dataset in which we have already pre-processed it into more relevant features for the data analysis and implementation. As the attributes before the data cleaning does not correlate with each other, we replaced it with new attributes that can correlate with each other.

https://doi.org/10.31436/ijpcc.v10i1.301

The new dataset consists of 'Year','D_male','F_male','L_30'.'M_30','MTD','MTS', and 'TD'. All data types excluding the column Year are in integer form for a better relationship between other features.

Figure 2 shows us that the datasets are divided in which some are correlated and some otherwise. This indeed can be used to measure and predict feature accuracy in the analysis.

Fig. 2 Correlation Heatmap for death dataset

.

C. *Modelling*

Following the exploratory data analysis stage, the next step is to move on to the modeling stage, which entails choosing three distinct algorithms and assessing their performance for each of the selected stations. There are 2 types of feature selection that we implemented with our selected algorithm.

Three distinct algorithms were chosen and taught in the same way under identical conditions. Random Forest, Naive Bayes, and Decision Tree were the algorithms used. A train-test split approach was utilized to assess the models, which works as a measure of how well the Machine Learning algorithm is doing when predictions are made on data that was not used to train the model. It is a time-saving strategy to use, with the results allowing one to assess the effectiveness of the chosen Machine Learning algorithms for their predictive modeling challenge. It's an excellent method for supervised learning techniques that include both classification and regression issues. The approach necessitates the division of a dataset into two subgroups. Non fit the model, the first subset, referred to as the training dataset, will be used. The second subset, referred to as the testing subset, will not be used to train the model; instead, it will be utilized as input to the model, with predictions produced and compared to predicted values based on the input [7]. For this research, the usual split rate was 70 percent for the Train dataset and 30 percent for the Test dataset.

One distinct algorithm was chosen to find out the relationship among these features that will affect the target variable and it is Linear Regression. An Ordinary least-squares (OLS) model is implemented as it suits the most in terms of relating variables. OLS helps in minimizing the sum of squared errors by fitting a model with a link between one or more explanatory factors and a continuous or at least interval outcome variable. The way we measure it is by getting the estimate of the coefficient.

III. Results and Discussions

The result and discussion is presented in such a way that each classifier is discussed in separate paragraph and then compared toward the end. Random Forest is a supervised learning technique that is used to solve regression and classification issues. During the training phase, it creates a huge number of decision trees and outputs the class that is the mean prediction (regression) of the individual trees or the mode of the classes (classification). Because of its simplicity and versatility in terms of working on both regression and classification issues, it has also become one of the most common algorithms [10]. According to the literature research, random forest is one of the top-performing algorithms for dealing with poverty-related data, thus it was chosen.

The Naive Bayes method is a supervised learning technique for addressing classification issues that are based on the Bayes theorem. The Naive Bayes Classifier is a simple and effective classification method that aids in the development of rapid machine learning models capable of making quick predictions. It's a probabilistic classifier, which means it makes predictions based on an object's likelihood [11]. We chose this algorithm because we want to know whether it is as good as other common algorithms on finding the prediction for poverty issues (based on previous works) or if it is just not suitable for these issues. This algorithm is mostly utilized in text classification tasks that need a large training dataset.

A Decision Tree is a supervised learning approach that may be used to solve both classification and regression problems, however, it is most commonly employed to solve classification issues [10]. Internal nodes contain dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier. The Decision Node and the Leaf Node are the two nodes of a Decision tree. Leaf nodes are the result of those decisions and do not include any more branches, whereas Choice nodes are used to make any decision and have several branches. It's a graphical depiction for obtaining all feasible answers to a problem/ decision depending on certain parameters. It is termed a decision tree because, like a tree, it starts with the root node and grows into a tree-like

https://doi.org/10.31436/ijpcc.v10i1.301

structure with additional branches. A decision tree simply asks a question and divides the tree into subtrees based on the answer (Yes/No) [9]. We chose this algorithm because our problems are basically decision-making problems. So, we think this algorithm is suitable for our research issues.

TABLE I
EVALUATION RESULTS (ACCURACY)

Target variable	Random Forest	Naive Bayes	Decision Tree
Awareness	0.95	0.95	0.95
B40 Status	0.8	0.8	0.75

What this confusion matrix tells us is that there is 0 value for the predicted 'Yes' value which means there is not correctly classified as such and also 0 value for incorrectly classified as 'No' answer. There is one 'No' answer being classified as False Positive. The value of True Negatives is one which means it represents data points that have a false ('No') real label that has been correctly classified as False by the model. This cell corresponds to the 'No' answer for the awareness that has been correctly classified as such. This model proves to us that based on gender, age, and strata, the prediction of the "Yes" answer for awareness is zero while the answer of "No" is one which means it shows us that from the selected predictors, people are not aware of the funeral expenses. The confusion matrix from 3 different algorithms gives us the same result with the same accuracy value. Therefore, we cannot see which algorithm is much better because all the selected algorithms give us the same result.

For this section, we select gender, job status, and strata as predictors and status as targets to know if the b40 group can be predicted through gender, job status, and strata. Maybe we can know if gender, job status, and strata can influence the accuracy of the prediction. If the results are likely positive after applying the algorithms, we can ensure that the status of the b40 group can be determined by the selected predictors. The 'Yes' on true positive value is equal to one which means it is correctly classified as such. 0.5 are incorrectly classified as 'No' (the false negatives block on the bottom left). This model is good because it gets all of the predictions correct as the value of true positives equals one. There is none of 'No' being classified as such (False Positive on the top-right corner).

The result is the same as the confusion matrix for random forest and naive Bayes because the value of True positives and False positives are literally the same. Since true positives and true negatives are data points that our model successfully identified, whereas false positives and false negatives represent data points that our model incorrectly categorized, it shows us that the model has successfully predicted the section of the true positive (correctly predict 'Yes' for B40 Status) while successfully predicted the true negatives section (correctly predict 'No' for B40 Status) with the value of 0.38. Moreover, there are 0.62 incorrectly classified as 'No' for B40 status.

Based on the accuracy value for all three selected algorithms, the random forest and naive Bayes give us the same value which is 80% of accuracy while the decision tree is 75% accuracy. We also can see that job status, strata, and age are related to B40 status in Malaysia with an accuracy value from 75% to 80%. Furthermore, in the confusion matrix, we can see the predicted for 'Yes' answer have a very accurate value while the predicted for 'No' value only had 0.38 to 0.5 percentage value. It proves that the predictors like age, job status, and strata are influencing the status of the B40 group in Malaysia.

A. *Target Variable: Median of Death by Year*

1) *Ordinary Least-squares (OLS):* Ordinary least squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model. The setup after an OLS model initiation is as follows:

The 'MTD' is the dependent variable where we want to measure. the standard error seems to have only 0.001% which shows why the R-squared equals 1. The feature 'D_male' shows significance to the dependent variable as the model captures 86.9% of the variance in the dependent variable. This shows that as 'D_male' increases by 1, 'MTD' will increase by 0.869.

2) *Linear Regression*: Linear regression helps in predicting relationships between features that are significant to each other. In this model, we can see that the model is able to predict almost accurately on the first five predictions of 'MTD'.
3) *Coefficient Estimate*: This estimate helps understand the significance of the feature towards the target variable. The table below shows the results of coefficient with its standard error values in different scenarios:

TABLE II
SCENARIOS AND COEFFICIENT

Scenarios	Coefficient	Standard Error
Male Death	0.869	0.001
Female Death	1.1774	0.002
Constant, Male Death	224.9100,0.8714	2151.147, 0.023
Male Death, Age 30 lower	0.8701, -0.0082	0.021,0.159
Male Death, Age 30 higher	0.6770,0.1196	0.191,0.119
Female Death, Age 30-lower	1.1691,0.0464	0.038,0.212
Female Death, Age 30 higher	0.8137,0.1672	0.339,0.156

The result explains the significance of the features and target variable. That way, we get to see what really affects the rising number of deaths which we can then relate to the B40 classification result. All scenario's R-squared value is surprisingly 1 which means it captures the variance of the dependent variable accurately. This shows that the dataset is undergoing overfitting which is not a good sign. Nevertheless, the results can still explain the relevance of the feature's relationship. The constant scenario is added to make sure there is still at least one scenario that does not have 1 for the value of R-squared. By looking at the table above, both male and female deaths show significance with the dependent variable 'MTD' as their coefficient estimates are 0.869 and 1.1774. This means that if the variables increase by 1, the predicted value of 'MTD' increases by 0.869 and 1.774. Other scenarios show good significance except for the 'Male death, Age Lower 30' scenario where one of the coefficients is a negative. This means as Male death increases by 1, 'MTD' increases by 0.8701 and as Age Lower 30 increases by 1, 'MTD' decreases by 0.0082. In a simpler interpretation we can say:

- The minority death of male in the population is likely aged below 30.
- In the subject of death of male population, is lower the age of them are likely to be below 30 years old.

B. *Relationship of both findings*

Both results have different purposes. To predict funeral poverty, classification of poor communities had to be done as the people that will undergo funeral poverty are in the lower income community. Next, the findings of what causes the death to increase helps in understanding how the B40 community got affected with funeral poverty. By understanding both results, we can see that people that are classified in the B40 category may be affected with funeral poverty due to their gender and age factor. This however can only be seen if we only take the two features in the death dataset result and therefore there is still room for more attributes and features that can help understand the causes of death that leads to funeral poverty.

IV. Conclusions

According to the results of the poll and the information gathered, we were able to conclude after modelling and exploring that the awareness level of the Muslim community in Malaysia is all significantly reliant on the data from the gender, age, and strata. Moreover, the status of B40 people is also heavily dependent on data from gender, job status, and strata. Model selection, feature selection, and model assessment were all carried out, and both methods produced acceptable results. Based on the analysis of our survey dataset, it is very well observed that the Random Forest and Naive Bayes algorithm are very good at predicting the outcomes for both feature selection with accuracy levels from 0.80 to 0.95. We can conclude that despite the Random Forest algorithm being the one common algorithm to be used in this kind of research, we also can use Naive Bayes as the predictive algorithm because it has similar outcomes with the Random Forest. We also find out that some features affect death to increase more. Coefficient estimation in the OLS model helps us differentiate and capture features that have the most influential variance in the target variable. However, there are limitations to what Linear Regressions can do. The prediction model can only predict the likelihood and the estimation of the target variable. Deep learning could solve the issue of predicting future death rates and death features far better than the Linear Regression model. We cannot also ignore the fact that the dataset itself is outdated as it showcases death cases latest by 2018. Nevertheless, Good insights and findings were achieved by this said research that could help us understand which and when people will be affected by funeral poverty.

V. Future Works and Recommendation

With all the findings and insights obtained throughout this research, it is safe to say that it did not stray away. This research will help in future studies on funeral poverty on a more bigger scale. Better results can be achieved if the two datasets combined together rather than separated. Results of the findings can also be more accurate by using a more variety of machine learning algorithms and implementation of deep learning. A model predicting funeral poverty can be made in a more cohesive and specific manner. Despite all that, this research may have flawed on the lack of utilising the capabilities of fusions of machine learning. Furthermore, the features and attributes can be improved by making more of it. Some attributes were also omitted as it was difficult to the relevancy of the research but can be implemented in future works.

Acknowledgement

We would like to express our gratitude to the International Islamic University Malaysia for its assistance. We would also like to thank Dr. Rawad Abdulghafor for his assistance with the research.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

[1] A.A. Mirza, I., Khairi, K. F., Shukor, S. A., Muhamed, N. A., Rasedee, A. F. N., Suhaimi, M., ... & Kamarubahrin, A. F. Funeral Poverty: An Exploratory Study for Micro-Takaful. Tazkia Islamic Finance and Business Review, (2021). 14(2).

https://doi.org/10.31436/ijpcc.v10i1.301

[2] M.A. Afif, Transforming the ‘khairat kematian’ (Mutual benevolent association) to mini cooperative. Vol. 43/2014, Dimensi Koop, Maktab Koperasi Malaysia. (2014), pp. 25-28

[3] M.S. Firrdhaus. Khairat kematian dalam komuniti Islam (Death Fund Concept in Muslim Society). (2015). Retrieved from https://www.academia.edu/8261480/Khairat_Kematian_Dalam_Komuniti_Islam_Death_Fund_Concept_in_Muslim_Society_.

[4] F. Omer, A Comprehensive Guide to Zakat Charity in Islam. The Light Publishing, (2007).

[5] S Sani, Nor & Abdul Rahman, Mariah & Bakar, Azuraliza & Sahran, Shahnorbanun & Sarim, Hafiz. Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification. International Journal on Advanced Science, Engineering and Information Technology. (2018), 8. 1698. 10.18517/ijaseit.8.4-2.6829.

[6] A. Alsharkawi, Al-Fetyani, M.; Dawas, M.; Saadeh, H.; Alyaman, M. Poverty Classification Using Machine Learning: The Case of Jordan. *Sustainability* 2021, *13*, 1412. https://doi.org/10.3390/su13031412

[7] A. Dildar Masood, Adnan Mohsin Abdulazeez, and Diyar Qader Zeebaree. "Machine learning supervised algorithms of gene selection: A review." Machine Learning 62, no. 03 (2020).

[8] F.Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi. "Supervised machine learning algorithms: classification and comparison." International Journal of Computer Trends and Technology (IJCTT) 48, no. 3 (2017): 128-138.

[9] V.S. Jackins, Vimal, Madasamy Kaliappan, and Mi Young Lee. "AI-based smart prediction of clinical disease using random forest classifiers and Naive Bayes." The Journal of Supercomputing 77, no. 5 (2021): 5198-5219.

[10] C. Bahzad, and Adnan Abdulazeez. "Classification based on decision tree algorithms for machine learning." Journal of Applied Science and Technology Trends 2, no. 01 (2021): 20-28.

[11] JavaTPoint. “Decision Tree Classification Algorithm”. JavaTPoint. (n.d). https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.