International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

INTERNATIONAL JOURNAL OF ENTREPRENEURSHIP AND MANAGEMENT PRACTICES (IJEMP)
www.ijemp.com

NEW ZAKAT DISTRIBUTION MODEL USING SUPERVISED MACHINE LEARNING MODEL: A CASE STUDY IN UITM CAWANGAN PERLIS.

Azlan Abdul Aziz [1], Nor Azriani Mohamad Nor[2*], Wan Nurshazelin Wan Shahidan[3], Siti Nor Nadrah Muhamad[4], Nur Syuhada Muhammat Pazil[5]

[1] Mathematical Sciences Study, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Perlis, 02600 Arau,Perlis, Malaysia
Email: azlan172@uitm.edu.my
[2] Mathematical Sciences Study, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Perlis, 02600 Arau,Perlis, Malaysia
Email: norazriani@uitm.edu.my
[3] Mathematical Sciences Study, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Perlis, 02600 Arau,Perlis, Malaysia
Email: shazelin804@uitm.edu.my
[4] Mathematical Sciences Study, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Perlis, 02600 Arau,Perlis, Malaysia
Email: nadrahmuhamad@uitm.edu.my
[5] Mathematical Sciences Study, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Melaka, 77300 Merlimau, Melaka, Malaysia
Email: syuhada467@uitm.edu.my
[1,2,3,4,5] Statistical Analytics, Forecasting & Innovation (SAFI) Research Interest Group, Universiti Teknologi MARA Cawangan Perlis, 02600 Arau,Perlis, Malaysia
[*] Corresponding Author

Article Info:

Article history:
Received date: 15.03.2023
Revised date: 10.04.2023
Accepted date: 31.05.2023
Published date: 06.06.2023

To cite this document:

Aziz, A, A., Nor, N. A. M., Shahidan, W. N. W., Muhamad, S. N. N., & Pazil, N. S. M. (2023). New Zakat Distribution Model Using Supervised

Abstract:

The new coronavirus (COVID-19) spread in early 2020 and affected health, economy, industry and education worldwide. To mitigate the effects of COVID-19, the Malaysian government imposed the Movement Control Order (MCO). Borders were closed, commercial activity ceased, and all academic institutions were closed. Students need to undergo online home learning, which causes difficulties, especially students from B40 groups. Concerned with the problems faced by UiTM Cawangan Perlis students, especially those from low-income families, the Unit Zakat, Sedekah & Wakaf (ZAWAF) has offered financial assistance in the form of zakat. However, the zakat distribution process is time-consuming. Therefore, this study aims to build a new zakat distribution model based on a Supervised Machine Learning model. The results show that parents' income and household size

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

Machine Learning Model: A Case Study In Uitm Cawangan Perlis.. *International Journal of Entrepreneurship and Management Practices, 6* (21), 87-95.

DOI: 10.35631/IJEMP.621008

significantly contributed to the success of students receiving zakat assistance. The results of this study suggest that developing a creative approach could help to streamline routine processes. Hence, the department of zakat can resolve the administrative and financial inefficiencies that have rendered zakat institutions ineffective.

Keywords:

Zakat, Supervised Machine Learning Model, Logistic Regression

Introduction

Zakat is one of the pillars of Islam and every Muslim is compulsory to pay zakat under the applicable terms and conditions. Zakat has the potential to close the poverty gap and improve the financial standing of the local Muslims. In Malaysia, the implementation and management of zakat has been managed by the Zakat Collection Center under the State Islamic Religious Council (Yusoff, 2012), which plays a crucial role in advancing the socio-economic objectives of the practice.

Zakat is not just about collecting money; it is also about distribution, which is still a major issue today. Several studies show that zakat institutions fail to relieve their responsibilities regarding the distribution of zakat to the asnaf. This is because there are some difficulties with the distribution process to qualified and non-qualified asnaf, uneven distribution (Zainal, Bakar & Ram Al Jaffri, 2016), wrong target recipient, some lack zakat funds while some have a surplus (Rahmat & Nurzaman, 2019; Lubis, Lubis & Almaarif, 2019).

The COVID-19 pandemic adversely affects health, the economy, industry, education, and society (Cohen, Chakravarthy, Bharathi & Narayanan, 2022). Online learning from home poses challenges for teachers, parents, and students. Low-income parents face difficulties meeting the need to provide their children with online learning facilities such as laptops and adequate internet data. Students from institutions of higher learning, such as UiTM Perlis Branch conducted open distance learning (ODL) are also facing the same situation. Due to this, any financial assistance is beneficial in alleviating the financial burden incurred while facing the COVID-19 pandemic.

Zakat, Waqaf, and Sadaqah Unit (ZAWAF) UiTM Perlis Branch provide zakat assistance to UiTM citizens and students. However, the conventional method of applying zakat assistance is less efficient since applications take a lengthy process, including filling the physical form and face-to-face interview before students receive zakat assistance. This situation affects the distribution of zakat where students need it urgently to continue their studies. Therefore, a new zakat distribution model using the binary logistic regression technique was proposed to improve and speed up the process of approving zakat assistance applications. And hence identify the significant factors (variables) towards zakat application status.

Literature Review

Zakat eligibility refers to the criteria and conditions determining who can receive zakat funds. It ensures that zakat funds are distributed to those most in need and deserving of assistance.

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

There are many factors contributing to zakat eligibility such as age, gender (Abdullah & Sapiei, 2018), level of education, site location, total monthly income, Cumulative Grade Point Average (CGPA) (Adibah, Hasan, Ahmad, Aina & Razak, 2017), marital status, size of family, total amount of aid received, etc. (Ramli, Ahmad, Wahid & Harun, 2011)

Logistic regression is one of the most popular tools for applied statistics and discrete data analysis. According to research by Zahi and Achchab (2020), logistical regression is a controlled machine-learning algorithm that examines one or more variables in a data set to produce a result. It is categorized as supervised model learning because it is a classification algorithm that uses real labels with a labelled data set during its training phases.

A study by Rahmatya and Wicaksono (2018) discussed the preferred system approach method, which is a structured approach based on the requirements to be implemented by the system or application. The structured analysis does not clarify how certain specifications can be satisfied or how applications can be deployed. The results of their study show that it is possible to complete the model of the information system for the receipt and distribution of zakat funds, and that this model can aid aamileen in the management of muzakki and mustahiq data as well as the distribution of zakat, allowing for the equitable and precise distribution of zakat funds.

Methodology

This study is divided into four stages. In the earlier stage, a total of 1934 data from four semesters from the ZAWAF UiTM Perlis online zakat application was collected from the COVID-19 pandemic from March 2020 until February 2022. The detailed variables used for this study are as follows:

Table 1: Variables Description

	Variable	Type
Y	Zakat eligibility	Categorical
X_1	Gender	Categorical
X_2	State	Categorical
X_3	Faculty	Categorical
X_4	Programme	Categorical
X_5	Semester	Categorical
X_6	Cumulative Grade Point Average (CGPA)	Quantitative
X_7	Students's status	Categorical
X_8	Head family's occupation	Categorical
X_9	Head family's status	Categorical
X_{10}	Head family's income	Quantitative
X_{11}	Mother's income	Quantitative
X_{12}	Other income	Quantitative
X_{13}	Household size	Quantitative
X_{14}	Sponsorship	Categorical
X_{15}	Zakat receiver	Categorical

Data splitting is the common approach in machine learning studies to evaluate forecast accuracy. Table 2 shows the splitting of data into training set (as a model fitting) and testing set (provide a final model).(Gupta, Saluja, Goyal, Vajpayee, & Tiwari, 2022)

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

Table 2: Training and Test Split

Dataset	Number of observations
Training set (semester 20202, 20204, 20212)	1515
Test set (semester 20214)	419
Total	1934

Next stage is developing the zakat distribution model using a supervised machine learning technique. Supervised machine learning using labelled data to train data and produce an outcome with testing data to predict categorical outcomes (Zahi & Achchab, 2020). The binary logistic regression technique is supervised machine learning that makes an estimation or prediction of the probability of occurrence of an event (Belavagi & Muniyal, 2016). Supervised learning is to approximate the mapping function when the new input variables (x) and can predict an output variable (Y). An algorithm to learn the mapping function from the input to the output. When the algorithm performs to an acceptable level of performance, learning ceases.

The Binary Logistic Regression Model equation (Hosmer Jr et al., 2013) used for this study is as follows:

$$log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \quad (1)$$

where p_i is the probability that $P(Y_i = 1)$. The left-hand side expression is referred to as the logit or log-odds. Then, it can solve the logit equation for p_i to obtain

$$p_i = \frac{\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})}{1+\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})} \quad (2)$$

By dividing both the numerator and denominator by the numerator itself, it can simplify further as:

$$p_i = \frac{1}{1+\exp[-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})]} \quad (3)$$

In the mathematical expression, this formula is called the logistic function and can be written as:

$$f(z) = \frac{1}{1+e^z} \quad (4)$$

where

$$z = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) \quad (5)$$

Four statistical criteria were used to determine the best fit model. First, the Wald Statistic will use the Chi-square test to determine whether the coefficient for that independent variable is significantly different from zero (Shappot & Gorgees, 2018). The Hosmer-Lemeshow (HL) test was used to determine whether observed event rates matched expected event rates in population subgroups and was considered significant when Chi-square p-values were greater than 5%. (Zahi & Achchab, 2020). Third, McFadden's R-Squared, McFadden's (Adj.) R-Squared, Cox & Snell R-Squared, and Nagelkerke R-Squared were calculated to determine the best model fit, with values close to 1 indicating a better model fit (Adejumo & Adetunji, 2013). Finally, the classification table will be examined. The model was evaluated using the classification table that tested sensitivity, specificity, and accuracy. The true positive rate of the model is measured by sensitivity (Rahmatya & Wicaksono, 2018). On the other hand, specificity measures the proportion of correctly identified negatives. The proportion of

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

correctly classified observations in comparison to all observations. A model with a high sensitivity percentage but a low specificity percentage is suitable for prediction. The AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria) will be used to compare which model performs better. The lowest values of AIC and BIC indicate the best model. (Fabozzi, Focardi, Rachev & Arshanapalli, 2014)

Results

According to ZAWAF UiTM Perlis, recorded data updated on February 2022 show that a total of 1476 students of UiTM Perlis Branch have received zakat financial assistance with a total amount of RM513,600. From the result obtained, most students are from the Faculty of Applied Sciences (23.9 percent), the Faculty of Computer and Mathematical Sciences (20.14 percent) and the Faculty of Plantation and Agrotechnology (16.1 percent) with the Cumulative Grade Point Average (CGPA) is between 3.03 to 3.55. Most family heads are self-employed, retirees, laborers, housewives, and rubber tappers. During COVID – 19 outbreaks, almost 97 family heads lost their jobs. The income of the family head is between RM478.10 to RM1500.00, and the average mother's income is RM320.60. The number of dependents of the head of the family is 3 - 5 people.

This study aims to build a new zakat distribution model to predict the eligibility status of zakat funding students. All the variables were included in the model when performing binary logistic regression. The data was analyzed using RStudio software. Using the stepwise method, it suggests that to add only seven out of 15 variables. The performance of various models is compared in this study, the best models are selected, and the predicted values are then combined to produce the final result. Table 3 below summarizes the best fit model analysis with the corresponding statistical measure. Based on the significant value of Wald test statistic, head of family income, mothers' income, household size and zakat recipient are significant at 5% level, while semester, gender, and CGPA are not significant (Model 1). We ran for the testing set and only got 7.23 percent of predictive accuracy (Table 4). That is a very low percentage of accuracy in prediction. Then we made a comparison of the model for every insignificant variable removed as shown in Table 3. Finally, we got 83.65 percent predictive accuracy of testing set for model 4.

Table 3: Summary Of Best Fit Model Analysis

Best Fit Model	Model			
	1	2	3	4
Parameter	Head family's income* Mother's income* Household size* Semester Gender CGPA Zakat recipient*	Head family's income* Mother's income* Household size* Gender CGPA Zakat recipient	Head family's income* Mother's income* Household size* Gender	Head family's income* Mother's income* Household size*

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

Hosmer-Lemeshow	Chi Square	3.5267	7.4131	6.0341	8.8684
	Sig.	0.8971	0.4928	0.6434	0.3535
Pseudo R - squared	Cox & Snell	0.068	0.042	0.04	0.037
	Negelkerke	0.293	0.181	0.172	0.16
	McFadden	0.267	0.163	0.155	0.143
	McFadden's Adj	0.185	0.118	0.123	0.118
Error measure	BIC	324.188	314.614	303.003	299.52
	AIC	258.006	278.978	277.549	279.156
Sensitivity		0.1429	0.0857	0.0857	0.0571
Specificity		0.9991	0.9983	0.9983	0.9983
Accuracy		0.9742	0.9717	0.9717	0.9709

*significance at 5% level.

Table 4 : Predictive Accuracy For Testing Set

Model 1	Model 4
0.0723	0.8365

The Hosmer and Lemeshow test assess the goodness of fit in the binary logistic regression model. The Hosmer-Lemeshow statistic indicates a good fit if the significance value of the chi-square is more than 0.05. Based on Table 3, the Chi-Square significance values are greater than 0.05. Thus, all the models adequately fit the data.

R square shows approximately how the model explains many variations in the outcome. For Cox & Snell, Nagelkerke, McFadden and McFadden Adjusted R square, the model fit the data better when the Pseudo R -Squared statistic was higher. The R-square values of tests shown in Table 3 indicate the models are relatively good.

Table 3 also summarizes the result of the sensitivity, specificity, and accuracy rate. The best fit model for the training set was model 1 because it had the highest value of sensitivity, specificity, and accuracy rate compared with another model. The value for sensitivity in Model 1 was 14.29% and specificity was 99.91% with an accuracy rate of 97.42%. However, if we observed the result of sensitivity, specificity, and accuracy for all models, there only have a slight difference among all models. Hence, to determine the best model, we check the value of the error measure. The results show that of the four models, the lowest BIC values were obtained by model 4.

In our regression model, Y represents the variable Eligibility Status, ' 1' represents eligible, and '0' is not. The logit, or natural log of the odds of having made one or the other decision, is predicted. Based on model 4, shows that head family income, mother's income and household size were significant after the insignificant variables had been removed (Table 5).

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

Table 5: Result of Final Binary Logistic Regression Model with Overall Statistical Explanation.

Explanatory Variables	Estimated parameters (b)	Std. error	Sig.
Constant	-3.7888	0.4327	0.00
Head family's income	0.0007	0.0002	0.0001*
Mother's income	0.0006	0.0001	0.0000*
Household size	-0.2465	0.1091	0.0239*

* significant at 5% level

Therefore, the new zakat distribution model using Binary Logistic Regression is as follows:

$$log\left[\frac{p_i}{1-p_i}\right] = -3.7888 + 0.0007(Head\ family's income) + 0.0006(Mother's\ income) - 0.2465(Household\ size)$$

The confusion matrix is used to evaluate the model quality of fit. Its evaluate the logit model performance, correctly and incorrectly fitted values prediction (Zahi & Achchab, 2020). Table 6 displays the comparison between the predicted outcome and the actual data outcome using the proposed binary logistic regression model. The classification rates indicate that the proposed binary logistic regression model is significantly better in predicting the successful distribution of zakat assistance to students. The accuracy of the data was 97.09% which indicates that the model is good in predicting the success of students receiving zakat assistance.

Table 6: Binary Logistic Regression Classification Table

		Predicted	
		Eligibility status	
Observed		Success	Fail
Eligibility status	Success	1164	33
	Fail	2	2
Overall Percentage		97.09%	

Conclusions

During the outbreak of COVID-19, ZAWAF has greatly helped UiTM students by providing zakat assistance to help students facing financial problems. However, the conventional application method of using the physical application form and a face-to-face interview is taking a long time to process. It has slowed down the process of distributing zakat assistance to students. Although ZAWAF has made new changes by using the online zakat assistance application, the zakat assistance still takes a long time to distribute. This is because, during COVID-19, there has been a high increase in the application of zakat assistance by students. Due to that, this study was conducted to develop a new zakat distribution model using the binary logistic regression technique to expedite the distribution of zakat assistance to eligible students. The statistical analysis results from the Hosmer-Lemeshow test, coefficient of determination (R-square), error measure, sensitivity and specificity show that our model is well

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

fitted and has a good correct classification rate. In conclusion, a binary logistic regression model can be used to classify students eligible for zakat assistance using predictor factors.

Acknowledgments

This research would not be able to be completed without permission and financial support from Unit Zakat, Sedekah dan Wakaf (ZAWAF) UiTM Cawangan Perlis (600-TNCPI 5/3/DDN (09) (002/2021)) to our team. Also, many thanks to Statistical Analytics, Forecasting & Innovation (SAFI) Research Interest Group team members for their effort, time, energy and willingness in completing this research.

References

Abdullah, M., & Sapiei, N. S. (2018). Do religiosity , gender and educational background influence zakat compliance ? The case of Malaysia. *International Journal of Social Economics*, *45*(8), 1250–1264. https://doi.org/10.1108/IJSE-03-2017-0091

Adejumo, A. ., & Adetunji, A. . (2013). Application of Ordinal Logistic Regression in the Study of Students ' Performance. *Mathematical Theory and Modeling*, *3*(11), 10–20.

Adibah, N., Hasan, A., Ahmad, N., Aina, N., & Razak, A. (2017). Factors That Significantly Affect College Students ' CGPA. *International Academic Research Journal of Social Science*, *3*(1), 77–81.

Belavagi, M. C., & Muniyal, B. (2016). *Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection* (pp. 117–123).

Cohen, S., Chakravarthy, S., Bharathi, S., & Narayanan, B. (2022). *Potential Economic Impact Of COVID-19-Related ADB Economics*. *657*.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B.G. (2014). *Model Selection Criterion : AIC and BIC*. *41*(1979), 399–403.

Gupta, S., Saluja, K., Goyal, A., Vajpayee, A., & Tiwari, V. (2022). *Measurement : Sensors Comparing the performance of machine learning algorithms using estimated accuracy*. *24*(August), 0–5.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Lubis, M., Lubis, A.R., & Almaarif, A. (2019). Comparison of the Approach in the Zakat Management System. *Journal of Physics: Conference Series*, *1235*(1). https://doi.org/10.1088/1742-6596/1235/1/012048

Rahmat, R. S., & Nurzaman, M. S. (2019). Assesment of zakat distribution: A case study on zakat community development in Bringinsari village, Sukorejo district, Kendal. *International Journal of Islamic and Middle Eastern Finance and Management*, *12*(5). https://doi.org/10.1108/IMEFM-12-2018-0412

Rahmatya, M. D., & Wicaksono, M. F. (2018). *Model of receipt and distribution of zakat funds information system*. *407*. https://doi.org/10.1088/1757-899X/407/1/012071

Ramli, R., Ahmad, S., Wahid, H., & Harun, F. (2011). Understanding Asnaf Attitude: Malaysia's Experience in Quest for an Effective Zakat Distribution Programme. *International Zakat Conference*.

Shappot, S. J., & Gorgees, H. M. (2018). View of Use of Logistic Regression Approach to Determine the Effective Factors Causing Renal Failure Disease. *Ibn AL-Haitham Journal For Pure and Applied Sciences*, *31*(3), 143–150.

Yusoff, M. B. (2012). Zakat Distribution and Growth in the Federal Territory of Malaysia. *Journal of Economics and Behavioral Studies*, *4*(8), 449–456. https://doi.org/10.22610/jebs.v4i8.346

International Journal of Entrepreneurship and Management Practices
IJEMP
EISSN : 2600-8750
Volume 6 Issue 21 (June 2023) PP. 87-95
DOI: 10.35631/IJEMP.621008

Zahi, S., & Achchab, B. (2020). Modeling car loan prepayment using supervised machine learning. *Procedia Computer Science*, *170*.

Zainal, H., Bakar, A. A., & Ram Al Jaffri, S. A. A. D. (2016). Reputation , Satisfaction of Zakat Distribution , and Service Quality as Determinant of Stakeholder Trust in Zakat Institutions. *International Journal of Economics and Financial Issues*, *6*(S7), 72–76.