P-18

Prediction of Depression among School-going Adolescents Using Machine Learning Techniques

Hamizatul Akmal Abd Hamid[1], Aida Wati Zainan Abidin[2], Muhammad Fadhli Mohd Yusoff[1].

1 Institute for Public Health, National Institutes of Health, Ministry of Health Malaysia
2 Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara (UiTM) Shah Alam, Selangor, Malaysia

NMRR-19-853-47604

Introduction

Globally, depression has become the ninth leading cause of illness and disability among adolescents. Early diagnosis of depression helps the professionals to treat it at an earlier stage and prevent it from become more complicated problems during adulthood. Machine learning techniques (MLT) are currently well suited for analysing big medical data yet the application is still scarce in Malaysia.[1] This study aims to evaluate the performances of MLT in predicting the depression status among school-going adolescent aged 12 to 18 years old in Malaysia.

Methodology

Secondary data from Adolescent Health Survey (AHS 2017), a cross-sectional stratified sampling study was used. The analysis involved the balanced sample of 9,503 observations and 19 predictors. The outcome, depression status was determined by the DASS-21 Score. Four predictive models; Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT) were applied and the classification performance of each models were evaluated. The selection of the best model was made based on the performance of each model by looking at the three (3) criteria's which are Receiver Operating Characteristics (ROC) curve, misclassification rate and accuracy rate.

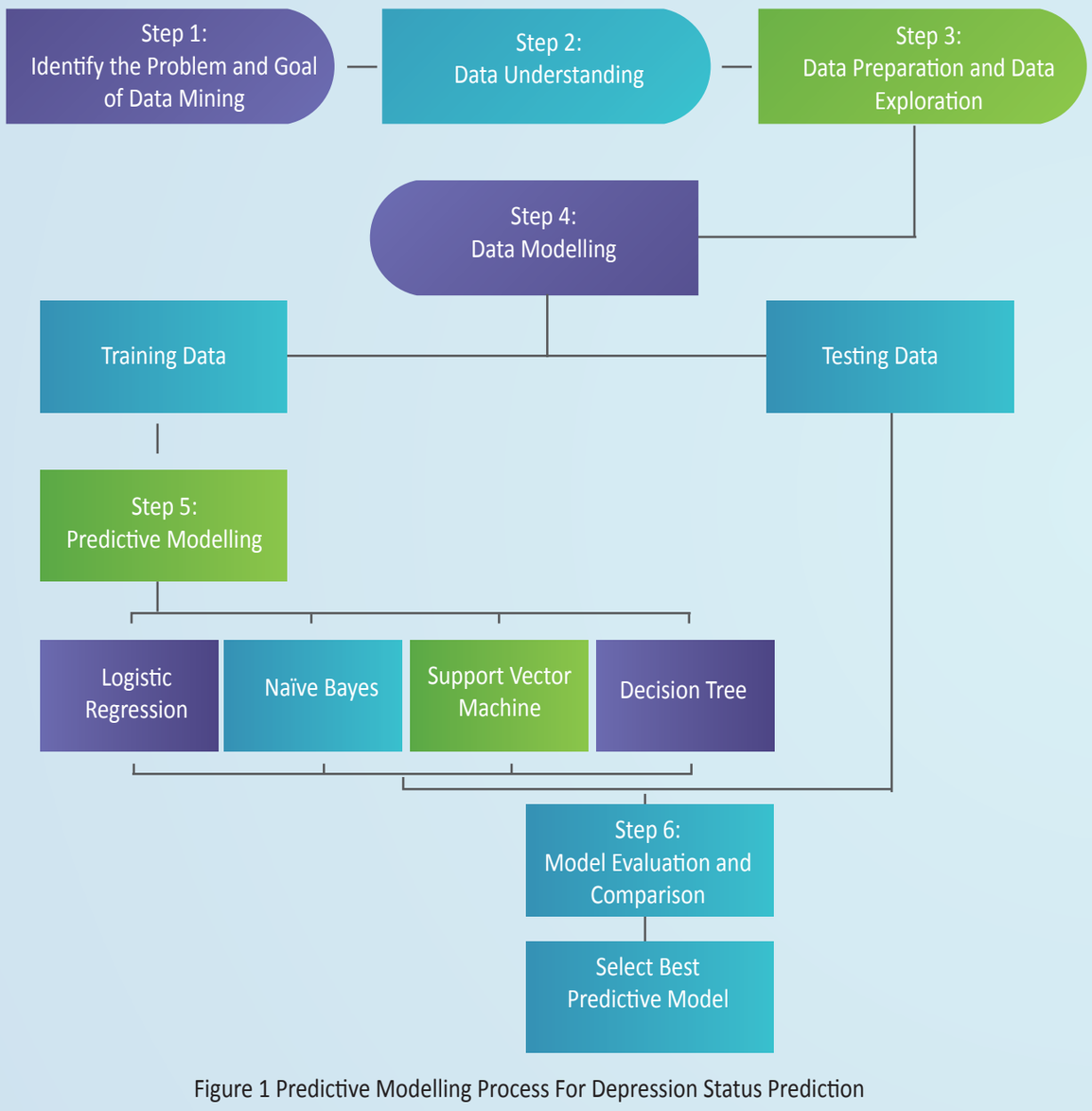


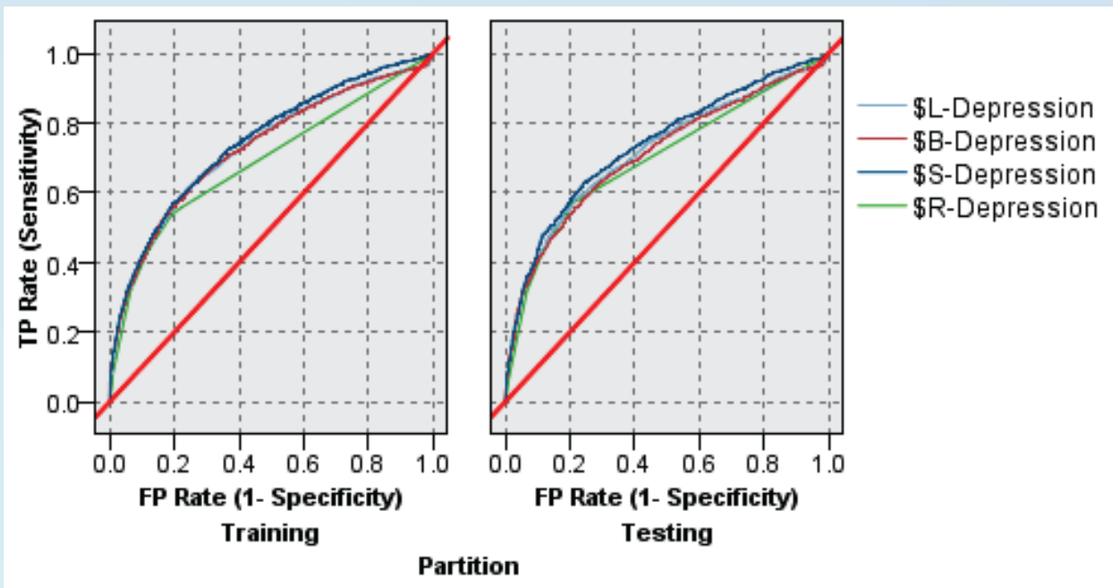
Figure 1 Predictive Modelling Process For Depression Status Prediction

Results

- A total of 4,778 (18.0%) adolescents had depression. The accuracy of models of MLT to predict the depression was best in LR Backward (68.9%), followed by SVM Linear (68.2%), NB (67.8%), and DT CART (67.7%) (Table 1).
- ROC chart (Figure 2) showed that the line for the LR Backward was more consistent in both training and testing sample compared with other models.
- Based on performance classification analysis results (Table 1) and ROC Chart (Figure 2), LR Backward model has been proved to have the highest rate of accuracy, the lowest misclassification rate and more consistent in training and testing sample compared to other models.
- The findings also highlighted the important variables that contribute the most to LR Backward model to predict depression status among adolescent were loneliness, sleep deprivation, bullied, suicidal ideation, parental connectedness, suicidal attempt, parental bonding, drug use, alcohol use and parent's marital status (Figure 3).

Table 1 Classification Performance Analysis

Models	Sample	Accuracy	Sensitivity	Specificity	Precision	Misclassification Rate
Logistic Backward	Training	69.09	59.99	78.27	73.59	30.91
	Testing	68.88	58.33	78.97	72.65	31.12
Naïve Bayes	Training	68.82	61.17	76.56	72.52	31.18
	Testing	67.84	59.16	76.15	70.35	32.16
SVM Linear	Training	67.23	50.73	84.24	76.83	32.77
	Testing	68.16	51.02	84.80	76.53	31.84
Decision Tree CART	Training	67.08	53.56	80.99	74.35	32.92
	Testing	67.72	53.76	81.28	73.63	32.28



Notes:
L-Depression - Logistic Regression (Backward)
B-Depression - Naïve Bayes
S-Depression - Support Vector Machine (Linear)
R-Depression - Decision Tree (CART)

Figure 2 ROC Chart of the Training and Testing Sample for All Models

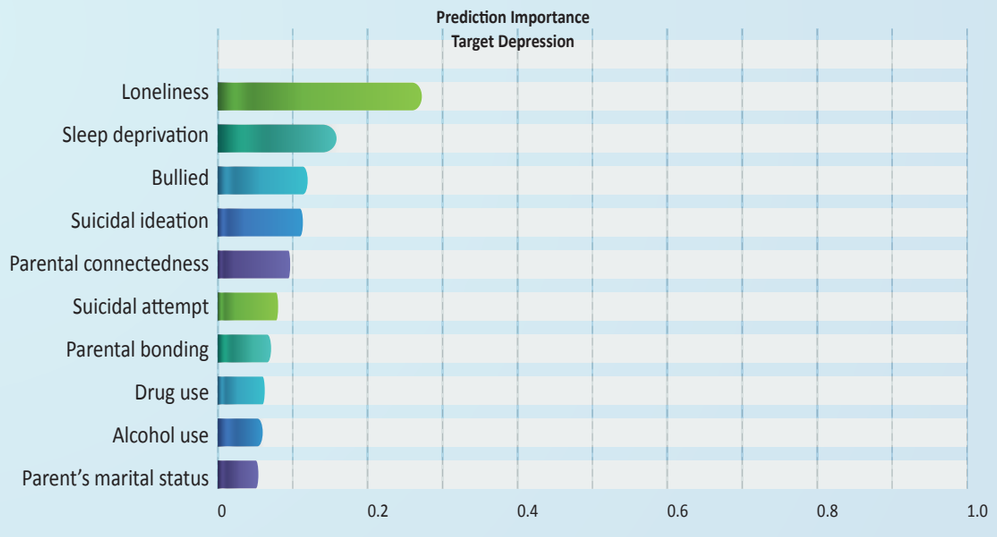


Figure 3 Predictor Importance for Logistic Regression (Backward) Model

Discussion

- This study highlighted the LR Backward as the best predictive model to predict the depression among school-going adolescent compared to other classification models. This result was supported by previous studies which indicated that LR have the best classification performance compared to other models[2-4].
- In term of classification performance analysis, LR Backward shows the moderate accuracy results with 68.88%. This results is similar with the study conducted by Camdeviren et al. (2007) where the results found that the accuracy rate, specificity and sensitivity for Logistic Regression model in predicting the postpartum depression women in Turkey was 65.4%, 95% and 16% respectively.
- However, according to Wolpert & Macready (1997) there is no single learning algorithm that universally performs best across all domains.
- This study used a standardized questionnaire based on Global School Based Health Survey (GSHS), thereby limiting the researcher to explore on other important factors that contributed to the depression among adolescent such as family history, history of abused at home, academic achievement, peer influence and parent's income.[7-11]

Conclusion

Although our top performing model have the moderate accuracy, however it lays important groundwork for future models to predict depression among adolescent. Future research should include more important factors that contributed to the depression among adolescent such as family history, history of abused at home, academic achievement, peer influence and parent's income. Other method such as Artificial Neural Network (ANN), Random Forest and Alternative Decision Tree (ADTree) can be used for improving classification performance of the study.

Keywords

depression, adolescent, machine learning techniques, Adolescent Health Survey

Acknowledgement

The authors would like to thank the Director-General of Health Malaysia for his permission to present this poster.

References

1) Islam, M., Hasan, M., Wang, X., Germack, H., & Noor-E-Alam, M. (2018). A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. Healthcare, 6(2), 54. https://doi.org/10.3390/healthcare6020054
2) Wallert, J., Tomasoni, M., Madison, G., & Held, C. (2017). Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. BMC Medical Informatics and Decision Making, 17(1), 1–11. https://doi.org/10.1186/s12911-017-0500-y
3) Shouval, R., Hadanny, A., Shlomo, N., Iakobishvili, Z., Unger, R., Zahger, D., … Beigel, R. (2017). Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An Acute Coronary Syndrome Israeli Survey data mining study. International Journal of Cardiology, 246, 7–13. https://doi.org/10.1016/j.ijcard.2017.05.067
4) Hou, Y., Xu, J., Huang, Y., & Ma, X. (2017). A big data application to predict depression in the university based on the reading habits. 2016 3rd International Conference on Systems and Informatics, ICSAI 2016, (Icsai), 1085–1089. https://doi.org/10.1109/ICSAI.2016.7811112
5) Camdeviren, H. A., Yazici, A. C., Akkus, Z., Bugdayci, R., & Sungur, M. A. (2007). Comparison of logistic regression model and classification tree: An application to postpartum depression data. Expert Systems with Applications, 32(4), 987–994. https://doi.org/10.1016/j.eswa.2006.02.022
6) Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation, 1(1), 67–82. https://doi.org/10.1109/4235.585893
7) Rodrigo, C., Welgama, S., Gurusinghe, J., Wijeratne, T., Jayananda, G., & Rajapakse, S. (2010). Symptoms of anxiety and depression in adolescent students; a perspective from Sri Lanka. Child and Adolescent Psychiatry and Mental Health, 4, 10. https://doi.org/10.1186/1753-2000-4-10
8) Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. Lancet (London, England), 379(9820), 1056–1067. https://doi.org/10.1016/S0140-6736(11)60871-4
9) Al-Sughayr, A. M., & Ferwana, M. S. (2012). Prevalence of mental disorders among high school students in National Guard Housing, Riyadh, Saudi Arabia. Journal of Family & Community Medicine, 19(1), 47–51. https://doi.org/10.4103/2230-8229.94015
10) Verma, N., Jain, M., & Roy, P. (2014). Assessment of Magnitude and Grades of Depression among Adolescents in Raipur City, India, 2(5), 10–13.
11) Umang P, S., Roy, N., Kumari, S., & Jugal, K. (2016). Prevalence and Factors Associated with Depression in School-Going Adolescents of India. Indian Journal of Youth and Adolescent Health (ISSN: 2349-2880), 3(4). Retrieved from https://medical.adrpublications.in/index.php/IndianJ-YouthandAdolescentHealth/article/view/920