

The Impact of Online Learning towards students during Covid-19 based-on Machine Learning Approach

Nahvin Muthusamy¹, Yoong Cheng Nee¹, Ding Xue Wing¹, Farabi Tasnim Ahmed¹, Sharifah Sakinah Syed Ahmad^{1,*}

¹) Faculty of Information and Communication Technology, University Teknikal Malaysia Melaka, 76100 Melaka Malaysia.

*Corresponding e-mail: sakinah@utem.edu.my

Keywords: Online learning; study stress; machine learning

ABSTRACT – This COVID-19 pandemic has changed so much of student's daily lives and gave a new norm called online learning. Although at first glance everything seems working well, upon deeper inspection how this online learning is taking a toll on student's mental health can be observed. It comes no surprise to constantly see students nagging about online learning on social medias which also corresponds to the increasing of student's mental health problem in Malaysia. Mental health is a very serious topic and this research is intended to understand the problem scope and to possibly give out some recommendations through the means of machine learning. The finding identifies a student's learning status and problems during online learning in terms of academic performance, physical and mental health and other issues such as financial aspects like paying extra for data and electricity during online learning.

1. INTRODUCTION

The year 2020 was definitely way different than how most of us imagined it to be. The main reason to that narrow down to 'COVID-19' and the pandemic it brought with it. It changed so much of our daily lives and introduced to new way of living. For many students, that meant online learning which left a huge mark on them. In simpler terms, online learning is essentially a virtual learning technique similar to video conferencing through various mediums such as Microsoft Teams, Google Meet & Cisco WebEx.

However, ongoing pandemic mixed with online learning was a recipe for disaster which applied immense amount of study related stress to students. Worryingly enough, according to the Health Ministry, a total of 465 suicide attempt cases received treatment in the ministry's hospitals between January to June 2020 as mentioned in previous study [1][2] which became the inspiration for this whole project.

The objectives of the project are:

- To identify student's learning status and problems during online learning in terms of academic performance, physical and mental health and other issues.
- To analyze the online learning behavioural patterns of different groups of students.
- To analyze the tendencies of contemporary learners about the use of online learning systems in learning process.
- To enhance the quality of learning and teaching so that the students and lecturers can perform

well.

- To provide appropriate initiatives to resolve the issues regarding online learning of both teachers and students on the basis of the key factors influencing the online learning platform.

2. METHODOLOGY

The methodology of the research consists of data collection, data pre-processing, exploratory data analysis and modelling using machine learning.

2.1. Data Collection

Since data about online learning was scarce, an online survey using Google Forms was conducted to accumulated data. This Google Form has been carefully constructed in a way to retrieve student's insight which consists of 23 questions related to opinions of students on online learning such as 'On average, how long do you spend on online learning every day except weekends and holidays?', 'What seems to be the drawbacks of online learning?' and 'How well could you manage time during online learning?'. The form then was distributed to university students through social medias. After a week, the form had 205 respondents (2021) which was above our initial target 150.

2.2. Data Pre-Processing

Since the data was collected from an online survey, the data was relatively clean with minor noises. To manage that, data cleaning was performed by removing irrelevant symbols together with control questions and data reshaping by dividing checkbox answers into own separate columns as previous study [3].

2.3. Exploratory Data Analysis

Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. One major advantage of Google Form is that it provides real-time graphical representation which can later be used to describe a particular attribute. Each attribute representation was studied carefully to understand what that particular attribute brings to context of online learning and any significant pattern has been noted down for further study.

2.4. Modelling

In machine learning, there are lots of algorithms that can be used to train the models on the data. For example, classification, regression and clustering. For

this project, classification has been chosen as the algorithms to train our models as supported by previous study [4][5][6][9]. It will utilize the input data and predict the categorical type of the data for the output. Since most of the attributes are binomial and polynomial in our data set, it will be easier for us to analyze the data and implement the model if classification method is used.

Through classification different models can be trained, such as k-Nearest Neighbours (kNN), Decision Tree, Random Tree, Random Forest and so on. For this project however, Random Tree model has been decided to train the data set on. Random Tree is easier to be implemented as it selects a random subset of attributes for each split of the tree. It will create a classification model that predicts the value of the label based on several input attributes of the data set which then will determine which attribute is the most important feature in our data set as mentioned in previous study [7] [8].

For the implementation of Random Tree, RapidMiner software along with few operators with randomly selected attributes were used as shown in Figure 1.

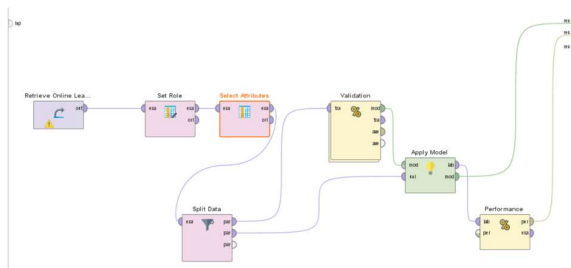


Figure 1 Process of RapidMiner

3. RESULTS AND DISCUSSION

Results of the project comes from both exploratory data analysis (EDA) and modelling. Both these processes give a different view of the data set which later can be used to come up with solutions to resolve named issues. From the EDA, some of the important visualizations which holds significant information were gathered as Figure 2, Figure 3, Figure 4 and Figure 5 shown below.

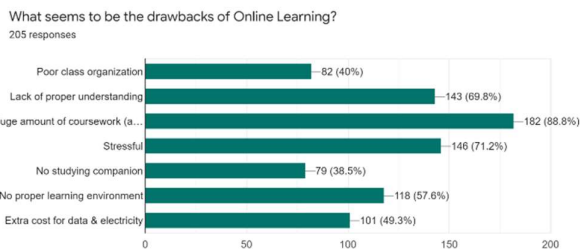


Figure 2 EDA for 'What seems to be the drawbacks of Online Learning?'

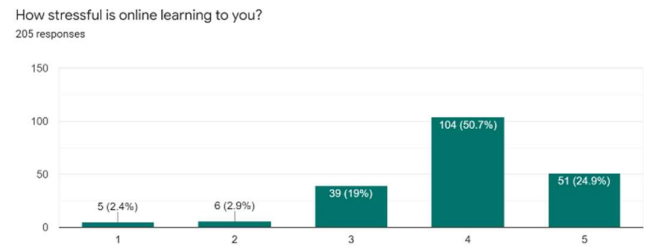


Figure 3 EDA for 'How stressful is online learning to you?'

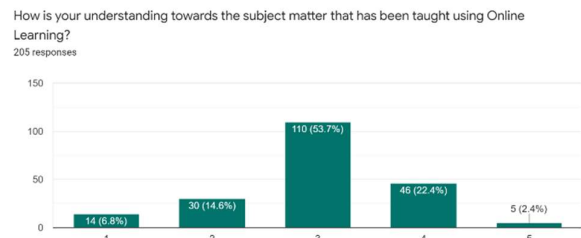


Figure 4 EDA for 'How is your understanding towards the subject matter that has been taught using online learning?'

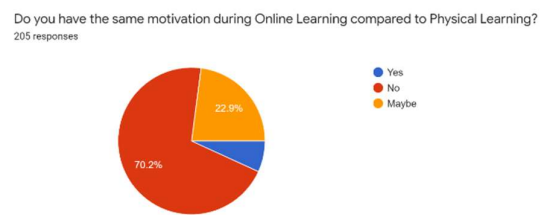


Figure 5 EDA for 'Do you have the same motivation during online learning compared to Physical Learning?'

As for the results of modelling, using attributes 'Are the concurrent deadlines affecting your eagerness to learn practically?', 'Do you agree that high quality learning cannot take place without face-to-face interaction?', 'Are you paying more for data and electricity compared to before Online Learning?', 'Do you have the same motivation during Online Learning compared to Physical Learning?', 'Do you have extra time to participate in other activities while doing Online Learning?', 'How difficult is it for you to stay focused on your coursework?' and 'How helpful are your teachers while studying online?' to train the Random Tree model gave the best performance.

Accuracy has been used as the performance metric to evaluate the performance of our models which is the total number of right predictions divided by the total number of predictions produced for one dataset, is one of the frequent evaluation metrics in classification problems. The accuracy of the Random Tree model with selected attributes is 74.41%.

The obtained Random Tree models have four decision trees with specified weightages as shown below in Figure 6, Figure 7, Figure 8 and Figure 9.

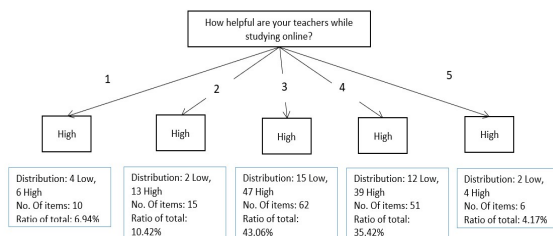


Figure 6 Model 1 [W=1.136].

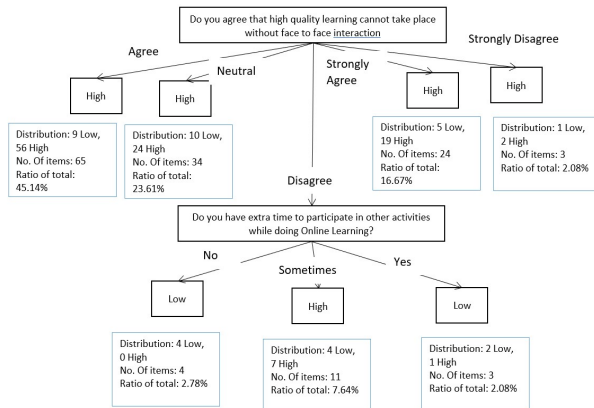


Figure 7 Model 2 [W=0.327].

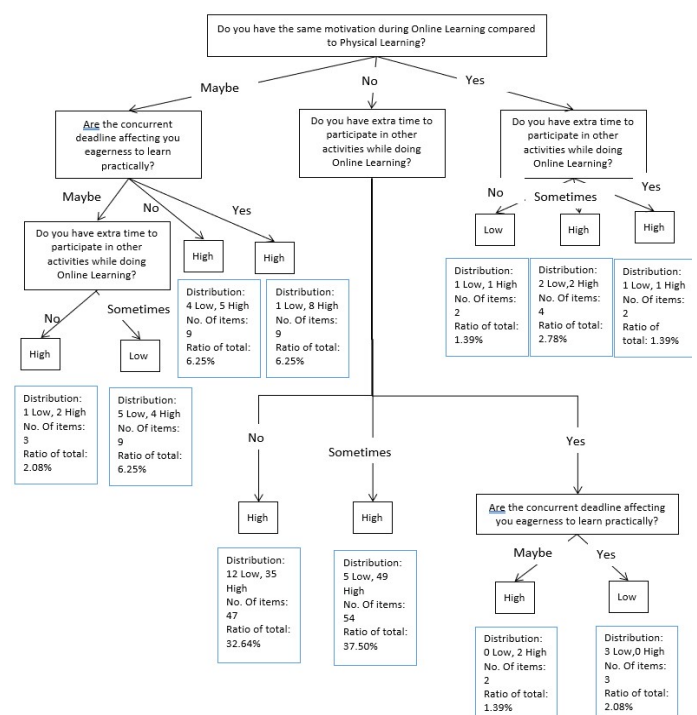


Figure 8 Model 3 [W=0.244].

Every single decision tree function in a similar way. However, associated weightages determine how valuable the prediction is. For example, in Figure 11 which is the third model of the Random Forest has a weightage of 0.244. The prediction process begins with the root of the tree following the specific branches and finally ending with leaves (labels). The predicted label together with the weightage will be compared with predicted labels from other models. The most frequent label will be selected as the final predicted value based on the weightages.

Low accuracy of the model can be due to number of reasons. The performance is based upon the data set that has been collected through online survey where there is scope of different opinions from different people about online learning. Also, the number of respondents were relatively low compared to conventional mainstream survey which did not register a lot of common traits among respondents. From the results, how online learning is impacting the students and what majority of the respondents are thinking of it can be observed. This gives a clear picture on the drawbacks of online learning. These selected attributes help to identify a student's learning status and problems during online learning in terms of academic performance, physical and mental health and other issues such as financial aspects like paying extra for data and electricity during online learning. It also shows the contribution of teachers in online learning.

It helps to visualize the effects of online learning to the quality of education due to lack of real-life interaction thus refraining students from achieving practical approaches to education. Also, the process emphasizes on their financial factors such as paying extra expenses for data and electricity. It also reflects the mental health issues of students during online learning.

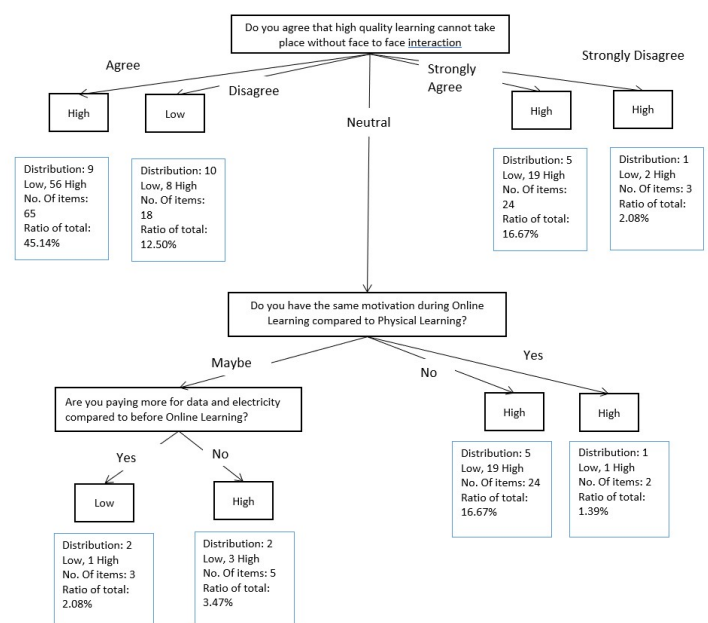


Figure 9 Model 4 [W=0.180].

Thus, helps to analyze the impacts of online learning from lecturer's and student's behavioral traits of understanding difficulties because of distance-based communication. The model also shows the feedback based on time management of students.

Figure 7 shows that 70.2% of the respondents do not have same motivation on online learning compared to physical learning which the attribute is also significant enough to be a tree root of Model 3 in Figure 11. This means that students could not retain attention throughout the entire class period which might be boring to them. This lack of focus issue can be easily solved by

implementing non-linear teaching techniques such as combination of synchronous and asynchronous classes. Sometimes practicing or learning alone without having to pay attention to ongoing online classes can be more interesting rather than staring at screen all the time.

Apart from that, since the attribute ‘*Do you agree that high quality learning cannot take place without face-to-face interaction?*’ is the tree root for Model 2 and Model 4 in Figure 10 and 12 respectively, this also falls under main drawback of online learning. Since students are very used to face-to-face learning, adapting to online learning which lacks the physical interactions seems alien to them. This issue can be combated by implementing hybrid teaching techniques which takes place in Sandbox Games. These Sandbox games will provide a virtual environment for teaching process to occur. This means that, students and lecturers can meet virtually through their game avatars. This would reduce the separation of teaching and learning process compared to typical online learning.

4. CONCLUSIONS

In summary, the implications of online learning and how it impacts the students has been discussed. The conducted survey definitely gave an insight of what vast majority of students are thinking about online learning. Processing has been done on those collected data and managed it properly to be used in modelling. This modelled random forest can be used alone to determine the stress level of students. However, using it as a means to distinguish the key factors of student’s study related stress by identifying the roots of tree models can give us an in-dept view of the situation. From these models, tailor made solutions can be developed to hopefully solve each drawback of online learning one by one. The objective of this project will be achieved if more students are aware of their mental health and drawbacks of online learning which are affecting them.

Learning is a crucial aspect in one’s life and online learning cannot be ignored just like that. Students have to manage their time well to make sure they are not falling into the stress pit. Until a better solution is found, online learning is all there is.

5. ACKNOWLEDGEMENT

We are thankful to *Universiti Teknikal Malaysia Melaka* for allowing us to leverage the university’s resources to conduct the initial stage of this research.

REFERENCES

- [1] Menon, S. (2021, April 4). Youth suicide. Retrieved from: <https://www.thestar.com.my/news/education/2021/04/04/youth-suicide>.
- [2] Abd Yusof N. F., Lin, C. & Guerin, F. (2018). Assessing the Effectiveness of Affective Lexicons for Depression Classification. In: Silberstein M., Atigui F., Kornysheva E., Métais E., Meziane F. (eds). *Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018, Paris, France, June 13-15, 2018, Proceedings*. Springer.
- [3] Kaur, J. (2018, December 21). Data Preparation Process Data Preprocessing and Data Wrangling. <https://www.xenonstack.com/blog/data-preparation>
- [4] Puyalnithi, T., Viswanatham, M. V. & Singh, A. (2016). Comparison of Performance of Various Data Classification Algorithms with Ensemble Methods Using RAPIDMINER. *International Journal of Advanced Research in Computer Science and Software Engineering* 6(5), 1-6.
- [5] Singh, A. & Sathiyaraj, R. (2016). A Comparison Between Classification Algorithms on Different Datasets Methodologies using RapidMiner. *International Journal of Advanced Research in Computer and Communication Engineering* 5(5), 560-563.
- [6] Syed Ahmad, S. S., Shaik Osman, A. N., & Basiron, H. (2018). The impact of social media on human interaction in an organisation based on real-time social media data. *International Journal of Data Science* 4(3), 260.
- [7] Donges, N. (2019). A complete guide to the random forest algorithm. Retrieved from: <https://builtin.com/data-science/random-forest-algorithm>.
- [8] Mbaabu, O. (2020, December 11). Introduction to Random Forest in Machine Learning. Retrieved from: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>.
- [9] Zulkarnain, N. Z., Basiron, H., & Abdullah, N. (2020). Writing Style and Word Usage in Detecting Depression in Social Media: A Review. *Journal of Theoretical and Applied Information Technology* 98(1), 124-135.