INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

www.hrmars.com
ISSN: 2222-6990

Machine Learning Approach to Classify Students' Mental Health During the COVID-19 Pandemic: A Web-Based Interactive Dashboard

Suraya Masrom, Nur Fatihah Jamaludin, Fadzilah Abdol Razak, Nor Rashidah Paujah @ Ismail

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v13-i7/17124 DOI:10.6007/IJARBSS/v13-i7/17124

Received: *25* April 2023***, Revised:*** *27* May 2023***, Accepted:*** *18* June 2023

Published Online: *30* June 2023

In-Text Citation: (Masrom et al., 2023)

To Cite this Article: Masrom, S., Jamaludin, N. F., Razak, F. A., & Ismail, N. R. P. @. (2023). Machine Learning Approach to Classify Students' Mental Health During the COVID-19 Pandemic: A Web-Based Interactive Dashboard. *International Journal of Academic Research in Business and Social Sciences*, *13*(7), 1192 – 1204.

Vol. 13, No. 7, 2023, Pg. 1192 – 1204

http://hrmars.com/index.php/pages/detail/IJARBSS | **JOURNAL HOMEPAGE**

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
www.hrmars.com
ISSN: 2222-6990

Machine Learning Approach to Classify Students' Mental Health During the COVID-19 Pandemic: A Web-Based Interactive Dashboard

Suraya Masrom[1], Nur Fatihah Jamaludin[2], Fadzilah Abdol Razak[3*], Nor Rashidah Paujah @ Ismail[4]
[1]Computing Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA Perak Branch, Tapah Campus, Perak, Malaysia, [2]Vision Technology Consulting Sdn Bhd, Menara Suezcap 2, KL Gateway, No 2, Jalan Kerinchi, Gerbang Kerinchi Lestari, Kuala Lumpur, Malaysia, [3,4]Mathematical Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA Perak Branch, Tapah Campus, Perak, Malaysia
Corresponding Author's Email: fadzilah.ar@uitm.edu.my

Abstract
The sudden shift to online learning caused by the COVID-19 pandemic has had a significant impact on students' mental health. Therefore, this paper presents a research project that utilized machine learning techniques to classify students' mental health during the COVID-19 pandemic and developed a web-based interactive dashboard for interactive data visualization. The project used Python as the programming language and adopted various machine learning models such as Support Vector Machine, Decision Tree, and Random Forest. The findings of the research reveal that the Support Vector Machine algorithm is the best algorithm in the mental health classification model, with an accuracy above 90%, as measured by a confusion matrix. The application of web-based interactive dashboard for data visualization assists educational institutions in identifying and supporting struggling students while developing effective mental health support strategies beyond the pandemic. While the project has limitations, including the inability to determine the root causes of mental health issues due to insufficient variables, it presents the immense potential of machine learning tools in addressing and mitigating mental health challenges during crises like the COVID-19 pandemic. Machine learning-based predictions can serve as valuable tools for mental health professionals in their decision-making processes, and interventions for students who may be at higher risk or in need of immediate support.
Keywords: COVID-19, Mental Health Classification, Machine Learning, Confusion Matrix, Interactive Visualization

Introduction

The ability of an individual to recognize their capabilities and manage everyday life's stresses is known as mental health (Bratman et al., 2019). It comprises intellectual, behavioral, and emotional factors that influence a person's thoughts, actions, and feelings.

Nevertheless, the global mental health has been considerably affected by the COVID-19 pandemic outbreak. Governments in most countries have imposed Movement Control Orders to reduce the spread of the virus, resulting in the closure of institutions and restrictions on social activities. The shift from face-to-face to online learning in schools and universities has been one of the responses to the pandemic. While this approach offers various perspectives, several students struggle to maintain motivation and concentration due to distractions in their surroundings, inadequate equipment, unstable internet connectivity, and uncertainty about the pandemic. These challenges have caused mild to severe psychological distress among many students, including anxiety, frustration, and boredom, which can have dramatic effects on physical and mental health (Chrikov et al., 2020; Wang et al., 2020).

When students engage in online learning, the absence of social interaction can impede their capacity to articulate ideas, evaluate their comprehension, and detect lapses in their reasoning, causing stress and potentially leading to untreated mental health issues (Mheidly et al., 2020). Consequently, it is vital to address these concerns and provide adequate medical intervention to those in need.

In response to the growing prevalence of mental health disorders and the necessity for appropriate medical intervention, machine learning has emerged as a potential solution in the field of mental health research. Machine learning, which is also referred to as artificial intelligence (AI), involves utilizing statistical methods to draw conclusions from existing data. It entails developing a function that can be generalized to novel data. Machine learning uses algorithms and analytical models in computer systems to perform a specific task using input data, without explicit programming. As a subset of AI, machine learning can recognize and comprehend meaningful patterns in data to create algorithms that can enhance learning performance over time, providing accurate, reliable, and efficient outcomes for various problems (Padala et al., 2019). Machine learning has evolved into a fundamental component of AI, utilized in problem-solving and identifying patterns, and is implemented in almost every domain of science and technology, including healthcare (Avula & Asha, 2018; Zoabi et al., 2021), where it has shown promising results in improving mental health research.

This paper discusses the implementation of a machine learning model inspired by a study from Shatte et al (2019), which can predict and classify students' mental health status during the COVID-19 pandemic. The study also proposes the development of a web-based dashboard that presents interactive visualizations of the mental health data collected from the students. The primary objective of this study is to categorize individuals into two groups: those who are experiencing mental health issues and those who are not. The findings of this research can aid in developing effective mental health support strategies for students during and beyond the pandemic.

The study's results have practical implications for mental health research, including the development of more precise and efficient tools for mental health assessment and support. By using machine learning to identify students experiencing mental health issues, appropriate medical intervention and counseling services can be delivered promptly. Educational institutions can leverage the predictive models to identify students at risk, provide timely support, and develop comprehensive mental health strategies during times of crisis. Additionally, the web-based dashboard developed in this study can help improve accessibility to mental health information as a supportive technology in promoting students' well-being and ensuring effective mental health care.

Literature Review

Machine learning is a powerful tool that has proven useful in solving real-world problems across various domains. It involves training computer algorithms to learn from data and make predictions without being explicitly programmed. With the advent of big data and advance progressions in computing power, machine learning has become increasingly prevalent in recent years for many areas.

In healthcare, machine learning has the potential to revolutionize the approach to patient care and treatment. Specifically, in mental health, machine learning has shown great promise in identifying patterns and predicting outcomes based on patient data. With the help of machine learning algorithms, mental health professionals can better understand the underlying factors contributing to various mental health conditions and develop more effective treatment plans.

There are several state-of-the-art research studies on mental health that utilize machine learning techniques. For example, researchers have used machine learning algorithms to predict depression Gao et al (2018), anxiety Priya et al (2020), and suicidal ideation from social media data (Ji et al., 2020). Other studies have focused on predicting mental health outcomes based on physiological data, such as heart rate (Panicker & Gayathri, 2019) and brain activity (Rashid & Calhoun, 2020). Machine learning has also been used to develop chatbots and virtual assistants that can provide mental health support and counseling services (Abd-Alrazaq et al., 2019). Overall, machine learning has enormous potential in the field of mental health research and treatment. With its ability to identify patterns and make accurate predictions, machine learning can provide valuable insights into various mental health conditions and inform the development of effective treatment strategies.

On the other hand, data visualization is the process of presenting data in a graphical or pictorial format to help understand complex information. It is an effective way to identify patterns and trends that may not be immediately apparent in raw data. Interactive data visualization takes this a step further by allowing users to interact with the data in real-time and explore it from multiple perspectives (Sarikaya et al., 2018). This allows for more efficient and effective decision-making, as well as better communication of complex information (Padilla et al., 2018).

In mental health, interactive data visualization can be a powerful tool for understanding patterns and trends in patient data. By presenting data in a visual format that is easy to understand, mental health professionals can quickly identify areas that require attention and develop more effective treatment plans (Vellido, 2020). Interactive data visualization also allows for real-time monitoring of patient progress, which can be especially useful in cases where frequent check-ins are required.

Data visualization plays a critical role in machine learning as it helps to identify patterns and trends within large datasets (Burkart & Huber, 2021). Machine learning models often require vast amounts of data to learn and make accurate predictions. Data visualization techniques allow researchers and data scientists to explore and understand complex datasets, which in turn can help them to improve the performance of their machine learning models.

Methodology

The study employs the Machine Learning Lifecycle with the Agile Model, as shown in Figure 1. The cycle consists of eight phases, which are discussed in detail below.

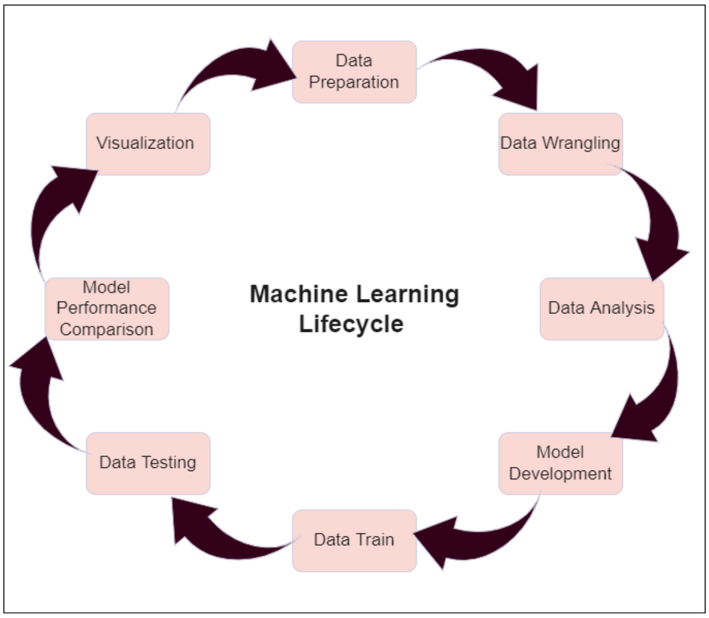

Figure 1. Machine Learning Lifecycle

The first two stages of the Machine Learning Lifecycle involve data preparation and wrangling, which includes data cleaning and formatting. For this project, a secondary dataset of COVID-19 Survey Student Responses uploaded into Kaggle.com was used, which was collected through a cross-sectional survey of 1182 students from different educational institutions in Delhi National Capital Region. The data cleaning process involved removing missing values, duplicate data, invalid data, and noise to ensure data quality. The cleaned data was then transformed into a usable format for analysis.

The machine learning algorithm for the mental health classification model was built using Python, and three supervised learning algorithms, namely, Support Vector Machine (SVM), Decision Tree, and Random Forest, were chosen based on their ability to optimize the specific purpose of minimizing prediction result error from the actual labels. The model was trained using 80% of the original dataset, and the remaining 20% was used for testing. A comparison was made between the prediction accuracy of the models by creating an accuracy table of their outputs, in order to identify the most dependable machine learning model for the given data and task. The model with the smallest mean square error value was ultimately chosen.

After obtaining precise outcomes, an interactive dashboard titled "The Impact of COVID-19 on Students' Mental Health Dashboard" was developed using Tableau software to visualize the analysis. The dashboard concentrated on the effects of COVID-19 on students' mental health. Utilizing interactive visualization not only made the insights more useful but also shortened the project development time and increased the dashboard's user engagement. In general, the top-level research activities involved in creating the interactive visualization dashboard for the students' mental health data are illustrated in Figure 2.

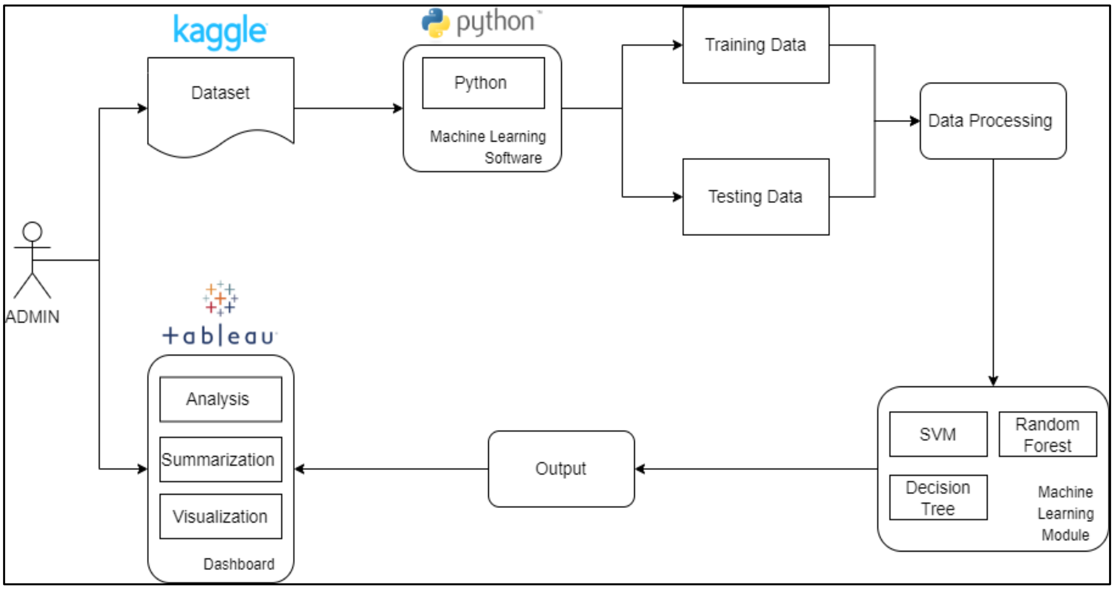

Figure 2. Supervised Machine Learning Project Architecture

To ensure the most precise representation of students' mental health, the findings were analysed and summarized on an interactive visualization dashboard created using Tableau software. This approach enabled users to interact with the data and obtain deeper insights, reduced the project development time, and made the insights more actionable. The system architecture employed in this project consisted of multiple phases, including data preparation, machine learning model development, data testing, and data visualization, all aimed at comprehending the effects of COVID-19 on students' mental health.

In evaluating the performance of machine learning classification models, a confusion matrix is a widely used table that assesses the model's performance on a set of test data with known actual values. The confusion matrix consists of four variables, namely True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). Table 1 provides a template of the confusion matrix table for case of mental health classification.

Table 1
Confusion Matrix of classifying the mental illness

		Actual Values	
		Having mental health illness	Not having mental health illness
Predicted Values	Having mental health illness	TP	FP
	Not having mental health illness	FN	TN

The count of each variable can be explained as the following

- True Positive (TP): The model predicts that a student has a mental illness, and it is true. In other words, the student has a mental illness and the model correctly identifies it as positive. In the confusion matrix, this is the count of correctly identified positive cases.
- False Positive (FP): The model predicts that a student has a mental illness, but it is false. In other words, the student does not have a mental illness but the model incorrectly

identifies it as positive. In the confusion matrix, this is the count of incorrectly identified positive cases. FP is also known as Type-I error.

- False Negative (FN): The model predicts that a student does not have a mental illness, but it is false. In other words, the student has a mental illness but the model incorrectly identifies it as negative. In the confusion matrix, this is the count of incorrectly identified negative cases. FN is also known as Type-II error.
- True Negative (TN): The model predicts that a student does not have a mental illness, and it is true. In other words, the student does not have a mental illness and the model correctly identifies it as negative. In the confusion matrix, this is the count of correctly identified negative cases.

The TP and TN values are desirable because they represent correct predictions, while FP and FN values are undesirable as they represent incorrect predictions. The goal is to maximize TP and TN and minimize FP and FN in order to achieve high classification accuracy. Based on the confusion matrix, multiple metrics were used to evaluate the machine learning models, including Precision, Recall, F1-score, and Accuracy. Accuracy measures the model's effectiveness in correctly categorizing data and can be used to compare the performance of different models. It is calculated as the fraction of correct positive predictions as given in Equation (1).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

In the context of mental health classification, precision refers to the proportion of true positives among all the positive predictions made by the model, calculated with Equation (2). In other words, it measures the accuracy of the model in predicting mental health issues among those who are predicted to have mental health issues. A high precision score indicates that the model is more accurate in correctly identifying students who have mental health issues, and is less likely to make false positive predictions. On the other hand, a low precision score indicates that the model is making many false positive predictions, which can lead to inappropriate interventions and waste of resources. Therefore, precision is an important metric to evaluate the performance of a machine learning model in the classification of mental health issues.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall as formulated with Equation (3) is another evaluation metric in machine learning classification models that measures the ability of the model to correctly identify all positive instances in the dataset. It is also called sensitivity and is calculated as the fraction of true positive predictions out of the total actual positive instances.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score is a metric that combines both precision and recall to provide a single score that represents the overall performance of the model. It is the harmonic mean of precision and recall, and it ranges from 0 to 1, with a higher score indicating better performance. The F1-score is a useful metric when there is an imbalance in the dataset, where the number of positive instances is significantly smaller than the number of negative instances. Equation (4) is the formula of F1-score.

$$F1 - score = \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}} \qquad (4)$$

Results and Discussion

The Performance of machine learning

Table 2-4 present the confusion matrix for SVM, Decision Tree, and Random Forest models, respectively. From the results, it was found that the Decision Tree model had the least Type-I error, while the SVM model had the least Type-II error. The best model will be discussed in the next section.

Table 2
Confusion Matrix of SVM

		Actual Values	
		Having mental health illness	Not having mental health illness
Predicted Values	Having mental health illness	0	20
	Not having mental health illness	0	216

Table 3
Confusion Matrix of Decision Tree

		Actual Values	
		Having mental health illness	Not having mental health illness
Predicted Values	Having mental health illness	15	5
	Not having mental health illness	72	144

Table 4
Confusion Matrix of Random Forest

		Actual Values	
		Having mental health illness	Not having mental health illness
Predicted Values	Having mental health illness	5	15
	Not having mental health illness	13	203

Based on the results of confusion matrix of each machine learning algorithm, the performance with different metrics is listed in Table 5. Accuracy, Precision, and Recall are all calculated as percentages since they represent the proportion of correct predictions made by the model over the total number of predictions. However, F1-score is not a percentage because it is a weighted average of Precision and Recall, which are themselves ratios of the number of correct predictions over the total number of relevant instances. Therefore, F1-score is a single number that ranges between 0 and 1, where a higher value indicates better performance.

Table 5
Accuracy, Precision, Recall, and F1-score

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
SVM	91.53	0	0	0
Decision Tree	67.37	75	17	28
Random Forest	88.14	25	28	26

According to Table 5, the SVM model had the highest accuracy percentage of 91.53% in predicting students' mental health issues, making it the best performing algorithm in this study. Despite having 0 Precision, Recall, and F1-score values, it was chosen because it accurately predicted that 91.5% of the students did not have mental health issues, making it the most reliable model for the task and data characteristics.

Interactive Visualization

This section provides a description of the interfaces of the interactive visualization dashboard. Figure 3 displays the main page of the website that visualizes the essential information on students' mental health during the COVID-19 pandemic.

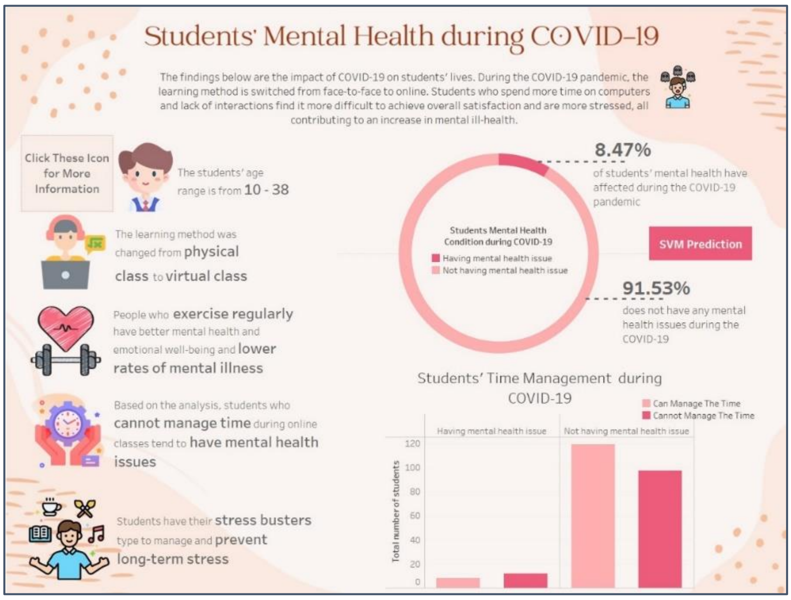

Figure 3. Main Page Dashboard

The data presented in Figure 3 reveals that there is a low prevalence of mental health issues among students during COVID-19, with only 8.47% of students experiencing such problems. Most students (91.53%) have been found to be in good mental health. Nonetheless, the data also suggests that both groups of students, those with and without mental health issues, might encounter time management difficulties amid the pandemic.

By clicking on the icons provided, users can interact with the data and access more information. For instance, when the Age icon is clicked, Figure 4 is displayed, which shows the mental health conditions of students by age group. Similarly, when the learning method icon is clicked, Figure 5 displays the information on learning methods.

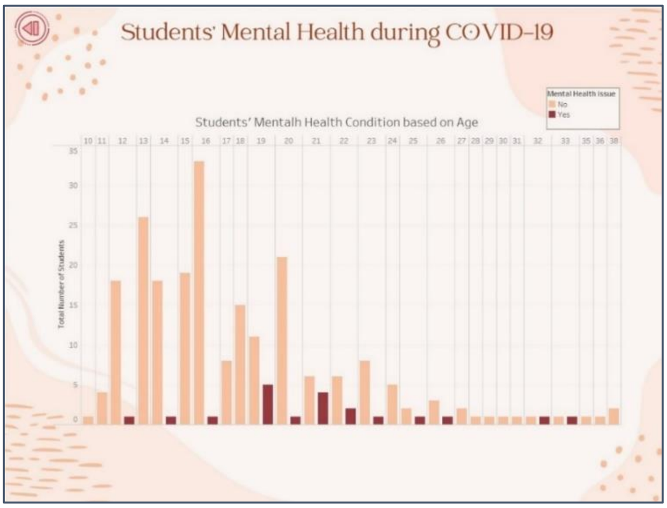

Figure 4. Students age categorization

The bar chart in Figure 4 shows that the highest number of students with mental illness is in the 19-year-old age group. Additionally, the chart suggests that the prevalence of mental health problems varies among different age groups, highlighting the importance of considering age as a significant factor when assessing the mental health of students during the COVID-19 pandemic. The findings presented in Figure 5 indicate that many students have rated their online class experience as 'Average', while 'Good' and 'Very Poor' ratings were also reported. These results imply that the students had a varied experience of online learning during the COVID-19 pandemic, with mixed feelings overall. In addition, the bar chart in Figure 5 shows that the average time spent on online classes indicates that there is no noteworthy variation in mental health outcomes among students who dedicate different durations of time to attend online classes. This discovery implies that the length of online classes might not be a crucial aspect in determining the mental well-being of students during the COVID-19 pandemic.

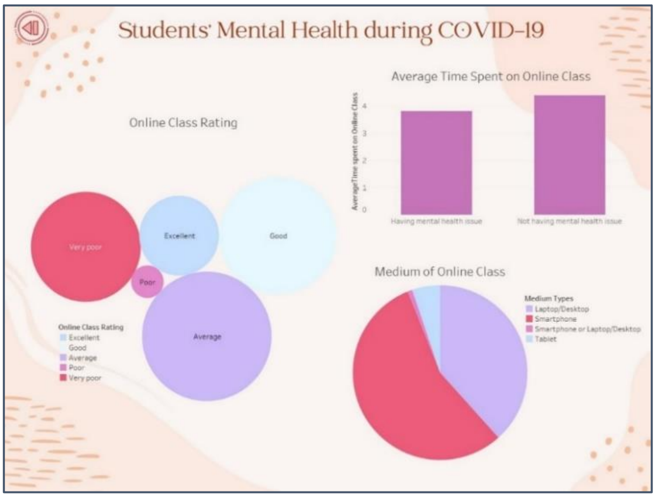

Figure 5. Information on learning methods

Moreover, Figure 6a, b, and c provide more detailed information on the factors contributing to students' mental health during the pandemic. Specifically, Figure 6a provides a visualization of the academic activities and involvement of students during the COVID-19 timeframe, while Figure 6b displays the average time spent on physical activities such as jogging, and Figure 6c represents the stress-busting activities of students.

a b

c

Figure 6. Factors contributing to the students' mental health

Finally, Figure 7 depicts the classification results of the students' mental health using SVM, displaying both the distribution of the actual data and the SVM prediction. Furthermore, it demonstrates a comparison of the accuracy of the two split ratios (70:30 and 80:20), indicating that the 80:20 split ratio has a higher accuracy than the 70:30 split ratio.

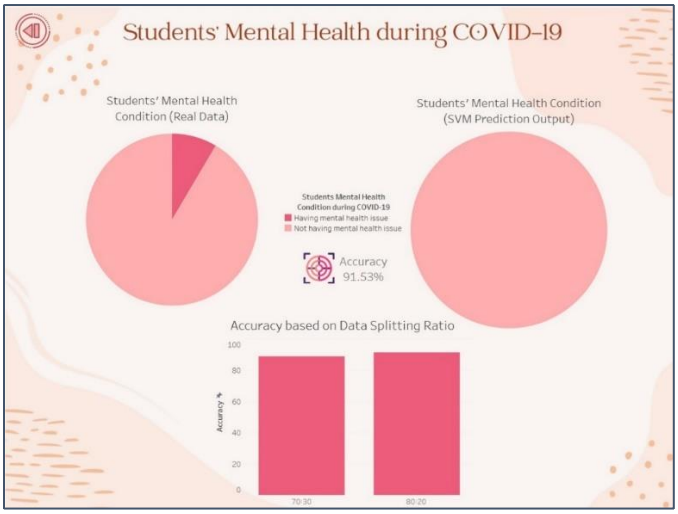

Figure 7. SVM Prediction Dashboard

An 80:20 split training:testing ratio logically has higher accuracy than the 70:30 because it provides the model with more data to train on. In an 80:20 split, 80% of the data is used for training the model, and the remaining 20% is used for testing. This means that the model has access to more data to learn from during the training phase, allowing it to better capture the patterns and relationships in the data. As a result, when the model is tested on the remaining 20% of data, it can make more accurate predictions compared to a model trained on only 70% of the data. However, one potential limitation of the study is the possibility of overfitting in the SVM model due to the small sample size. Further research with a larger and more diverse dataset is needed to determine if the results generalize to a broader population and more split ratios can be observed.

Conclusion

This research effectively utilized machine learning techniques to classify the mental health conditions of students during the COVID-19 pandemic. The SVM model yielded the highest accuracy in predicting mental health issues, especially when using an 80:20 data split ratio. The results were presented on a Tableau dashboard, offering a clear and concise way of visualizing the SVM predictions. Although the research found that the COVID-19 pandemic did not significantly impact the mental health of students, it is important to acknowledge that some students may still be struggling with mental health issues and require support. Therefore, it is essential to address these issues proactively to prevent them from worsening. This study's limitations regarding the potential for overfitting in the 80:20 data split ratio require further research. We hope that our findings will serve as a valuable resource for educational institutions to recognize and support students who may be experiencing challenges with their mental health during the COVID-19 pandemic and beyond. Ultimately, we aspire that our project will contribute to the development of effective mental health support strategies for students.

Acknowledgements

The authors gratefully acknowledge the Universiti Teknologi MARA (UiTM), Perak branch.

Conflict of Interest

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

Anitha, A., and Asha, A. (2018). "Improving Prediction Accuracy Using Hybrid Machine Learning Algorithm on Medical Datasets." *International Journal of Scientific & Engineering Research* 9(October 2018).

Burkart, N., and Marco, F. H. (2021). "A Survey on the Explainability of Supervised Machine Learning." *Journal of Artificial Intelligence Research* 70:245–317.

Chrikov, I., Krista, M. S., Horgos, B., and Jones-White, D. (2020). "Undergraduate and Graduate Students' Mental Health during the COVID-19 Pandemic."

Shuang, G., Calhoun, V. D., and Sui, J. (2018). "Machine Learning in Major Depression: From Classification to Treatment Outcome Prediction." *CNS Neuroscience & Therapeutics* 24(11):1037–52.

Mheidly, N., Mohamad, Y.. Fares, and Fares, J. (2020). "Coping with Stress and Burnout Associated with Telecommunication and Online Learning." *Frontiers in Public Health* 672.

Padala, V. S., Gandhi, K., and Pushpalatha, D. V. (2019). "Machine Learning : The New Language for Applications." *IAES International Journal of Artificial Intelligence (IJ-AI)* 8(4):411–21. doi: 10.11591/ijai.v8.i4.

Padilla, L. M., Sarah, H., Creem-Regehr, M. H., and Jeanine, K. S. (2018). "Decision Making with Visualizations: A Cognitive Framework across Disciplines." *Cognitive Research: Principles and Implications* 3(1):1–25.

Panicker, S. S., and Gayathri, P. (2019). "A Survey of Machine Learning Techniques in Physiology Based Mental Stress Detection Systems." *Biocybernetics and Biomedical Engineering* 39(2):444–69.

Priya, A., Garg, S., and Tigga, N. P. (2020). "Predicting Anxiety, Depression and Stress in Modern Life Using Machine Learning Algorithms." *Procedia Computer Science* 167:1258–67.

Barnaly, R., and Calhoun, V. (2020). "Towards a Brain-Based Predictome of Mental Illness." *Human Brain Mapping* 41(12):3468–3535.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D. (2018). "What Do We Talk about When We Talk about Dashboards?" *IEEE Transactions on Visualization and Computer Graphics* 25(1):682–92.

Shatte, A. B. R., Delyse, M. H., and Samantha, J. T. (2019). "Machine Learning in Mental Health: A Scoping Review of Methods and Applications." *Psychological Medicine* 49(9):1426–48.

Vellido, A. (2020). "The Importance of Interpretability and Visualization in Machine Learning for Applications in Medicine and Health Care." *Neural Computing and Applications* 32(24):18069–83.

Zoabi, Y., Deri-Rozov, S., and Shomron, N. (2021). "Machine Learning-Based Prediction of COVID-19 Diagnosis Based on Symptoms." *Npj Digital Medicine* 4(1):1–5.