

Phân tích hành vi di chuyển của sinh viên cho ứng dụng nhận diện stress hằng ngày

Người trình bày

Phan Hoàng Khang (2111460)



Thầy cô và các bạn có thể xem bài viết tại đây

1. Giới thiệu

- Theo WHO, căng thẳng tâm lý (stress) là hiện tượng đặc trưng có tinh thần của một người bị căng thẳng, hoặc lo lắng từ nhiều nguồn khác nhau như công việc và học tập [1]
- 37% người Mỹ trong năm 2022 đã gặp phải stress (tăng 6% so với 2021) [2]
- Hậu quả là thế giới đã mất 55% số ngày làm việc vì sức khỏe tâm lý [3]



[1] World Health Organization, "Stress," <https://www.who.int/news-room/questions-and-answers/item/stress/>

[2] P. N. Alert, Psych News Alert, Dec. 22, 2022, <https://alert.psychnews.org/2022/12/more-americans-rate-their-mental-health.html>

[3] American Psychiatric Association, <https://www.psychiatry.org/news-room/news-releases/americans-anticipate-higher-stress-at-the-start-of>.

1. Giới thiệu

Vậy làm sao để sớm
nhận diện các vấn đề
tinh thần?

room/question

[2] P. N. Alert, R.

[3] American Psychiatric Association
the-start-of.

g-ann... self-mental-health.html
ses/americans-anticipate-higher-stress-at-

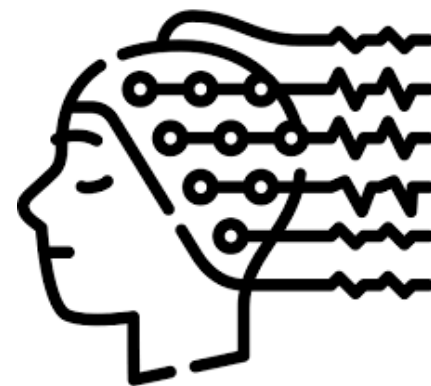
1. Giới thiệu



Việc điền các
bảng câu hỏi đôi
khi mất thời gian,
thiếu tính định
lượng



Nhận diện bằng khuôn
mặt là một giải pháp tốt
nhưng khá tốn công sức
và rườm rà



Nhận diện bằng EEG là
một giải pháp đưa được
nhiều thông tin nhưng
khá đắt đỏ và không thể
sử dụng hằng ngày

1. Giới thiệu

Nó không thật sự tiện lợi!

thiết
lượng

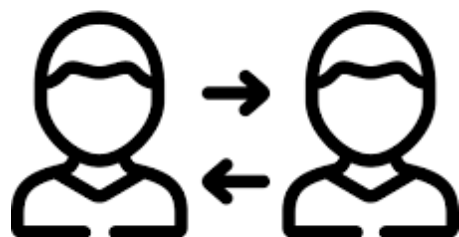
Su... ngày

g thể

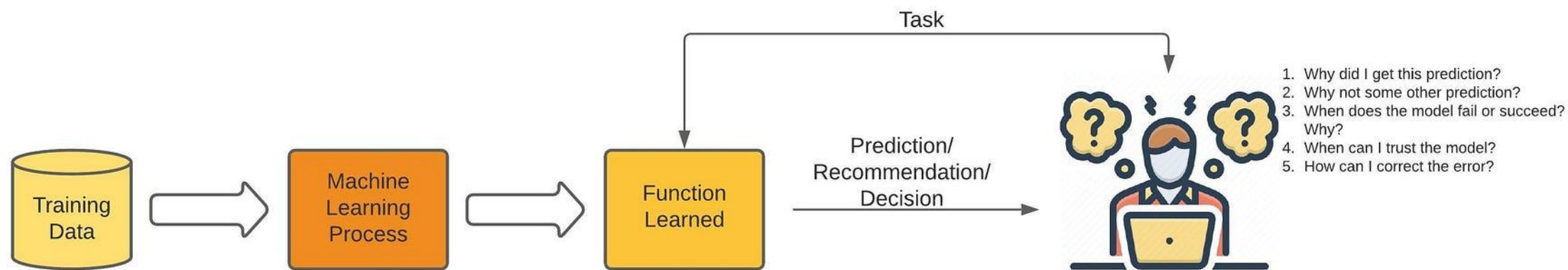


- **Ubiquitous Computing** là câu trả lời cho việc thu thập tín hiệu liên tục, phân tích các hành vi của con người và hiểu thêm về con người

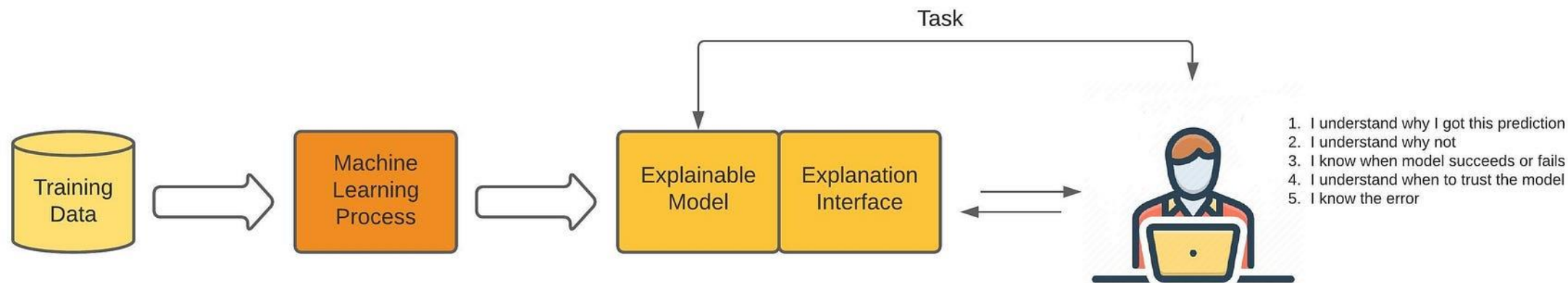
1. Giới thiệu



AI



XAI



XAI (eXplainable Artificial Intelligence) giải quyết được việc hiểu về các giá trị thu thập được trước đó, góp phần lý giải các tác động của hành động và hành vi đến đối tượng khảo sát (căng thẳng tâm lý)

2. Nghiên cứu liên quan

**Nhưng các nghiên cứu
này nghiên cứu trong
môi trường có kiểm soát!**

- [4] M. Salai, I. Vassányi, et al., "A Survey on the Role of Wearable Devices in Smart Homes," *IEEE Access*, vol. 11, pp. 10111–10124, 2023.
- [5] H. Jebelli et. al, "Automation in Smart Homes: A Survey," *IEEE Access*, vol. 9, pp. 10111–10124, 2021.
- [6] Y. Xiao et. al, "Proc. ACM Interact. Mob. Wearable Ubiquitous Computing," vol. 8, pp. 1–14, 2024.

2. Nghiên cứu liên quan



Nghiên cứu của Salai trên tín hiệu nhịp tim đã mang lại một phương pháp để nhận diện đạt độ chính xác 75% [4]



Houtan và các đồng nghiệp đã phát triển hệ thống nhận diện thông qua tín hiệu EEG đạt độ chính xác tối đa là 80% [5]



Nhóm của Xiao đã phát triển hệ thống nhận diện dựa trên nhiệt hồng ngoại. Với phương pháp này, nhóm của Xiao đã đạt được độ chính xác 76% [6]

[4] M. Salai, I. Vassányi, and I. Kósa, J. Healthc. Eng., vol. 2016, pp. 1–13, 2016.

[5] H. Jebelli et. al, Automation in Construction, vol. 93, pp. 315–324, 2018.

[6] Y. Xiao et. al Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 7, Jan. 2024.

2. Nghiên cứu liên quan

Mục tiêu nghiên cứu:

- Sử dụng dữ liệu GPS liên tục phong phú được ghi lại trước để trích xuất các tính năng mạnh mẽ từ dữ liệu GPS.
- Tận dụng các tính năng được trích xuất, nghiên cứu này nỗ lực thực hiện dự đoán căng thẳng hàng ngày bằng các mô hình học máy.
- Bằng cách xem xét kỹ lưỡng các tính năng mới được trích xuất từ dữ liệu GPS, nghiên cứu này đi sâu vào mối tương quan giữa mức độ căng thẳng và hành vi của học sinh trong tự nhiên.

3. Phương pháp

Đặc trưng	Số lượng
Đặc trưng thời gian	2
Đặc trưng học tập	9
Đặc trưng vị trí	39

Bảng 1: Bảng đặc trưng và thống kê số lượng đặc trưng

3. Phương pháp

3.1. Bộ dữ liệu

Nghiên cứu này sử dụng nhiều tập dữ liệu trong StudentLife Dataset (thu thập vào năm 2014) [7] bao gồm:

- Thời gian, nhà cung cấp, loại mạng, vĩ độ và kinh độ từ tập dữ liệu GPS StudentLife
- Lớp học của sinh viên từ tập dữ liệu lớp StudentLife
- Thông tin lớp học từ tập dữ liệu thông tin lớp StudentLife
- Hạn chót, từ tập dữ liệu hạn chót StudentLife
- Thông tin căng thẳng từ tập dữ liệu EMA StudentLife

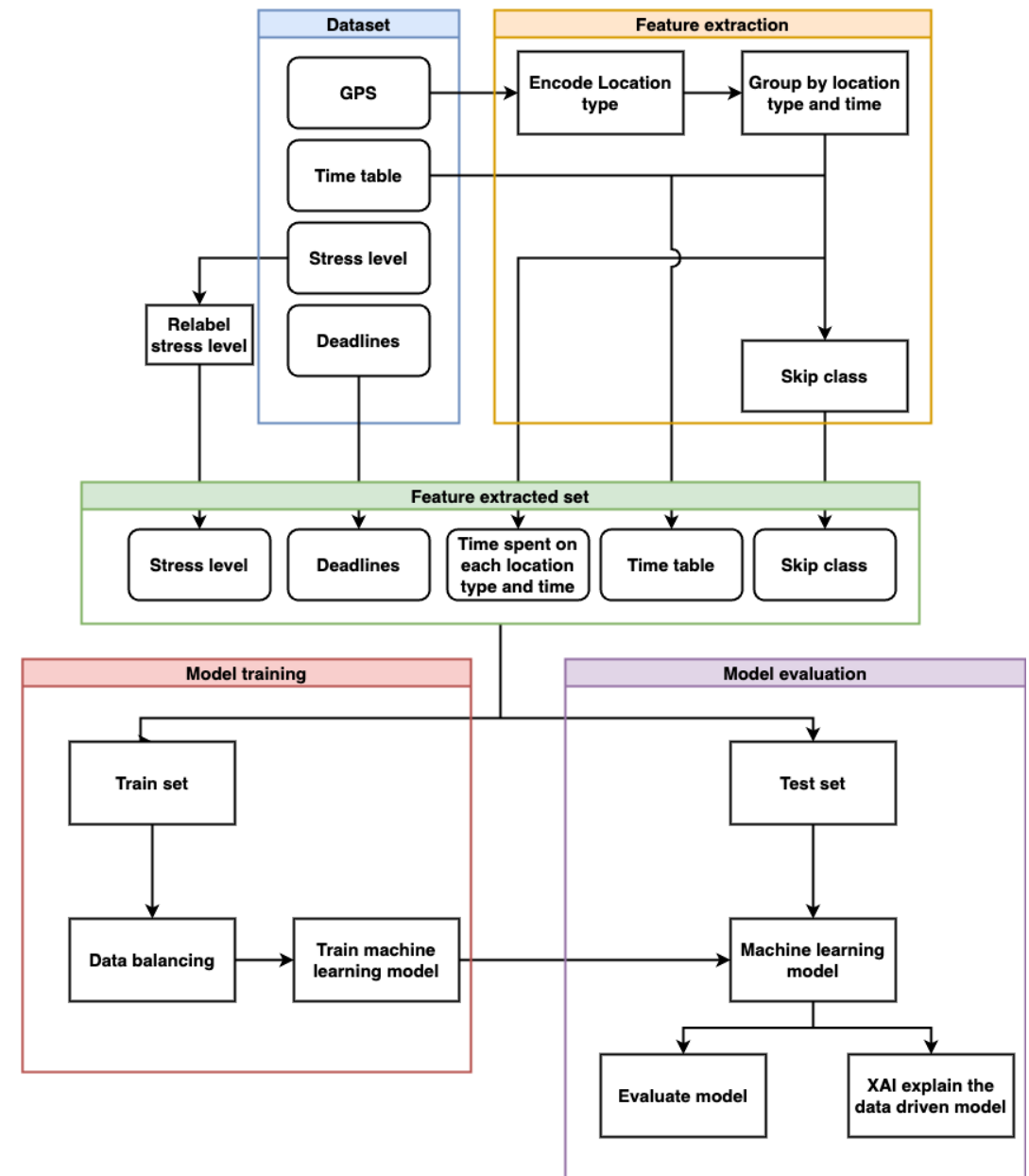
[7] R. Wang et al., in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3-14, 2014.

3. Phương pháp

3.2. Lưu đồ xử lý

Phương pháp gồm 4 phần chính:

- **Dataset** là bộ dữ liệu đầu vào
- **Feature extraction** là khối trích xuất đặc trưng
- **Model training** là khối dùng để tạo dữ liệu cho huấn luyện
- **Model evaluation** là khối dùng để đánh giá kết quả khối model training, và dùng để giải thích cách mô hình nhìn nhận vấn đề tâm lý của sinh viên bằng XAI



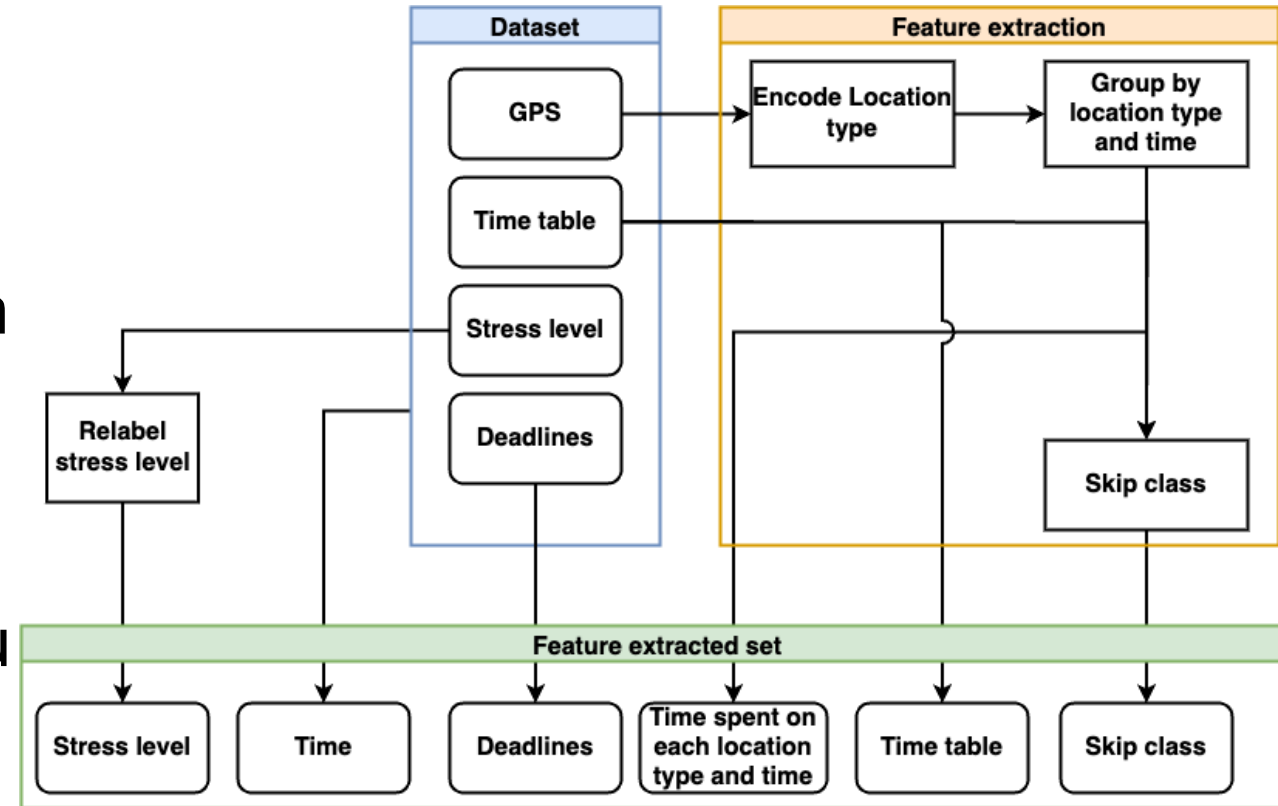
Hình 1: lưu đồ hệ thống

3. Phương pháp

3.3. Trích xuất đặc trưng

Nghiên cứu sẽ tập trung vào 3 nhóm đặc trưng chính bao gồm:

- Đặc trưng học tập thể hiện những yếu tố xoay quanh việc học của sinh viên
- Đặc trưng vị trí thể hiện rõ sinh hoạt hàng ngày của sinh viên
- Đặc trưng thời gian thể hiện các yếu tố theo thời gian

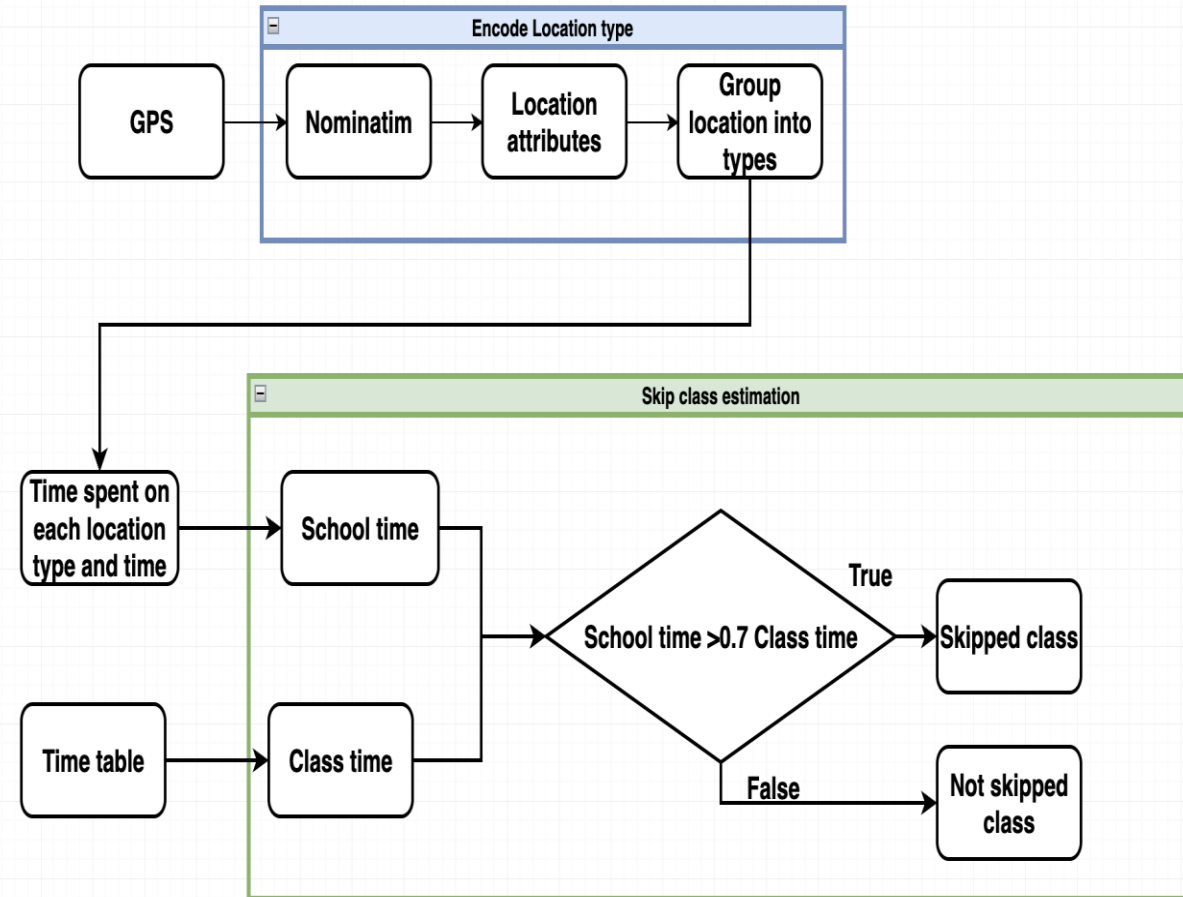


3. Phương pháp

3.3. Trích xuất đặc trưng

Đặc trưng vị trí, đặc trưng cho sự nghỉ học được trích xuất như sơ đồ bên:

1. Sử dụng kinh độ và vĩ độ từ GPS đưa vào Nominatim API
2. Thu thập giá trị trả về từ API
3. **Chuyển** các giá trị này về dạng các **thuộc tính** như nhà, trường, quán bar, ...
4. Gộp các thuộc tính trên lại thành nhóm và **tính tổng thời gian** (theo giây) ở các nhóm.
5. Kết hợp yếu tố thời gian từ 4. và thời khoá biểu của sinh viên, đưa ra *kết luận sinh viên có nghỉ học hay không?*

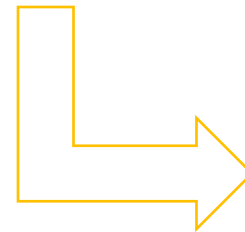


3. Phương pháp

3.4. Tái đánh dấu mẫu stress



Sắp xếp lại bộ dữ liệu
chia stress thành 5 mức
độ từ vui vẻ đến căng
thẳng nặng

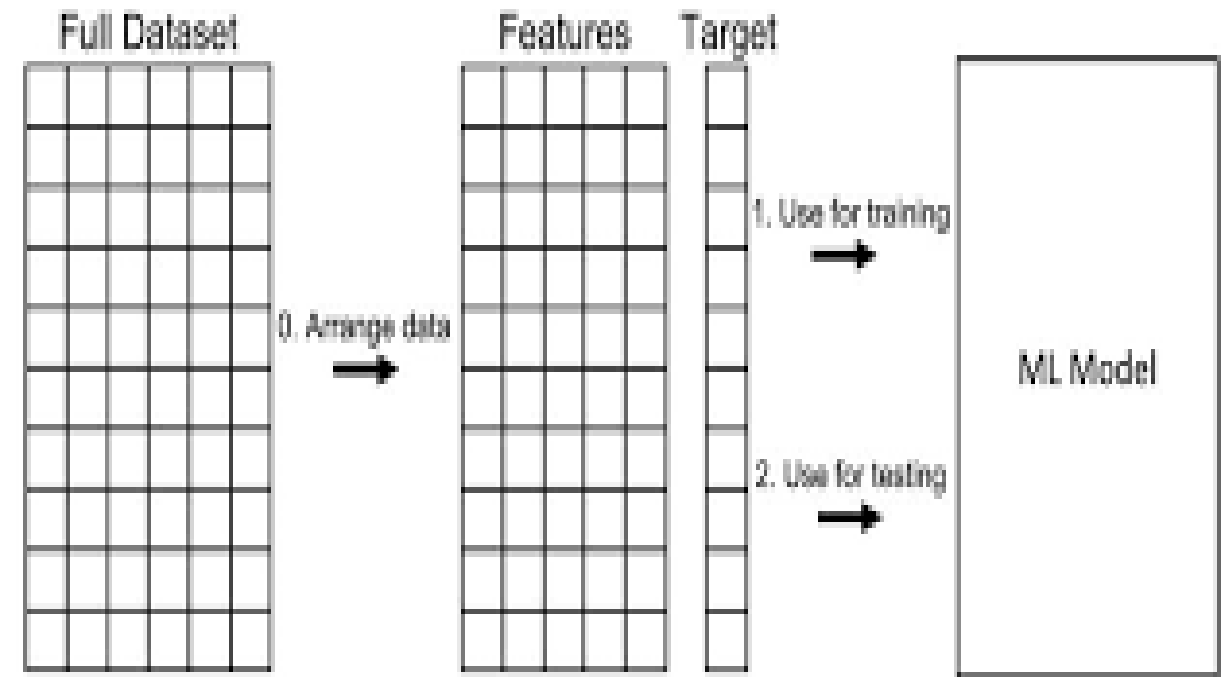


- Chia làm 2. mức độ căng thẳng (có, không) và 3 mức độ (không, có vừa, có cao)

3. Phương pháp

3.5. Huấn luyện mô hình

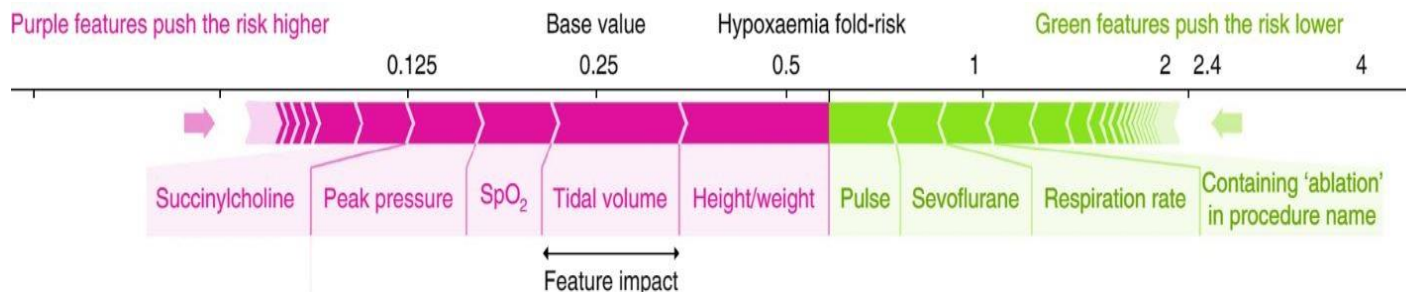
- Bộ dữ liệu đã xử lý được tách ngẫu nhiên riêng lẻ thành các tệp gấp Train và Test với tỷ lệ 7:3.
- 70% dữ liệu huấn luyện được cân bằng bằng SMOTE+ENN (trường hợp 2 lớp) hoặc SMOTE+Tomek-Links (trường hợp 3 lớp)
- Mô hình XGBoost đã được sử dụng với focal loss để đào tạo.
- Hơn nữa, mô hình Random Forest cũng đã được sử dụng để so sánh với mô hình XGBoost



3. Phương pháp

3.6. Phân tích mô hình

- Trong công trình này, chúng tôi đã sử dụng các giá trị SHAPley để đánh giá các mô hình của mình nhằm đo lường hệ số tương quan giữa căng thẳng và hành vi của học sinh trong một ngày.
- Các giá trị SHAPley là một phương pháp để đo lường sự đóng góp của từng tính năng vào đầu ra phân loại một cách công bằng



4. Kết quả và Phân tích

4.1. Hiệu suất Phân loại

- Có thể thấy rằng mô hình Random Forest đạt được độ chính xác cao hơn một chút và điểm F1 so với mô hình XGBoost
- Độ chính xác của mô hình Random Forest đạt 79% cao hơn 4% so với phương pháp thu thập chủ động [4]

	2 class classification scenario			3 class classification scenario		
	XGB	Random Forest	SVM	XGB	Random Forest	SVM
Accuracy	79	79	39	65	66	33
Weighted F1 score	80	81	45	61	62	37
Macro F1 score	60	63	37	45	51	31
F1 score for class stress	88	87	48	80	78	43

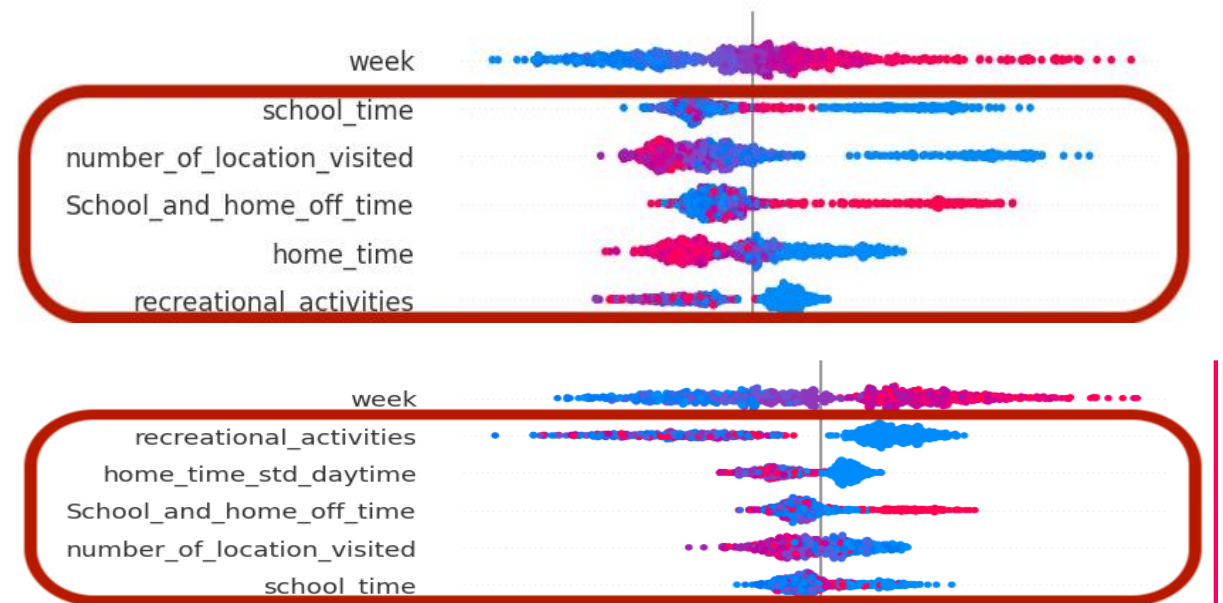
Bảng 2: Báo cáo phân loại các mô hình

[4] M. Salai, I. Vassányi, and I. Kósa, 'Stress detection using low cost heart rate sensors', J. Healthc. Eng., vol. 2016, pp. 1–13, 2016.

4. Kết quả & Phân tích

4.2. Phân tích đặc trưng

- Tuần là đặc điểm tác động nhiều nhất đến nhận dạng căng thẳng
- Về thời gian học, thời gian ở nhà và thời gian không ở nhà và ở trường có tác động lớn đến quyết định của mô hình

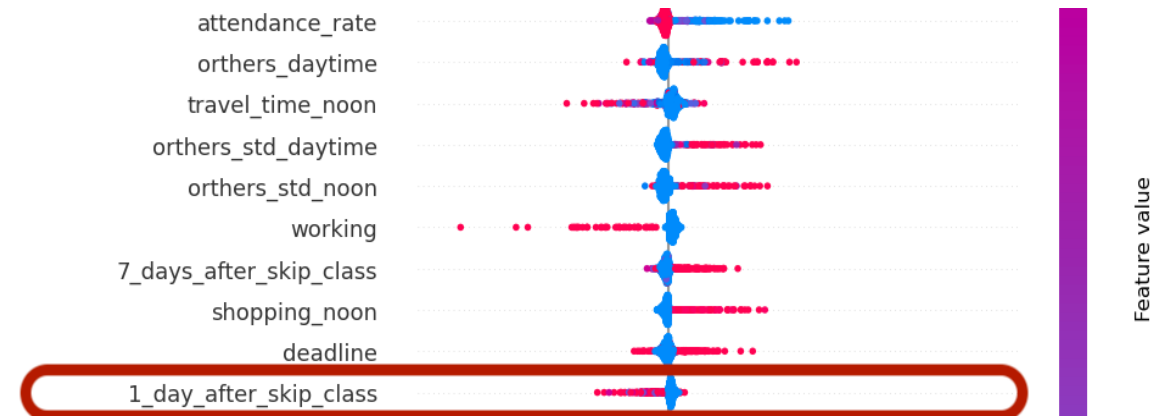


Hình 2: Biểu đồ điểm của giá trị SHAPley của kịch bản 3 lớp (trên) và kịch bản 2 lớp (dưới)

4. Kết quả & Phân tích

4.2. Phân tích đặc trưng

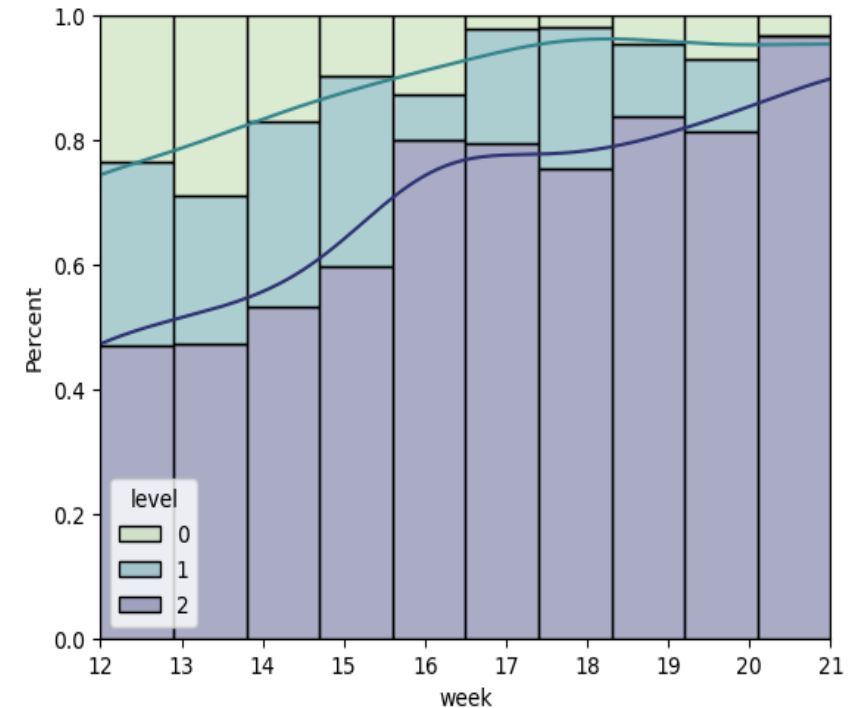
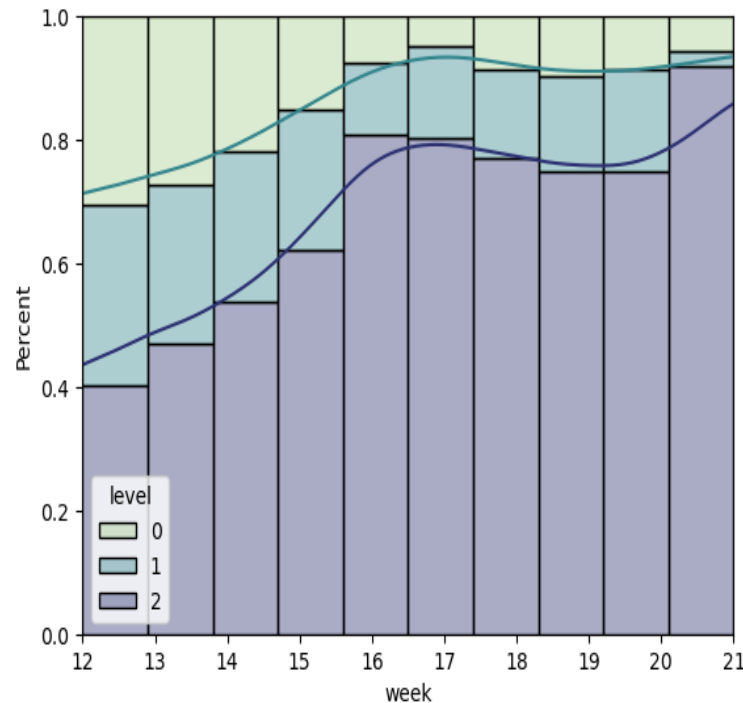
- Về việc nghỉ học, phân tích cho thấy học sinh nghỉ học có ít khả năng bị căng thẳng cao hơn
- Nhưng 7 ngày sau khi nghỉ học, sinh viên lại có khả năng bị stress cao hơn.



Hình 3: Biểu đồ điểm của giá trị SHAPley của kịch bản 3 lớp

5. Bàn luận

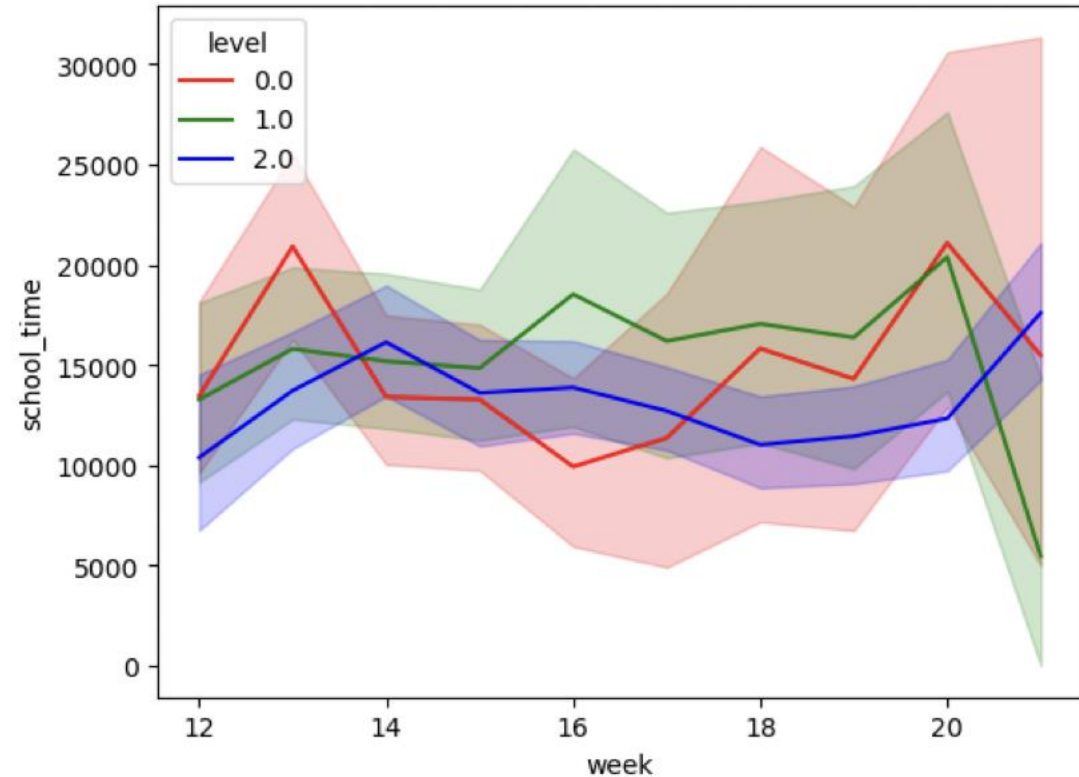
- Trong phân tích sâu hơn, trong nhóm bỏ học, tỷ lệ học sinh căng thẳng tăng theo thời gian
- Trong khi đó, tỷ lệ học sinh không bỏ học tăng đến hết tuần 16-17 (tuần cuối của kỳ thi giữa kỳ) sau đó giảm nhẹ



Hình 4: Tỷ lệ học sinh gặp vấn đề căng thẳng trong nhóm không trốn tiết (trái) và trốn tiết (phải)

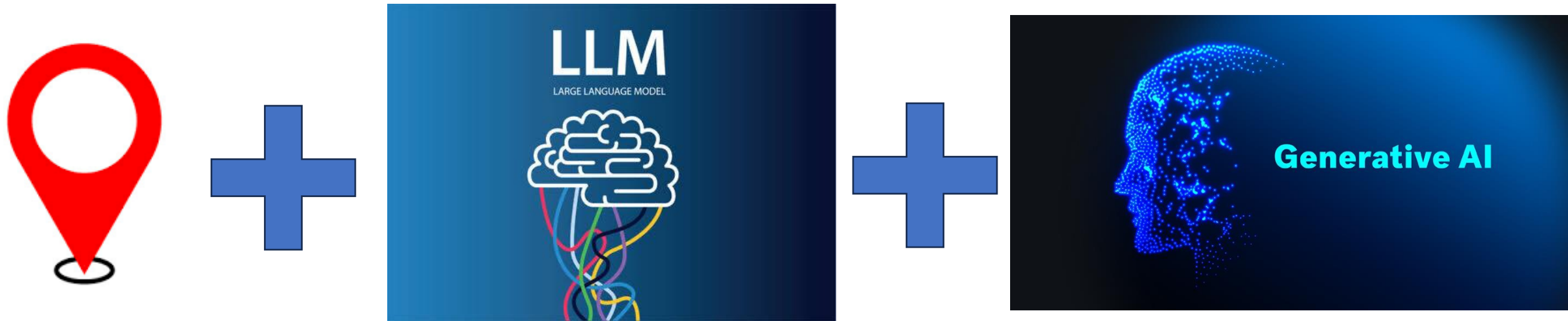
5. Bàn luận

- Sinh viên có thời gian trên trường **trung bình (khoảng 3-4 giờ trên trường)** thường có xu hướng cảm thấy vui vẻ
- Càng về cuối kì sinh viên có **xu hướng trên trường** nhiều hơn
- Sinh viên có mức độ **căng thẳng vừa** có thời gian lên trường nhiều nhất, trong khi sinh viên bị **căng thẳng nặng** lên trường ít nhất



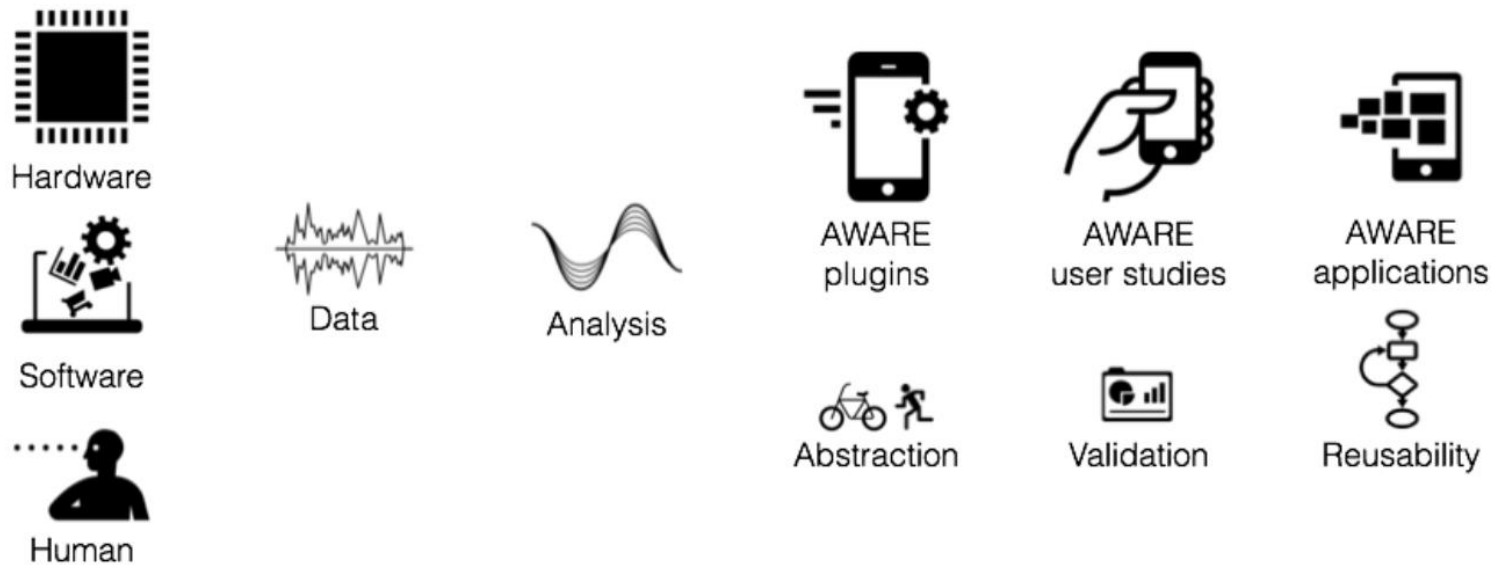
Hình 5: Biểu đồ biểu diễn thời gian đi học của sinh viên theo tuần chia theo mức độ căng thẳng

6. Hướng phát triển



- Đối với các công trình trong tương lai, chúng tôi đề xuất tích hợp kiến thức LLM để xác định hoạt động của mọi người chính xác hơn.
- Chúng tôi đề xuất các phương pháp học tăng cường có thể học hỏi từ từng mô hình cá nhân để cải thiện kết quả đánh giá sức khỏe tâm thần.

Hướng phát triển



Hình 6: Hệ thống thu thập dữ liệu đề xuất

Vì dữ liệu hiện tại được thu thập từ năm 2014 nên tại thời điểm hiện tại việc các yếu tố này có thể không còn chính xác ở thực tế.

Vậy nên tôi đề xuất dung hệ thống thu thập như hình bên để thu thập được thêm các dữ liệu.

7. Kết luận

- Trong công trình này, 50 đặc điểm liên quan đến hoạt động, hành vi và thời gian của sinh viên được trích xuất. Chỉ số tuần nào có tác động lớn nhất đến căng thẳng của sinh viên
- Kết quả phân tích đặc điểm cho thấy giá trị quan trọng cao trong các đặc điểm được trích xuất, đặc biệt là các đặc điểm dựa trên vị trí và đặc trưng học tập. Đặc biệt, chúng tôi đã phát hiện ra mối tương quan giữa hành vi của sinh viên với trạng thái căng thẳng của họ.
- Mô hình Rừng ngẫu nhiên đã được sử dụng và thể hiện hiệu suất đáng kể trong các nhiệm vụ dự đoán căng thẳng trong bệnh nhân trên tập dữ liệu lớp mất cân bằng.

Reference

- [1] World Health Organization, "Stress," <https://www.who.int/news-room/questions-and-answers/item/stress/>
- [2] P. N. Alert, "More Americans Rate Their Mental Health Worse Compared With a Year Ago, Poll Finds," Psych News Alert, Dec. 22, 2022, <https://alert.psychnews.org/2022/12/more-americans-rate-their-mental-health.html>
- [3] American Psychiatric Association, "Americans Anticipate Higher Stress at the Start of 2023 and Grade Their Mental Health Worse," <https://www.psychiatry.org/news-room/news-releases/americans-anticipate-higher-stress-at-the-start-of>.
- [4] M. Salai, I. Vassányi, and I. Kósa, 'Stress detection using low cost heart rate sensors', J. Healthc. Eng., vol. 2016, pp. 1–13, 2016.
- [5] H. Jebelli et. al, "Eeg-based workers' stress recognition at construction sites," Automation in Construction, vol. 93, pp. 315–324, 2018.
- [6] Y. Xiao et. al, "Reading between the heat: Co-teaching body thermal signatures for non-intrusive stress detection," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 7, Jan. 2024.
- [7] R. Wang et al., "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3-14, 2014.



Questions & Answers



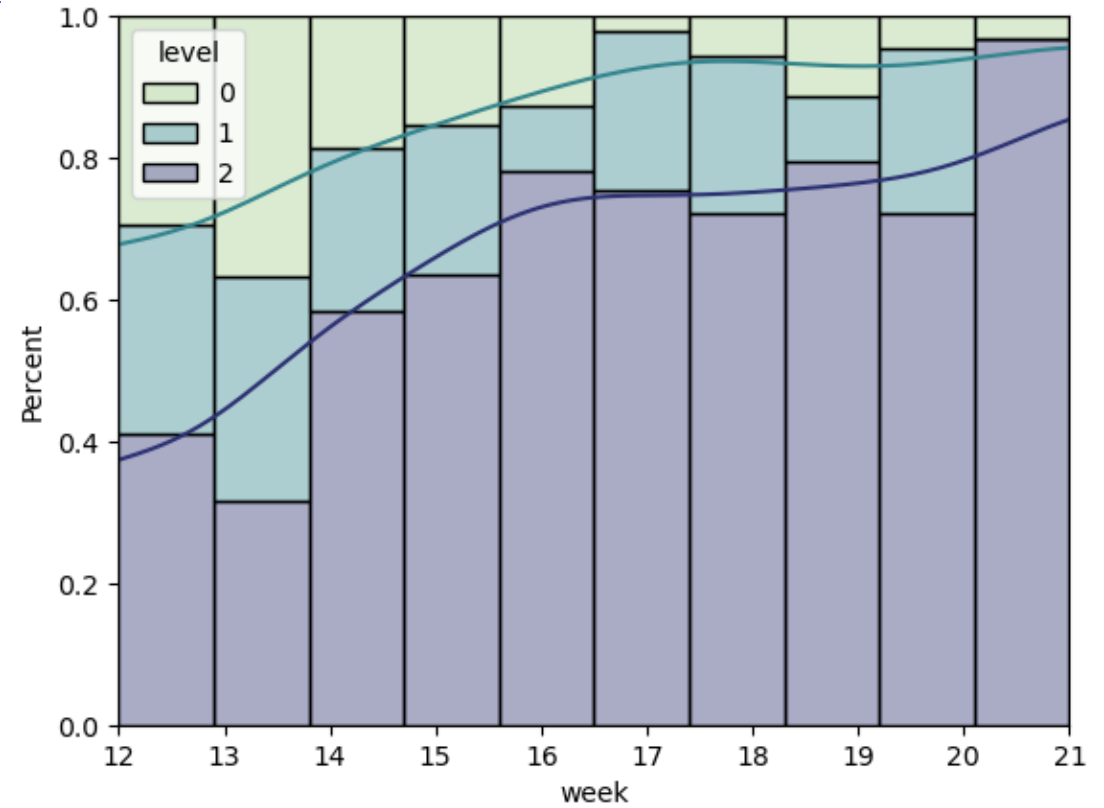
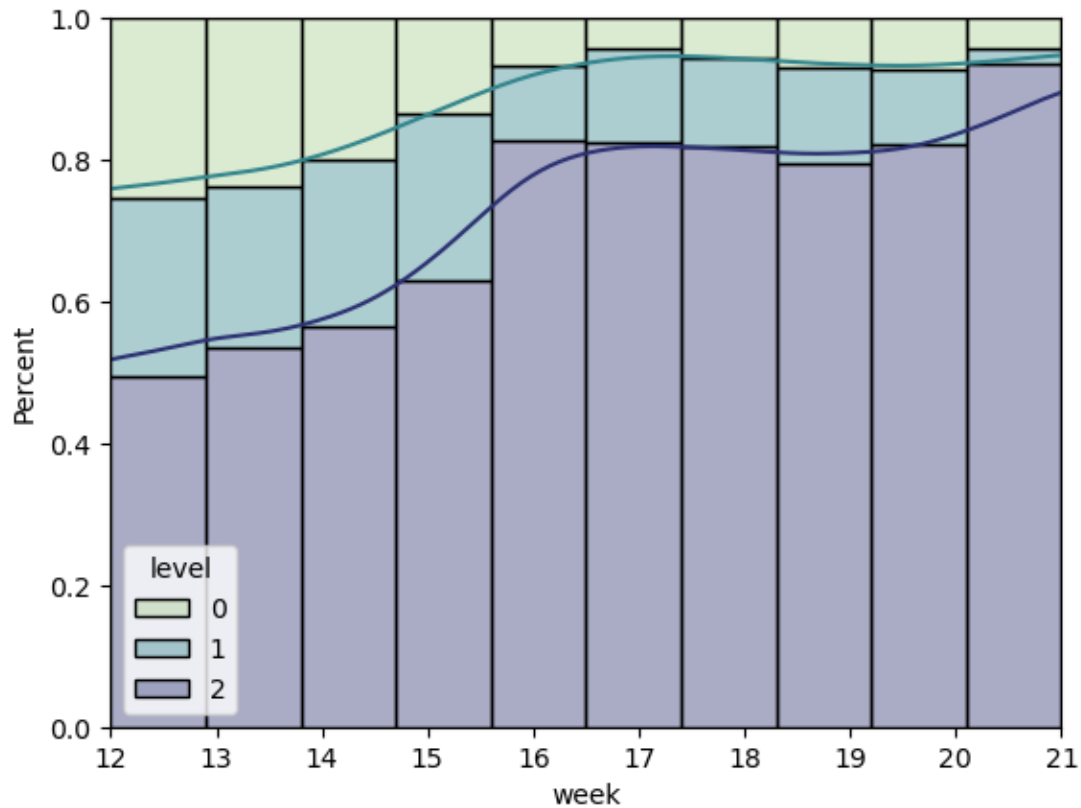
Các bạn có thể xem
bài viết tại đây

Cảm ơn mọi người đã lắng nghe
__Phan Hoàng Khang__

Bài nghiên cứu này đã được trình bày tại ABC 2024 và ISAS 2024.

End of slide show.

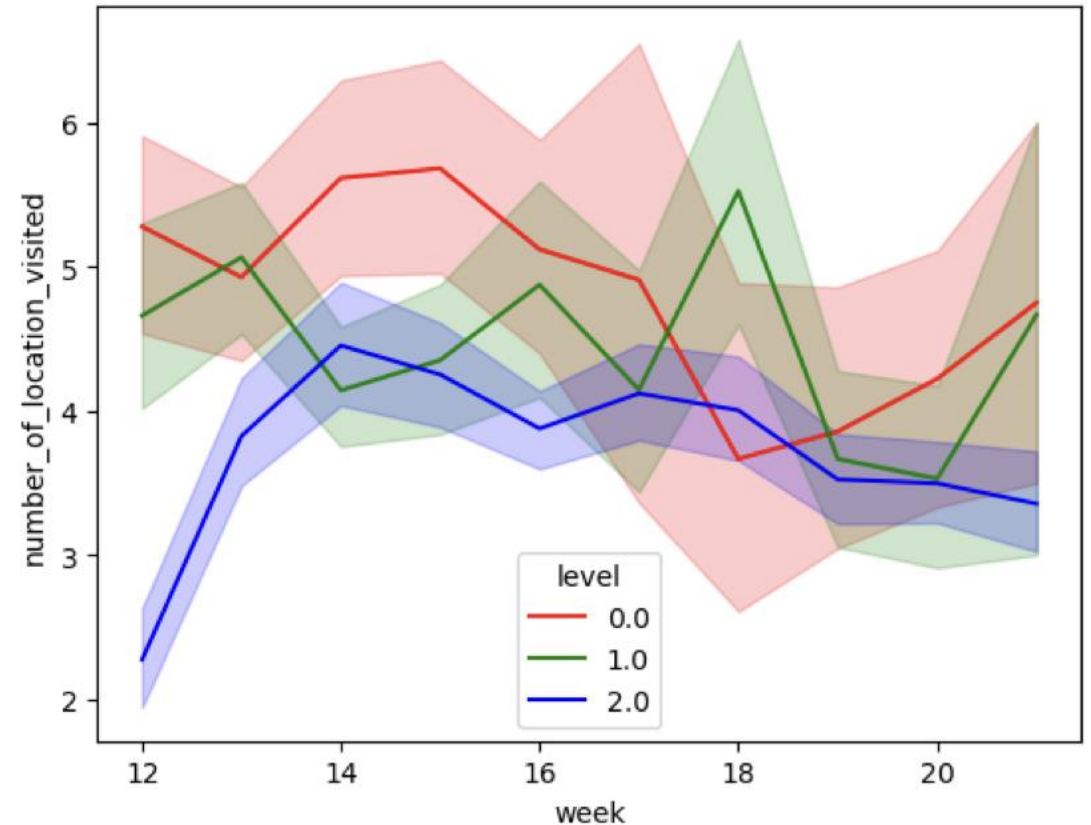
Appendix



Hình Tỷ lệ học sinh gặp vấn đề căng thẳng, trong nhóm không trốn tiết ngày hôm trước (trái) và trốn tiết ngày hôm trước (phải)

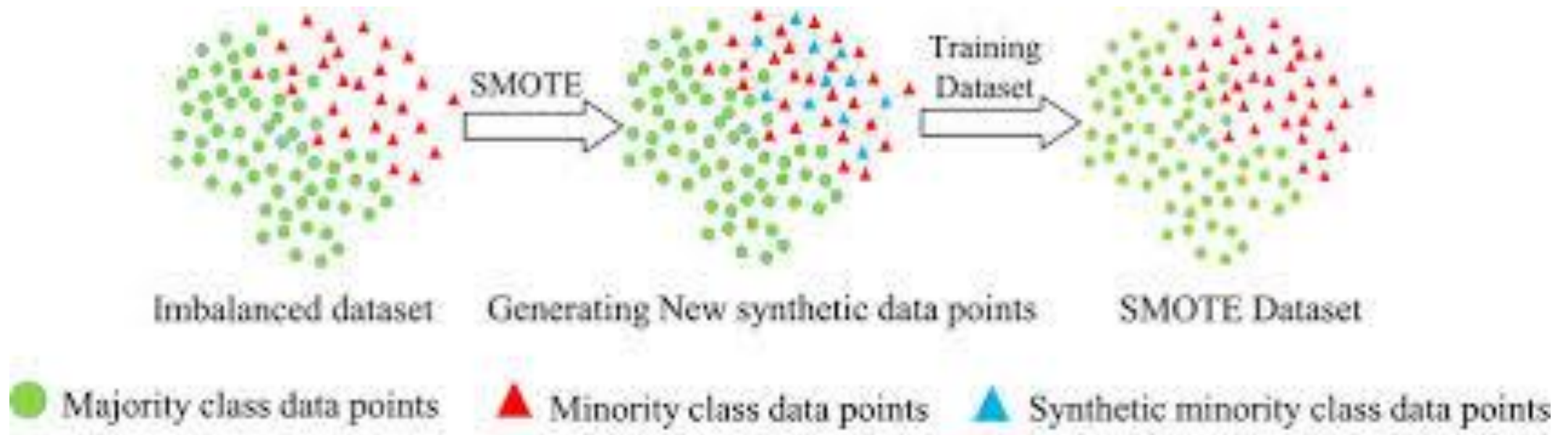
Appendix

- Số địa điểm sinh viên đến **tỉ lệ nghịch** với mức độ nghiêm trọng của stress
- Sinh viên có xu hướng chỉ đi đến 4 đến 5 địa điểm hằng ngày. Tôi đề xuất khai phá rõ hơn về vì sao sinh viên lại chỉ đến vài địa điểm hằng ngày chứ không nhiều hơn hoặc ít hơn



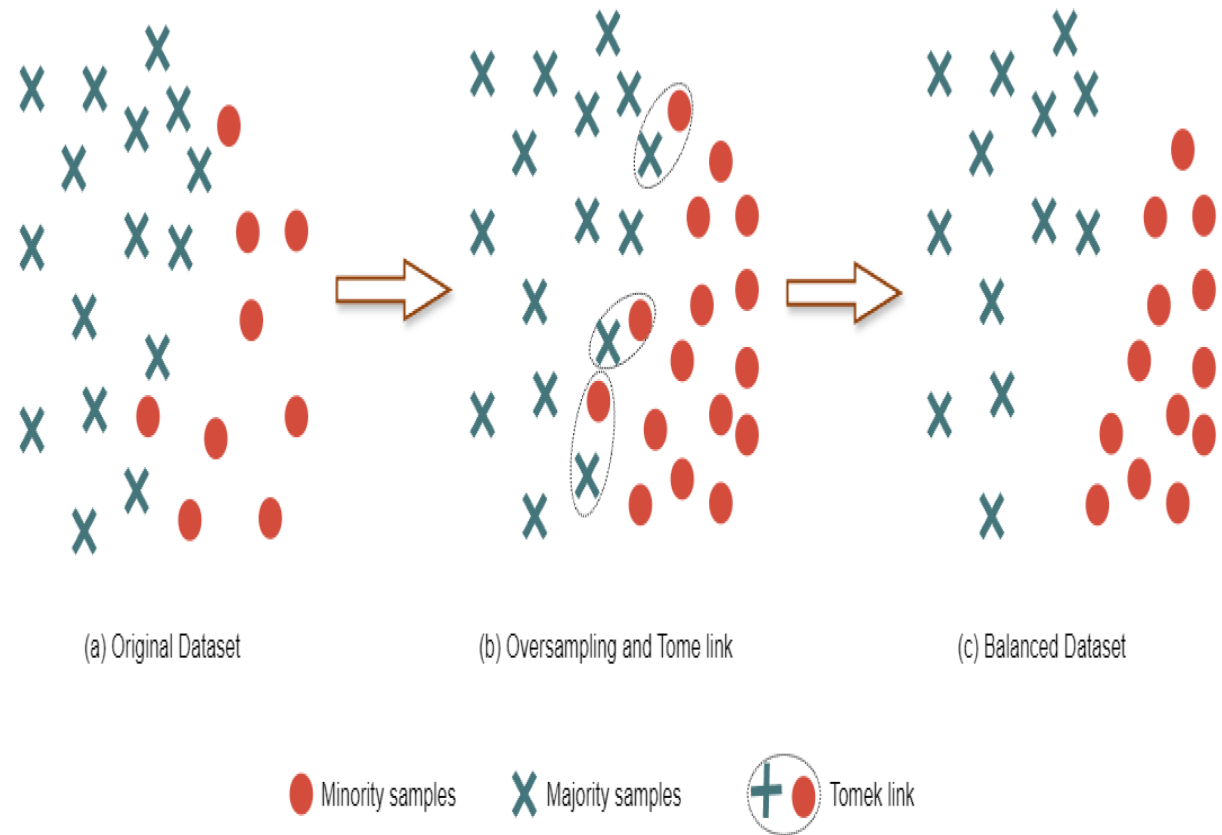
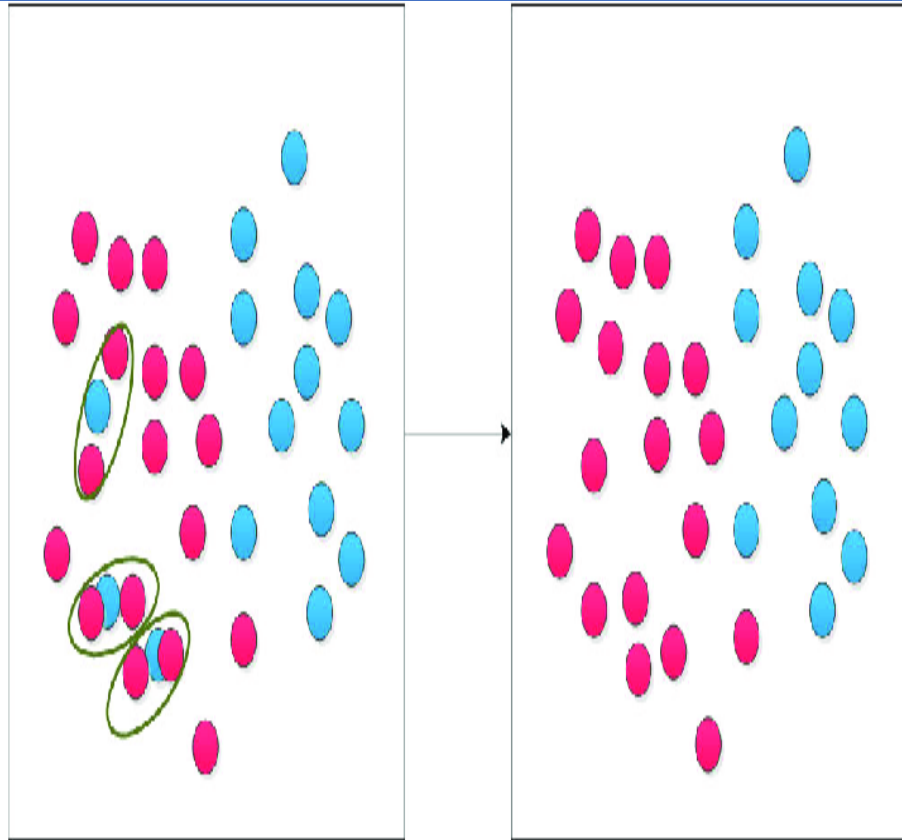
Hình Biểu đồ biểu diễn số lượng địa điểm đến của sinh viên theo ngày chia theo tuần chia theo mức độ căng thẳng

Appendix



Hình Phương pháp SMOTE(Synthetic Minority Over-sampling)

Appendix



Hình ENN (Edited Nearest Neighbors) (trái) và SMOTE-Tomek-Links (phải)