

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN**  
**DỰ ĐOÁN CHẤT LƯỢNG GIẤC NGỦ**  
**LỚP: CS117.O22**

**GVHD: Ngô Đức Thành**

**Nhóm: 14**

**Trần Đình Khánh Đăng - 22520195**

**Lê Minh Nhựt - 22521060**

**Huỳnh Danh Đạt - 22520211**

**Lưu Đoàn Ngọc Phát - 22521070**

## MỤC LỤC

<b>I. Problem .....</b>	<b>3</b>
<b>1. Introduction.....</b>	<b>3</b>
<b>2. Description .....</b>	<b>3</b>
<b>II. Computational Thinking.....</b>	<b>5</b>
<b>1. Decomposition .....</b>	<b>5</b>
<b>2. Pattern Recognition .....</b>	<b>6</b>
<b>2.1. Mapping.....</b>	<b>6</b>
<b>2.2. Grouping.....</b>	<b>7</b>
<b>3. Abstraction / Genelization .....</b>	<b>9</b>
<b>3.1. Abstraction .....</b>	<b>9</b>
<b>3.2. Generalization .....</b>	<b>10</b>
<b>4. Evaluation.....</b>	<b>11</b>
<b>III. Conclusion .....</b>	<b>12</b>
<b>IV. Demo .....</b>	<b>13</b>
<b>V. References .....</b>	<b>14</b>

# I. Problem

## 1. Introduction

Giấc ngủ từ xưa đến nay luôn là một nhu cầu thiết yếu trong cuộc sống con người. Giấc ngủ tốt sẽ đem lại cho ta một cảm giác thoải mái, tràn đầy năng lượng để làm việc. Để có một giấc ngủ tốt đòi hỏi rất nhiều yếu tố, và xác định xem các yếu tố đó ảnh hưởng ra sao tới giấc ngủ là một trong những vấn đề quan trọng. Trong bối cảnh hiện đại, với sự phát triển của công nghệ và khoa học dữ liệu, việc dự đoán và cải thiện chất lượng giấc ngủ thông qua các mô hình máy học thật sự là một vấn đề đáng để nghiên cứu nhằm nâng cao sức khỏe con người.



## 2. Description

Đề tài này nhằm mục đích phát triển một mô hình dự đoán chất lượng giấc ngủ dựa trên phương pháp phân loại từ đó giúp người dùng có cái nhìn tốt hơn về tình trạng giấc ngủ hiện tại. Mô hình này sẽ sử dụng các thuật toán học máy để phân tích và dự đoán chất lượng giấc ngủ từ dữ liệu thu thập được. Các bước triển khai mô hình bao gồm thu thập dữ liệu, tiền xử lý dữ liệu, xây dựng mô hình, huấn luyện mô hình và đánh giá hiệu suất của mô hình.

### Input

- Giới tính, Tuổi, Chỉ số BMI.
- Thời gian làm việc, Chất lượng giấc ngủ, Mức độ vận động, Mức độ căng thẳng.
- Thói quen sử dụng thực phẩm có chất gây mất ngủ, Thói quen sử dụng thiết bị điện tử trước khi ngủ.
- Thời gian ngủ trong 1 đêm, Thời gian thường bắt đầu ngủ.
- Huyết áp, Nhịp tim.

- Nhiệt độ, Tiếng ồn.

## Output

- **7 nhãn:** ‘Bình thường’, ‘Mất ngủ nhẹ’, ‘Có nguy cơ rối loạn giấc ngủ’, ‘Rối loạn giấc ngủ’, ‘Hội chứng ngưng thở khi ngủ’, ‘Chứng ngủ rũ Narcolepsy’, ‘Giấc ngủ tốt’.

## Constraints

- **Giới tính:** ‘Nam’, ‘Nữ’.
- **Tuổi:** số nguyên dương bé hơn hoặc bằng 100.
- **Chỉ số BMI:** 1 số thực dương từ 10 tới 50, làm tròn 1 chữ số thập phân ( $\text{kg/m}^2$ ).
- **Thời gian làm việc:** ‘Ban ngày’, ‘Ban đêm’, ‘Không cố định’, ‘Không có’.
- **Chất lượng giấc ngủ, Mức độ vận động, Mức độ căng thẳng:** 1 thang đánh giá từ 1 đến 10 theo cảm nhận từ người dùng.
- **Thói quen sử dụng thực phẩm có chất gây mất ngủ, Thói quen sử dụng thiết bị điện tử trước khi ngủ:** ‘Không’, ‘Có’.
- Thời gian ngủ trong 1 đêm: số nguyên dương chỉ khoảng thời gian ngủ từ 2 đến 12 tiếng.
- **Thời gian thường bắt đầu ngủ (giờ):** 1 số nguyên dương từ 7 đến 12 nếu ngủ vào đêm hôm trước, từ 1 đến 6 nếu ngủ vào rạng sáng hôm sau.
- **Huyết áp tâm thu:** số nguyên dương từ 50 đến 200 (mmHg).
- **Nhịp Tim:** 1 số nguyên dương từ 30 đến 200 (bpm).
- **Nhiệt độ:** 1 số nguyên ( $^{\circ}\text{C}$ ).
- **Tiếng ồn:** ‘Không có’, ‘Nhỏ’, ‘To’.

## Requirements

- Độ chính xác của mô hình dự đoán **lớn hơn 85% trên tập test** thu thập từ 500 người.
- Thời gian sau khi đưa input vào mô hình đến khi đưa ra kết quả dự đoán phải **nhỏ hơn 2s trên cloud**, với tốc độ internet tối thiểu là 2MB/s.

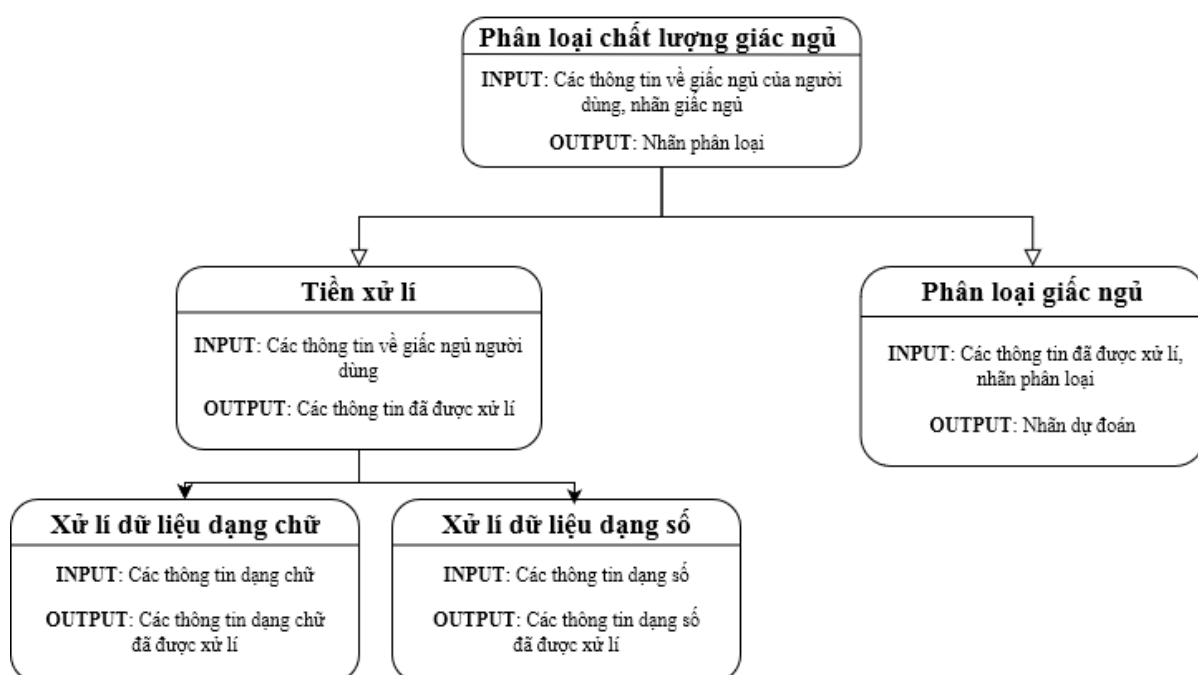
## II. Computational Thinking

### 1. Decomposition

Đối với những bài toán lớn, ta không thể giải ngay lập tức mà cần phải giải các toán nhỏ hơn để từ đó có được lời giải cho bài toán cuối cùng. Kỹ thuật đó được gọi là Decomposition.

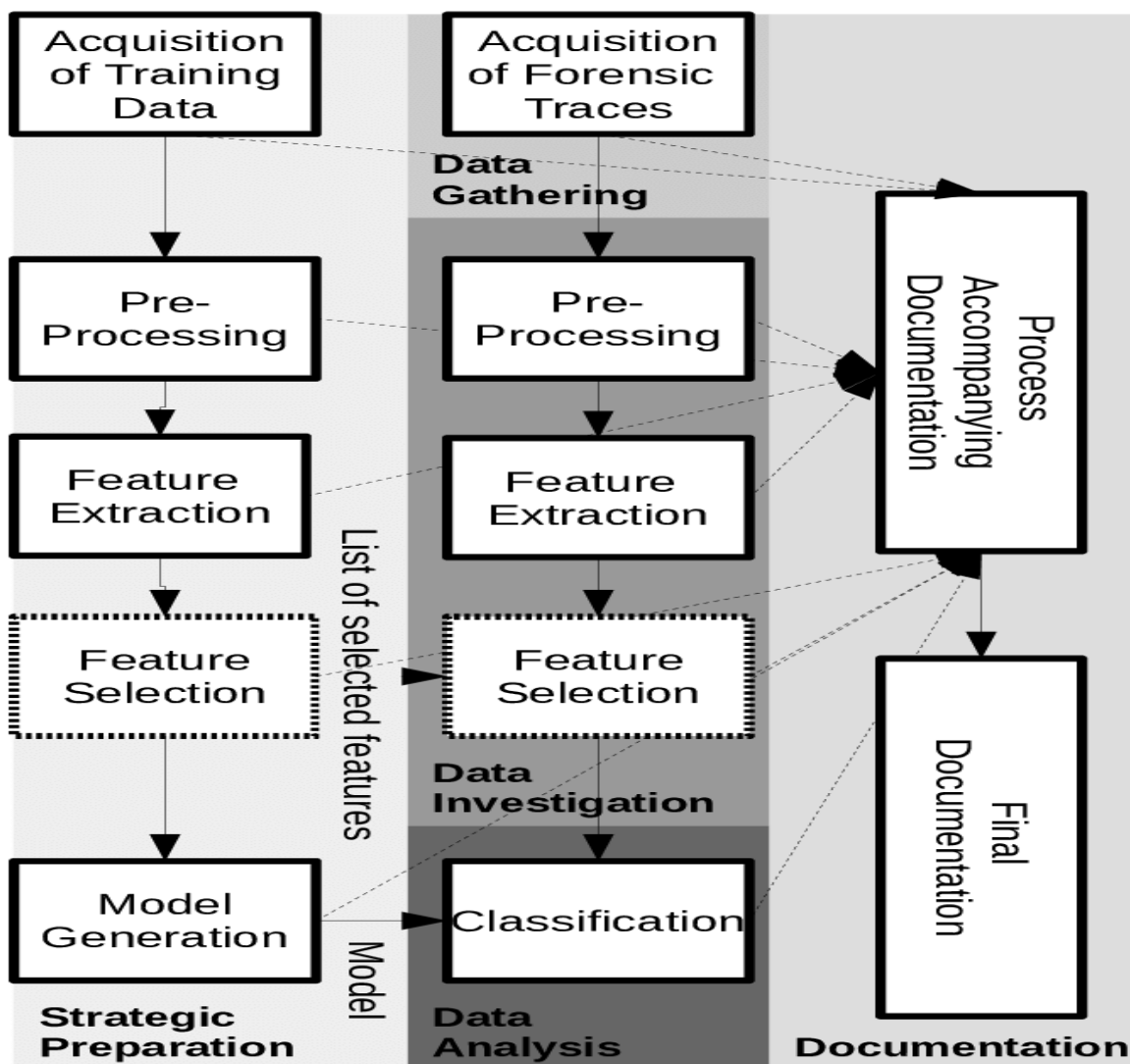
Đối với bài toán trên, chúng em áp dụng Decomposition để chia nhỏ thành các bài toán sau:

- **Phân loại chất lượng giấc ngủ**
  - **INPUT:** Các thông tin về giấc ngủ của người dùng, nhân giấc ngủ.
  - **OUTPUT:** Nhân phân loại giấc ngủ (VD: ‘Giấc ngủ tốt’, ‘Thiếu ngủ’...).
- **Tiền xử lý dữ liệu**
  - **Xử lý dữ liệu dạng chữ**
    - **INPUT:** Các thông tin dạng chữ (VD: Xử dụng chất kích thích gây mất ngủ nào,...).
    - **OUTPUT:** Các thông tin dạng chữ đã được xử lý.
  - **Xử lý dữ liệu dạng số**
    - **INPUT:** Các thông tin dạng số (VD: Nhiệt độ phòng (°C), cường độ ánh sáng (lux),...).
    - **OUTPUT:** Các thông tin dạng số đã được xử lý.
- **Phân loại giấc ngủ**
  - **INPUT:** Các thông tin đã được xử lý, nhân phân loại (Gồm thông tin dạng chữ và thông tin dạng số).
  - **OUTPUT:** Nhân giấc ngủ (VD: ‘Giấc ngủ tốt’, ‘Thiếu ngủ’...).



## 2. Pattern Recognition

### 2.1. Mapping



Để hiểu rõ và tối ưu hóa chất lượng giấc ngủ của người dùng, chúng ta cần xây dựng một mô hình ánh xạ hiệu quả giữa các đặc điểm cá nhân, chỉ số sức khỏe, thói quen sinh hoạt, môi trường sống và tình trạng giấc ngủ. Quá trình này, được gọi là 'mapping', sẽ giúp chúng ta phân tích và dự đoán chính xác tình trạng sức khỏe giấc ngủ dựa trên các dữ liệu đầu vào phong phú và đa dạng

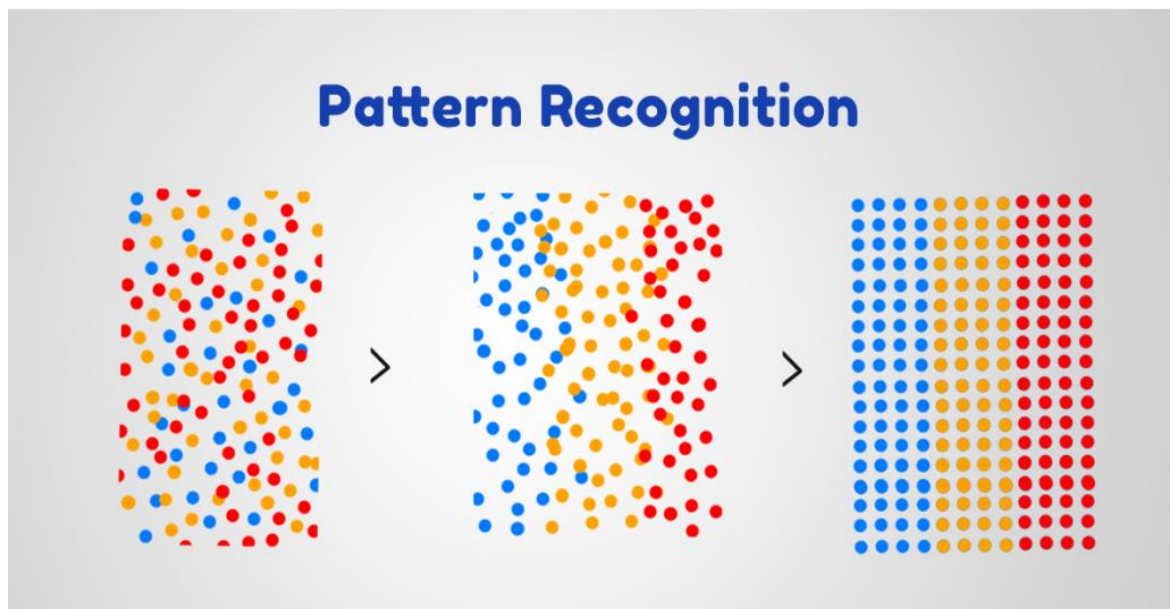
- **Tiền xử lý (Preprocessing):**

- **Làm sạch dữ liệu:** Loại bỏ các giá trị ngoại lệ, xử lý dữ liệu thiếu, và chuẩn hóa dữ liệu.
- **Chuyển đổi dữ liệu:** Mã hóa các biến phân loại (giới tính, thời gian làm việc, thói quen sử dụng thực phẩm, thói quen sử dụng thiết bị điện tử, tiếng ồn) thành các số liệu có thể xử lý bằng mô hình học máy.

- **Trích xuất đặc trưng (Feature Extraction):**

- **Đặc trưng thông tin cá nhân:** Giới tính, Tuổi.
- **Đặc trưng chỉ số sức khỏe:** Chỉ số BMI, Huyết áp tâm thu, Nhịp tim.
- **Đặc trưng thói quen sinh hoạt:** Thời gian làm việc, Mức độ vận động, Mức độ căng thẳng, Thói quen sử dụng thực phẩm, Thói quen sử dụng thiết bị điện tử.
- **Đặc trưng môi trường:** Nhiệt độ, Tiếng ồn.
- **Đặc trưng chỉ số giấc ngủ:** Thời gian ngủ, Thời gian bắt đầu ngủ, Chất lượng giấc ngủ.
- **Ánh xạ (Mapping):**
  - **Lựa chọn mô hình:** Các mô hình học máy phù hợp như mạng nơ-ron nhân tạo (ANN), máy vector hỗ trợ (SVM), rừng ngẫu nhiên (Random Forest) hoặc mô hình hồi quy logistic.
  - **Huấn luyện mô hình:** Sử dụng các đặc trưng đã trích xuất để huấn luyện mô hình dự đoán tình trạng giấc ngủ.
  - **Xác nhận và hiệu chỉnh mô hình:** Sử dụng dữ liệu kiểm tra để đánh giá độ chính xác và tinh chỉnh các tham số của mô hình.
- **Phân loại hoặc dự đoán (Classification or Prediction):**
  - **Áp dụng mô hình:** Dùng mô hình đã huấn luyện để dự đoán tình trạng sức khỏe giấc ngủ dựa trên dữ liệu mới.
  - **Phản hồi cho người dùng:** Cung cấp báo cáo về tình trạng giấc ngủ dựa trên kết quả dự đoán.

## 2.2. Grouping



- **Nhóm dữ liệu đầu vào:**
  - **Thông tin cá nhân**
    - **Giới tính:** Nam, Nữ

- **Tuổi:** số nguyên từ 1 đến 100
- **Chỉ số sức khỏe**
  - **Chỉ số BMI:** số thực dương từ 10 đến 50 (làm tròn 1 chữ số thập phân).
  - **Huyết áp tâm thu:** số nguyên dương từ 50 đến 200 (mmHg).
  - **Nhịp tim:** số nguyên dương (bpm).
- **Thói quen sinh hoạt**
  - **Thời gian làm việc:** ban ngày, ban đêm, không cố định, không có.
  - **Mức độ vận động:** mức đánh giá từ 1 đến 10.
  - **Mức độ căng thẳng:** mức đánh giá từ 1 đến 10.
  - **Thói quen sử dụng thực phẩm có chất gây mất ngủ:** Có, Không.
  - **Thói quen sử dụng thiết bị điện tử trước khi ngủ:** Có, Không.
- **Môi trường**
  - **Nhiệt độ:** số nguyên (độ C).
  - **Tiếng ồn:** không có, nhỏ, to.
- **Chỉ số giấc ngủ**
  - **Thời gian ngủ trong 1 đêm:** số nguyên dương từ 2 đến 12 tiếng.
  - **Thời gian thường bắt đầu để ngủ:** số nguyên dương từ 1 đến 12 (phụ thuộc vào thời điểm ngủ).
  - **Chất lượng giấc ngủ:** mức đánh giá từ 1 đến 10.
- **Nhóm dữ liệu đầu ra:**
  - Gồm *tình trạng sức khỏe giấc ngủ* có thể phân loại thành **7 nhãn**: ‘Bình thường’, ‘Mất ngủ nhẹ’, ‘Có nguy cơ rối loạn giấc ngủ’, ‘Rối loạn giấc ngủ’, ‘Hội chứng ngưng thở khi ngủ’, ‘Chứng ngủ rũ Narcolepsy’, ‘Giấc ngủ tốt’.

### Nhận xét:

Grouping như vậy được xác định dựa trên việc nhận diện các đặc tính quan trọng và có thể ảnh hưởng đến chất lượng giấc ngủ của một người. Quá trình nhận dạng và phân nhóm dữ liệu có thể diễn ra như sau:

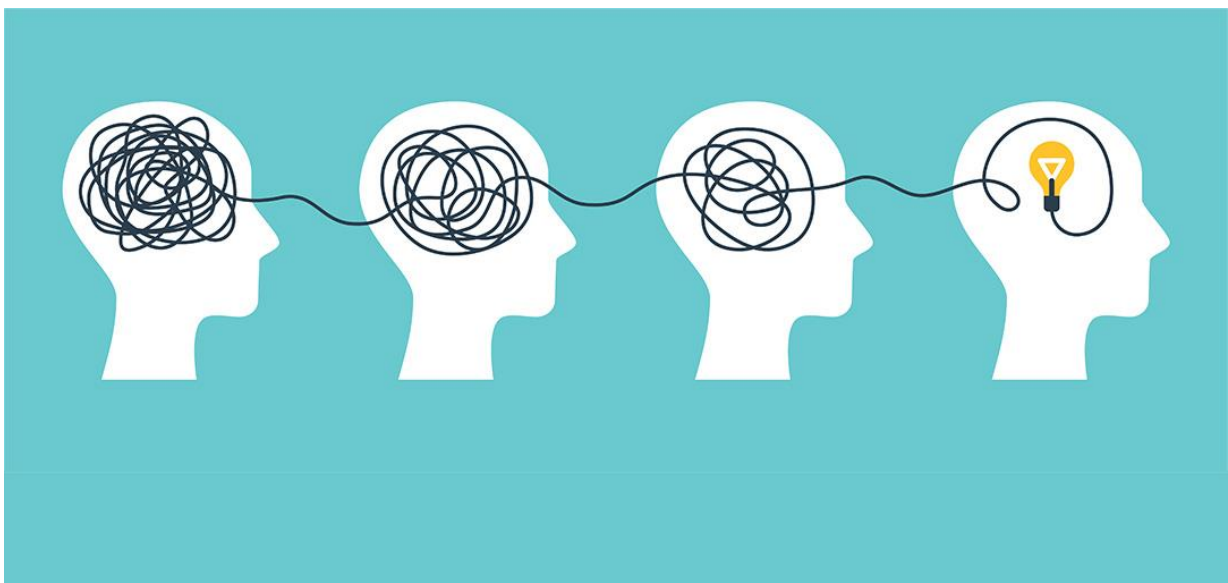
- **Nhận dạng các đặc tính quan trọng:**
  - **Thông tin cá nhân:** Giới tính và tuổi tác ảnh hưởng khá lớn tới giấc ngủ. Chẳng hạn như 1 người già sẽ giấc ngủ ngắn hơn so với 1 đứa trẻ hay đàn ông thường có nhiều áp lực công việc hơn so với phụ nữ, những công việc nặng nhọc, phải trực đêm ảnh hưởng tới giấc ngủ.
  - **Chỉ số sức khỏe:** Chỉ số BMI, huyết áp tâm thu, và nhịp tim là các chỉ số sức khỏe quan trọng có thể phản ánh trạng thái cơ thể và sức khỏe tổng thể của người dùng. Những người có chỉ số sức khỏe xấu như BMI cao, huyết áp cao, hay nhịp tim không ổn định sẽ dễ bị mất ngủ hơn so với người có chỉ số bình thường.



- **Thói quen sinh hoạt:** Thời gian làm việc, mức độ vận động, mức độ căng thẳng, thói quen sử dụng thực phẩm có chất gây mất ngủ, và thói quen sử dụng thiết bị điện tử trước khi ngủ có thể ảnh hưởng đến quá trình chuẩn bị và cơ chế giấc ngủ.
- **Môi trường:** Nhiệt độ và tiếng ồn trong môi trường ngủ có thể làm ảnh hưởng lớn đến chất lượng giấc ngủ. Việc nhiệt độ quá lạnh hoặc quá nóng, âm thanh xung quanh quá to có thể gây mất ngủ.
- **Chỉ số giấc ngủ:** Thời gian ngủ trong một đêm, thời gian thường bắt đầu để ngủ, và chất lượng giấc ngủ là các thông tin cụ thể về quá trình và chất lượng giấc ngủ của người dùng. Các chỉ số này sẽ đánh giá trực tiếp được tình trạng ngủ khi nó phản ánh rõ thời gian ngủ hiện tại cũng như cảm nhận của người dùng về sức khỏe giấc ngủ của bản thân.
- **Mục đích của việc phân nhóm:**
  - **Cải thiện hiệu suất dự đoán:** Việc nhóm các thông tin có liên quan lại với nhau giúp cho mô hình học máy có thể học được các mẫu phức tạp và đưa ra dự đoán chính xác hơn về tình trạng giấc ngủ của người dùng.
  - **Dễ dàng quản lý và hiệu quả:** Việc phân nhóm giúp cho việc quản lý và xử lý dữ liệu trở nên dễ dàng hơn, đồng thời cải thiện khả năng hiểu và diễn giải kết quả dự đoán từ mô hình học máy.
- **Tiêu chuẩn phân nhóm:**
  - Mỗi nhóm dữ liệu được xác định dựa trên tính liên quan chặt chẽ và tác động của các thông tin tới chất lượng giấc ngủ.
  - Mục đích là tạo ra các nhóm có tính đồng nhất về mặt chức năng, giúp cho mô hình học máy có thể học và dự đoán hiệu quả.

### 3. Abstraction / Generalization

#### 3.1. Abstraction



Bài toán dự đoán chất lượng giấc ngủ có thể rất phức tạp nếu ta xét chi tiết tới từng yếu tố ảnh hưởng đến chất lượng giấc ngủ vì vậy để đơn giản hoá bài toán và làm cho quá trình giải bài toán dễ dàng hơn ta sẽ lược bỏ bớt các chi tiết không cần thiết và tập trung vào các yếu tố chính, quan trọng:

Một số yếu tố chính ảnh hưởng đến giấc ngủ:

- Nhịp tim, huyết áp
- Mức độ vận động của cơ thể
- Mức độ căng thẳng
- Điều kiện môi trường xung quanh (nhiệt độ, tiếng ồn)

Một số ràng buộc cho các yếu tố trên:

- **Mức độ vận động, Mức độ căng thẳng:** 1 thang đánh giá từ 1 đến 10 theo cảm nhận từ người dùng.
- Huyết áp tâm thu: số nguyên dương từ 50 đến 200 (mmHg)
- Nhịp Tim là 1 số nguyên dương (bpm)
- Nhiệt độ là một số nguyên (độ C)
- Tiếng ồn: “Không”, “Nhỏ”, “To”

=> Từ các yếu tố trên ta tổng hợp một giải thuật để đưa ra dự đoán chất lượng giấc ngủ của người dùng ở tương lai

Với việc bỏ qua các chi tiết không quan trọng hay ảnh hưởng đáng kể tới giấc ngủ như các loại, mẫu mã hay màu sắc chăn, giường, gối, nệm khác nhau, hay việc nằm ở tư thế nào, giúp chúng ta có một bài toán bớt phức tạp và dễ giải quyết hơn nhiều so với bài toán gốc.

### 3.2. Generalization

Sau khi có xác định được cốt lõi của bài toán dự đoán chất lượng giấc ngủ của người dùng thông qua quá trình Abstraction, ta tiến hành mở rộng bài toán, rút ra những nguyên tắc quy luật chung từ mô hình trừu tượng, để dự đoán chất lượng giấc ngủ của nhiều người dùng khác nhau và các dữ liệu khác nhau:

- Tiến hành thu thập dữ liệu đa dạng: Thu thập dữ liệu từ nhiều độ tuổi, đa dạng về giới tính độ tuổi cũng như tình trạng sức khỏe, thói quen ngủ hoặc từ các nguồn, thiết bị khác nhau
- Trích xuất, chọn lọc đặc trưng hay thiết kế các đặc trưng mới để mô hình có thể học tốt hơn, tìm hiểu thêm được nhiều nguyên tắc để phù hợp, đưa ra dự đoán chính xác nhiều trường hợp dữ liệu đầu vào khác nhau
- Xây dựng huấn luyện mô hình và tiến hành thử nghiệm và đánh giá trên tập dữ liệu mới

- Liên tục điều chỉnh, cập nhật, tối ưu để mô hình có thể hoạt động tốt trên nhiều tập dữ liệu khác nhau.

Quá trình Generalization giúp cho mô hình nâng cao hiệu suất, dự đoán chính xác trên nhiều tập dữ liệu khác nhau, giúp ta có một mô hình có thể áp dụng được cho nhiều trường hợp khác nhau giúp chúng ta tiết kiệm thời gian mà không cần phải thiết kế lại.

#### 4. Evaluation

- Độ chính xác của mô hình dự đoán **lớn hơn 85% trên tập test** thu thập từ 500 người.

$$F1\ score = \frac{Precision * Recall}{Precision + Recall}$$

Trong đó:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

		Model dự đoán	
		Positive	Negative
Thực tế	Positive	True Positive (Dự đoán Đúng là Positive)	False Negative (Dự đoán Sai là Negative)
	Negative	False Positive (Dự đoán Sai là Positive)	True Negative (Dự đoán Đúng là Negative)

- Thời gian sau khi đưa input vào mô hình đến khi đưa ra kết quả dự đoán phải **nhỏ hơn 2s trên cloud**, với tốc độ internet tối thiểu là 2MB/s.

$$Time = t_{end} - t_{start}$$

Trong đó:

$t_{start}$ : thời gian sau khi nhập input.

$t_{end}$ : thời gian đưa ra kết quả dự đoán.

### III. Conclusion

Đề tài này đã đề xuất các phương pháp để **dự đoán chất lượng giấc ngủ**. Việc áp dụng những kết quả này sẽ một phần nào đó giúp người dùng có thể **đánh giá sớm** tình trạng giấc ngủ, cải thiện **sức khỏe tổng thể** và nâng cao **hiệu quả làm việc**, từ đó mang lại **hiệu quả kinh tế** cao cho người sử dụng.

Việc áp dụng **Computational Thinking** đóng vai trò quan trọng trong việc chia nhỏ vấn đề thành các phần dễ quản lý và thiết kế các thuật toán hiệu quả cho từng nhiệm vụ như phát hiện, phân loại, và theo dõi giấc ngủ. Điều này giúp xác định các thành phần cốt lõi và phát triển các giải pháp hệ thống một cách có tổ chức.

**Pattern Recognition** là yếu tố then chốt khi xử lý dữ liệu giấc ngủ. Bằng cách nhận diện các mẫu trong dữ liệu, hệ thống có thể cải thiện độ chính xác và tin cậy trong việc đánh giá giấc ngủ. Điều này rất cần thiết để phân biệt giữa các giai đoạn của giấc ngủ và đưa ra chính xác chất lượng giấc ngủ.

**Abstraction** cũng là một khái niệm quan trọng khi áp dụng các giải pháp phát hiện, phân loại, và theo dõi vào vấn đề quản lý giấc ngủ. Bằng cách trừu tượng hóa các chi tiết của thuật toán và tập trung vào ứng dụng thực tế, hệ thống có thể trở nên linh hoạt và mạnh mẽ hơn.

Tổng kết lại, sự tích hợp các kỹ thuật học sâu và tư duy tính toán không chỉ nâng cao độ chính xác và hiệu quả của hệ thống quản lý giấc ngủ mà còn góp phần cải thiện sức khỏe và chất lượng cuộc sống của người dùng.

## **IV. Demo**

### **\*Source code demo:**

<https://drive.google.com/drive/folders/1DYFYgrQT2916RW5WI16M4VeliYc7RhTX?usp=sharing>

### **\*Video demo:**

[https://drive.google.com/file/d/1kHk-ZLhrLcPu2JDwszwvLLCVru\\_5UygB/view?usp=drive\\_link](https://drive.google.com/file/d/1kHk-ZLhrLcPu2JDwszwvLLCVru_5UygB/view?usp=drive_link)

## V. References

- <https://tamanhhospital.vn/roi-loan-giac-ngu/>
- <https://codevidhya.com/computational-thinking/>
- [Mapping figure-computational-thinking](#)
- <https://www.learning.com/blog/decomposition-in-computational-thinking/>
- <https://ueh.edu.vn/cuoc-song-ueh/tin-tuc/polyphasic-sleep-phuong-phap-ngu-nhieu-giac-giup-tiet-kiem-thoi-gian-59326?app=true&app=true>