1. Lịch sử và Nguồn gốc

Tác giả: Practice Data Analysis With Me

Ngày tạo: 2/9/2024

2. Công dụng

Bộ dữ liệu được sử dụng phổ biến để:

Phân tích hiệu suất học tập: Tìm hiểu các yếu tố chính ảnh hưởng đến điểm số của học sinh.

Ứng dụng trong học máy (Machine Learning): Huấn luyện mô hình dự đoán điểm số hoặc phân loại học sinh dựa trên các yếu tố đầu vào.

Nghiên cứu giáo dục: Đưa ra các chính sách hoặc giải pháp cải thiện chất lượng giảng dạy và học tập.

3. Danh sách các đặc trưng

Student\_ID: mã số của học sinh(có thể:1,2,3,…)

Age: Tuổi của học sinh (các giá trị có thể: 18, 19, 20, 21, 22, 23, …)

Gender: Giới tính của học sinh (các giá trị có thể: Male, Female, Other)

University\_Year: Năm học của học sinh (các giá trị có thể: 1, 2, 3, 4)

Sleep\_Duration: Thời gian ngủ mỗi ngày (các giá trị có thể: 4-5, 6-7, 8-9, 10+ giờ)

Study\_Hours: Số giờ học mỗi ngày (các giá trị có thể: 1-2, 3-4, 5-6, 7+ giờ)

Screen\_Time: Thời gian sử dụng thiết bị điện tử mỗi ngày (các giá trị có thể: <2, 3-4, 5-6, >7 giờ)

Caffeine\_Intake: Lượng caffeine tiêu thụ mỗi ngày (các giá trị có thể: 0-1, 2-3, 4-5, 6+ tách cà phê)

Physical\_Activity: Mức độ hoạt động thể chất của học sinh (các giá trị có thể: 1,2,3,…)

Sleep\_Quality: Chất lượng giấc ngủ của học sinh (các giá trị có thể:1,2,3…,10)

Weekday\_Sleep\_Start: Thời gian bắt đầu ngủ vào các ngày trong tuần (các giá trị có thể: 0-1,1-2,20-21,21+…)

Weekend\_Sleep\_Start: Thời gian bắt đầu ngủ vào cuối tuần (các giá trị có thể: : 0-1,1-2,20-21,21+ …)

Weekday\_Sleep\_End: Thời gian thức dậy vào các ngày trong tuần (các giá trị có thể: : 0-1,1-2,8-9,9+ …)

Weekend\_Sleep\_End: Thời gian thức dậy vào cuối tuần (các giá trị có thể: : 0-1,1-2,9-10,10+…)

4. Kiểu dữ liệu

Kiểu dữ liệu phân loại (Categorical): Các giá trị của những đặc trưng như Gender, Sleep\_Duration,Study\_Hours,Screen\_Time,Caffeine\_Intake,Sleep\_Quality,Weekday\_Sleep\_Start,Weekend\_Sleep\_Start, Weekday\_Sleep\_End,và Weekend\_Sleep\_End đều là dữ liệu phân loại, phản ánh các mức độ, khoảng thời gian, hoặc đặc điểm hành vi có thể phân chia thành các nhóm.

Kiểu dữ liệu số nguyên (Integer): Các đặc trưng như Student\_ID, Age, University\_Year, Physical\_Activity là kiểu số nguyên, mô tả các giá trị có thể đo lường hoặc đếm được.

5. Mục đích

Mục đích của bộ dữ liệu là hỗ trợ phân tích mối quan hệ giữa các yếu tố như giấc ngủ, thời gian học và kết quả học tập của học sinh, từ đó phát triển các hệ thống giáo dục thông minh, xây dựng mô hình dự đoán hiệu quả học tập và tối ưu hóa môi trường học tập để cải thiện chất lượng giáo dục.

6.Lĩnh vực ứng dụng

Giáo dục: Đánh giá và cải thiện chất lượng học tập.

Khoa học dữ liệu: Nghiên cứu các thuật toán học máy trong lĩnh vực giáo dục.

Chính sách công: Phân tích tác động của môi trường xã hội và gia đình đối với học sinh.

7. Hạn chế

\* Hạn chế:

Thiếu dữ liệu đầy đủ: Bộ dữ liệu có thể thiếu thông tin cho một số học sinh, làm giảm tính đầy đủ và khả năng phân tích.

Sự chủ quan trong việc tự báo cáo: Các đặc trưng như giấc ngủ, giờ học, và hoạt động thể chất dựa vào tự báo cáo của học sinh, có thể dẫn đến sự thiên lệch do không chính xác.

Không có thông tin về thành tích học tập: Bộ dữ liệu thiếu thông tin về điểm số, ảnh hưởng đến khả năng phân tích tác động của các yếu tố đến kết quả học tập.

Thiếu thông tin về yếu tố môi trường: Bộ dữ liệu không cung cấp thông tin về các yếu tố như căng thẳng, điều kiện sống, hay hỗ trợ gia đình, ảnh hưởng đến thói quen học tập và giấc ngủ.

Chưa chuẩn hóa dữ liệu: Các đơn vị đo lường chưa được chuẩn hóa, gây khó khăn trong việc xử lý và phân tích.

8. Khiếm khuyết

\*Khiếm khuyết:

Giới hạn mẫu học sinh: Bộ dữ liệu chỉ có thể đại diện cho một nhóm học sinh cụ thể, làm giảm tính khái quát và khả năng áp dụng cho toàn bộ học sinh.

Thiếu thông tin về các yếu tố khác: Thiếu các yếu tố như dinh dưỡng, sức khỏe, và mức độ xã hội hóa, cần thiết để đánh giá đầy đủ ảnh hưởng đến thói quen học tập và giấc ngủ.

Không có đánh giá chất lượng dữ liệu: Bộ dữ liệu thiếu thông tin về nguồn gốc và quy trình thu thập, làm khó khăn trong việc đánh giá độ tin cậy.

Không có thông tin về tác động dài hạn: Dữ liệu chỉ phản ánh thói quen học tập và sức khỏe tại một thời điểm, thiếu thông tin về tác động dài hạn của các yếu tố.

Khả năng phân tích có hạn: Thiếu thông tin về thành tích học tập làm hạn chế khả năng phân tích mối quan hệ giữa các yếu tố hành vi và kết quả học tập.