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

Hours\_Studied: Số giờ học mỗi tuần (các giá trị có thể: số nguyên dương).

Attendance: Tỷ lệ tham gia lớp học (các giá trị có thể: từ 0% đến 100%).

Parental\_Involvement: Mức độ tham gia của phụ huynh vào việc học của học sinh (các giá trị: Low, Medium, High).

Access\_to\_Resources: Sự sẵn có của các tài nguyên giáo dục (các giá trị: Low, Medium, High).

Extracurricular\_Activities: Tham gia các hoạt động ngoại khóa (các giá trị: Yes, No).

Sleep\_Hours: Số giờ ngủ trung bình mỗi đêm (các giá trị: số nguyên dương).

Previous\_Scores: Điểm số từ các kỳ thi trước (các giá trị: số thực từ 0 đến 100).

Motivation\_Level: Mức độ động lực học tập của học sinh (các giá trị: Low, Medium, High).

Internet\_Access: Sự sẵn có của kết nối internet (các giá trị: Yes, No).

Tutoring\_Sessions: Số buổi học thêm mỗi tháng (các giá trị: số nguyên không âm).

Family\_Income: Mức thu nhập gia đình (các giá trị: Low, Medium, High).

Teacher\_Quality: Chất lượng của giáo viên (các giá trị: Low, Medium, High).

School\_Type: Loại trường học mà học sinh theo học (các giá trị: Public, Private).

Peer\_Influence: Ảnh hưởng của bạn bè đến hiệu suất học tập (các giá trị: Positive, Neutral, Negative).

Physical\_Activity: Số giờ hoạt động thể chất trung bình mỗi tuần (các giá trị: số nguyên không âm).

Learning\_Disabilities: Sự hiện diện của các khó khăn học tập (các giá trị: Yes, No).

Parental\_Education\_Level: Trình độ học vấn cao nhất của phụ huynh (các giá trị: High School, College, Postgraduate).

Distance\_from\_Home: Khoảng cách từ nhà đến trường (các giá trị: Near, Moderate, Far).

Gender: Giới tính của học sinh (các giá trị: Male, Female).

Exam\_Score: Điểm thi cuối kỳ (các giá trị: số thực từ 0 đến 100).

4. Kiểu dữ liệu

Số nguyên (Integer):Hours\_Studied, Sleep\_Hours, Tutoring\_Sessions, Physical\_Activity: Đại diện cho các số lượng hoặc thời gian cụ thể.

Số thực (Float):Attendance, Previous\_Scores, Exam\_Score: Biểu thị các giá trị phần trăm hoặc điểm số.

Chuỗi(String):Parental\_Involvement,Access\_to\_Resources,Motivation\_Level,Family\_Income,Teacher\_Quality,School\_Type,Peer\_Influence,Parental\_Education\_Level,Distance\_from\_Home, Gender: Các giá trị mô tả mức độ, loại hình, hoặc thuộc tính.

Boolean (Đúng/Sai):Extracurricular\_Activities, Internet\_Access, Learning\_Disabilities: Các giá trị dạng Yes/No.

5. Mục tiêu:

Bộ dữ liệu này được thiết kế để hỗ trợ trong lĩnh vực giáo dục bằng cách:

Phân tích hiệu suất học tập: Xác định mối quan hệ giữa các yếu tố như thời gian học, giờ ngủ, hoặc mức độ tham gia hoạt động ngoại khóa với điểm thi.

Phát triển mô hình dự đoán: Xây dựng các mô hình để dự đoán thành tích học tập của học sinh dựa trên các yếu tố ảnh hưởng.

Cải thiện hệ thống giáo dục: Đưa ra các gợi ý cải thiện về chất lượng giáo dục, phương pháp dạy học, và hỗ trợ cá nhân hóa cho học sinh.

Nghiên cứu học thuật: Hỗ trợ nghiên cứu về hành vi và thói quen học tập của học sinh trong các bối cảnh khác nhau.

6. Lĩnh vực:

Giáo dục: Tối ưu hóa phương pháp giảng dạy và hỗ trợ học sinh.

Khoa học dữ liệu: Ứng dụng trong phân tích và xây dựng các mô hình học máy.

Tâm lý học giáo dục: Hiểu rõ hơn các yếu tố ảnh hưởng đến động lực học tập và kết quả học tập.

Xã hội học: Phân tích mối liên hệ giữa điều kiện kinh tế xã hội và hiệu suất học tập.

7. Hạn chế:

Dữ liệu không đầy đủ:

Thiếu thông tin quan trọng về các yếu tố như sức khỏe tinh thần, căng thẳng, hoặc cảm xúc học tập.

Không có dữ liệu theo dõi dài hạn về ảnh hưởng của các yếu tố lên hiệu suất học tập.

Chủ quan trong tự báo cáo:

Nhiều dữ liệu như số giờ học, động lực, hoặc mức độ tham gia hoạt động ngoại khóa dựa trên tự đánh giá, dễ dẫn đến sai lệch.

Thiếu sự đại diện:

Bộ dữ liệu có thể chỉ đại diện cho một nhóm học sinh cụ thể, làm giảm tính khái quát khi áp dụng vào các nhóm khác.

Không chuẩn hóa:

Các giá trị không thống nhất về đơn vị đo lường hoặc phạm vi (ví dụ: cách đo thời gian hoặc mức độ ảnh hưởng).

8. Khiếm khuyết

Hạn chế mẫu học sinh:

Chỉ bao gồm một nhóm đối tượng cụ thể, như trường công hoặc tư, hoặc khu vực nhất định, làm giảm tính khái quát.

Thiếu các yếu tố môi trường và xã hội:

Không có thông tin về sự hỗ trợ từ gia đình, điều kiện sống, hoặc áp lực xã hội ảnh hưởng đến học sinh.

Không đánh giá chất lượng dữ liệu:

Bộ dữ liệu thiếu các thông tin về quy trình thu thập, làm khó khăn trong việc đánh giá độ chính xác và độ tin cậy.

Giới hạn về thời gian:

Dữ liệu chỉ ghi nhận tại một thời điểm cố định, không phản ánh xu hướng hoặc thay đổi dài hạn.