

Mental Health Dataset¹

Giới thiệu về tập dữ liệu

Bộ dữ liệu này cung cấp thông tin chi tiết toàn diện về sức khỏe tâm thần. Tập dữ liệu cung cấp một bộ sưu tập dữ liệu sức khỏe tâm thần ẩn danh của 1000 cá nhân, đại diện cho nhiều lứa tuổi, giới tính, nghề nghiệp và quốc gia, nhằm mục đích làm sáng tỏ các yếu tố khác nhau ảnh hưởng đến sức khỏe tâm thần, cung cấp những hiểu biết có giá trị về mức độ căng thẳng, kiểu ngủ, cân bằng giữa công việc và cuộc sống và hoạt động thể chất.

Tập dữ liệu này có thể được sử dụng cho các mô hình nghiên cứu, phân tích hoặc học máy để dự đoán xu hướng sức khỏe tâm thần, khám phá mối tương quan giữa cân bằng giữa công việc và cuộc sống và sức khỏe tinh thần, đồng thời khám phá tác động của căng thẳng và hoạt động thể chất đối với sức khỏe tâm thần.

Các tính năng chính của tập dữ liệu

1. **Demographics** (Nhân khẩu học): Tập dữ liệu bao gồm các cá nhân từ nhiều quốc gia khác nhau như Hoa Kỳ, Ấn Độ, Anh, Canada và Úc. Mỗi mục nhập nắm bắt thông tin nhân khẩu học quan trọng như tuổi tác, giới tính và nghề nghiệp (ví dụ: CNTT, Chăm sóc sức khỏe, Giáo dục, Kỹ thuật).
2. **Mental Health Conditions** (Tình trạng sức khỏe tâm thần): Tập dữ liệu chứa dữ liệu về việc các cá nhân có báo cáo bất kỳ vấn đề sức khỏe tâm thần nào hay không (Có/Không), cùng với mức độ nghiêm trọng của các tình trạng này được phân loại thành Low, Medium, High (Thấp, Trung bình hoặc Cao).
 - (i).- **User ID** (ID người dùng, kiểu integer): Mã định danh duy nhất cho mỗi người dùng.
 - (ii).- **Age** (Tuổi, kiểu integer): Tuổi của người dùng
 - (iii).- **Sex** (Giới tính, là thuộc tính dạng danh nghĩa – categorical): cung cấp nhận dạng giới tính của người dùng gồm Male (nam), Female (nữ), Non-binary (phi nhị giới²), Prefer not to say (Không muốn nói).
 - (iv).- **Occupation category** (Danh mục nghề nghiệp, là thuộc tính dạng danh nghĩa – categorical): gồm IT, Finance, Healthcare, Education, Engineering, Sales, Other (CNTT, Tài chính, Y tế, Giáo dục, Kỹ thuật, Bán hàng, Khác)
 - (v).- **Country of residence** (Quốc gia cư trú, là thuộc tính dạng danh nghĩa – categorical): gồm USA, India, UK, Canada, Australia, Germany, Other (Hoa Kỳ, Ấn Độ, Anh, Canada, Úc, Đức, Khác).
3. **Consultation History** (Lịch sử tư vấn): Đối với những cá nhân có tình trạng sức khỏe tâm thần, tập dữ liệu ghi chú xem họ đã tham khảo ý kiến của chuyên gia sức khỏe tâm thần hay chưa.
 - (vi).- **Mental Health Condition** (là thuộc tính dạng boolean): cho biết người dùng có báo cáo vấn đề sức khỏe tâm thần hay không (Yes/No).

¹ <https://www.kaggle.com/datasets/bhadramohit/mental-health-dataset?resource=download>

² https://vi.wikipedia.org/wiki/Phi_nh%E1%BB%8B_nguy%C3%AAn_gi%E1%BB%9Bi

- (vii).- **Severity** (là thuộc tính dạng danh nghĩa – categorical): cho biết mức độ nghiêm trọng của tình trạng sức khỏe tâm thần, gồm: Low, Medium, High, None (Thấp, Trung bình, Cao, Không có).
- (viii).- **Consultation_History** (Lịch sử Tư vấn, là thuộc tính dạng boolean): cho biết người dùng đã tham khảo ý kiến chuyên gia sức khỏe tâm thần hay chưa (Yes/No).
4. **Stress Levels** (Mức độ căng thẳng): Mức độ căng thẳng của mỗi cá nhân được phân loại là Low, Medium, High (Thấp, Trung bình hoặc Cao), cung cấp thông tin chi tiết về các yếu tố khác nhau như giờ làm việc hoặc giấc ngủ có thể tương quan như thế nào với sức khỏe tinh thần.
- (ix).- **Stress_Level** (là thuộc tính dạng danh nghĩa – categorical): cho biết mức độ căng thẳng được báo cáo , gồm: Low, Medium, High (Thấp, Trung bình, Cao)
5. **Lifestyle Factors** (Các yếu tố về lối sống): Tập dữ liệu bao gồm thông tin về thời lượng ngủ, số giờ làm việc mỗi tuần và số giờ hoạt động thể chất hàng tuần, cung cấp một bức tranh chi tiết về các yếu tố lối sống góp phần như thế nào đối với sức khỏe tâm thần.
- (x).- **Sleep_Hours** (kiểu float): cho biết số giờ ngủ trung bình mỗi đêm.
- (xi).- **Work_Hours** (kiểu integer): cho biết số giờ làm việc/tuần.
- (xii).- **Physical_Activity_Hours** (kiểu integer): cho biết số giờ hoạt động thể chất/tuần.