

# MENTAL HEALTH DETECTION DATASET

Nomes: Augusto Hansen Mentz e Joice Aline Colling

Disciplina: Inteligência Artificial e Computacional

Data: 01/07/2024

## 1. APRESENTAÇÃO DO DATASET

O conjunto de dados selecionado contém respostas de uma pesquisa com pacientes sobre diversos sintomas associados à depressão, como sono, apetite, interesse em atividades físicas, entre outros. Cada paciente respondeu a 14 perguntas, e as respostas são codificadas de 1 a 6, de acordo com a frequência com que experimentaram cada sintoma.

O dataset possui 16 colunas: 14 com dados das entrevistas, uma coluna com o identificador do entrevistado e uma coluna com o resultado. No total, há 813 registros, mas após a remoção dos dados em branco, restam 491 registros. A coluna que vamos considerar como nosso atributo classe é a coluna que indica o estado geral de depressão do paciente. O dataset pode ser acessado através deste [link](#).

## 2. CARACTERIZAÇÃO DOS DADOS

Nos próximos tópicos, abordaremos a caracterização dos dados presentes no dataset escolhido.

### 2.1 DADOS

A tabela abaixo apresenta as 16 colunas do dataset, o significado e o tipo de cada uma.

Coluna	Significado	Tipo
Number	Identificador do paciente	Float64
Sleep	Frequência de distúrbios do sono	Float64
Appetite	Alterações no apetite	Float64
Interest	Perda de interesse em atividades	Float64
Fatigue	Sensação de fadiga ou baixa energia	Float64
Worthlessness	Sentimentos de inutilidade ou culpa excessiva	Float64
Concentration	Dificuldade de concentração	Float64
Agitation	Agitação física	Float64
Suicidal Ideation	Pensamentos de auto-lesão ou suicídio	Float64
Sleep Disturbance	Problemas para dormir	Float64
Aggression	Sentimento de agressão	Float64
Panic Attacks	Experiência de ataques de pânico	Float64
Hopelessness	Sentimentos de desesperança	Float64
Restlessness	Inquietude	Float64
Low Energy	Baixa Energia	Float64
Depression State	Estágio de depressão	Object

## 2.2 CLASSIFICAÇÃO DOS DADOS

As respostas da pesquisa foram classificadas de acordo com a frequência com que os pacientes sentiram cada sintoma. A tabela a seguir mostra o significado de cada valor.

1	Never (Nunca)
2	Always (Sempre)
3	Often (Frequentemente)
4	Rarely (Raramente)
5	Sometimes (Às vezes)
6	Not at all (De forma alguma)

## 2.3 ATRIBUTO CLASSE

O atributo classe do dataset é a coluna **Depression State**, no qual pode ser classificada por 4 categorias: “No depression”, “Mild”, “Moderate” e “Severe”. A variável com maior desbalanceamento é a “No depression” com 157 registros. Na tabela abaixo, apresentamos as categorias e sua respectiva quantidade de registros no dataset.

Categoria	Quantidade de Registros
No depression (Sem depressão)	157
Mild (Leve)	117
Moderate (Moderado)	109
Severe (Severo)	108

## 3. PROCEDIMENTOS DE PRÉ-PROCESSAMENTO

Os três algoritmos escolhidos foram:

- Árvore de Decisão (DecisionTreeClassifier)
- Floresta Aleatória (RandomForestClassifier)
- KNN (KNeighborsClassifier)

Para obtermos os melhores resultados, foram necessárias algumas adequações:

- a. Remover os registros que não possuem dados.
- b. Transformar o atributo classe para variável categórica.
- c. Eliminar a coluna Number, pois ela não contém dados úteis.

Após as adequações, foram obtidos os resultados e mostramos na tabela abaixo:

Algoritmo	Acerto	Precisão	Recall	F-Measure	Tempo de processamento
DecisionTreeClassifier	0,318	0,872	0,304	0,434	0.30 segundos
RandomForestClassifier	0,257	0,835	0,258	0,383	0.42 segundos
KNeighborsClassifier	0,338	0,674	0,363	0,396	0.10 segundos

#### 4. CONCLUSÕES

**Acurácia:** Os algoritmos DecisionTreeClassifier e KNN apresentaram acurácias próximas, sendo os mais elevados entre os três algoritmos testados. Isso indica que eles tiveram a maior taxa de acertos na classificação de casos positivos e negativos de depressão, evidenciando sua eficácia em fazer previsões precisas.

**Precisão:** O algoritmo DecisionTreeClassifier se destacou com uma precisão superior em relação aos outros algoritmos. Isso significa que, das instâncias previstas como positivas para depressão, uma maior proporção realmente correspondia a casos verdadeiros de depressão, resultando em menos falsos positivos.

**Recall:** O recall, ou sensibilidade, do algoritmo KNN foi o mais alto comparado aos demais algoritmos. Isso demonstra que ele foi capaz de identificar uma maior proporção de casos reais de depressão entre todos os casos positivos existentes no dataset.

**F-Measure:** O F-Measure, uma métrica que considera tanto a precisão quanto o recall do modelo, foi mais elevado para o algoritmo DecisionTreeClassifier. Isso indica um equilíbrio entre precisão e recall, o que é crucial para a qualidade geral do modelo.

Com base nessas métricas e na análise detalhada dos resultados, destacamos os algoritmos DecisionTreeClassifier e KNN. Contudo, o algoritmo DecisionTreeClassifier possuiu um tempo de treinamento menor. Concluímos que o algoritmo DecisionTreeClassifier foi o mais eficaz na previsão da depressão, sendo que deste algoritmo, o atributo classe que obteve o melhor resultado foi “No depression”, com 0.880 de precisão e o pior resultado foi “Moderate”, com 0.857 de precisão.