

**CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA**  
**Faculdade de Tecnologia do Estado de São Paulo – FATEC Deputado Ary**  
**Fossen**  
Curso Superior de Tecnologia em Ciência de Dados

## **CIÊNCIA DE DADOS APLICADA À NOSOLOGIA PSIQUIÁTRICA**

ARIEL LADISLAU REISES

JUNDIAÍ  
2024

## RESUMO

Neste artigo, exploramos como a ciência de dados pode ser utilizada na nosologia psiquiátrica, com base nas técnicas de agrupamento de clusters. Três clusters principais de transtornos mentais foram identificados com base em dados clínicos e biomarcadores, destacando padrões ocultos e sobreposições diagnósticas não capturadas pelas classificações sintomáticas convencionais como o DSM-5. Uma abordagem dimensional e não categórica implementada por ciência de dados pode garantir uma maior precisão diagnóstica e personalização terapêutica. Porém, os resultados devem ser interpretados com cautela e, futuramente, estudos com amostras maiores e mais diversas, bem como dados genéticos e ambientais, são necessários para modelar sistemas mais robustos e dinâmicos de nosologia psiquiátrica.

## PALAVRAS-CHAVE

*Ciência de dados; nosologia psiquiátrica; agrupamento de clusters; DSM-5; biomarcadores; precisão diagnóstica; personalização de tratamento.*

## ABSTRACT

*In this article, we explore how data science can be used in psychiatric nosology, based on clustering techniques. Three main clusters of mental disorders were identified based on clinical data and biomarkers, highlighting hidden patterns and diagnostic overlaps not captured by conventional symptomatic classifications like the DSM-5. A dimensional and non-categorical approach implemented through data science can ensure greater diagnostic accuracy and therapeutic personalization. However, results should be interpreted with caution, and future studies with larger and more diverse samples, as well as genetic and environmental data, are necessary to model more robust and dynamic systems of psychiatric nosology.*

## KEYWORDS

*Data science; psychiatric nosology; clustering; DSM-5; biomarkers; diagnostic accuracy; personalized treatment.*

## INTRODUÇÃO

Nos últimos anos, a união entre a ciência de dados e a saúde mental tem trazido novas esperanças para melhorar a precisão dos diagnósticos e criar tratamentos mais personalizados. A psiquiatria, com sua complexidade e desafios, ainda encontra dificuldades na forma de categorizar e tratar os transtornos mentais, devido à grande variação nos sintomas e à sobreposição entre diferentes diagnósticos. Até hoje, a nosologia psiquiátrica, que é área que estuda a classificação das doenças mentais, tem se baseado em sistemas tradicionais, como o Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5) e a Classificação Internacional de Doenças (CID-11), que utilizam critérios clínicos e entrevistas para definir os diagnósticos. Porém, esses métodos muitas vezes falham em distinguir claramente transtornos parecidos, sendo considerados subjetivos e até arbitrários.

Com essa realidade em mente, surge uma necessidade urgente por abordagens mais objetivas e confiáveis na classificação das doenças mentais. É aí que entram as ferramentas de ciência de dados, como os algoritmos de agrupamento de clusters. Esses algoritmos, projetados para identificar padrões em grandes volumes de dados, podem trazer uma nova perspectiva. Com o agrupamento de clusters, é possível identificar grupos de transtornos que compartilham características semelhantes, permitindo uma reclassificação dos diagnósticos psiquiátricos com base em evidências e não apenas em critérios subjetivos.

Este artigo explora como a ciência de dados, em especial o uso de técnicas de agrupamento de clusters, pode ajudar a reclassificar a forma como entendemos os transtornos mentais. A principal hipótese é que essas ferramentas tecnológicas podem refinar os sistemas diagnósticos atuais, trazendo mais precisão e oferecendo uma base mais sólida para identificar e tratar diferentes condições psiquiátricas. O uso de grandes volumes de dados, aliado a algoritmos de aprendizado de máquina, tem o potencial não só de melhorar o diagnóstico, mas também de personalizar os tratamentos, aumentando a eficácia das terapias e, consequentemente, os resultados para os pacientes.

## **CIÊNCIA DE DADOS NA SAÚDE**

O uso de ciência de dados na área da saúde tem crescido exponencialmente nas últimas duas décadas, impulsionado pelo aumento na disponibilidade de dados biomédicos e pelos avanços nas capacidades computacionais. Inicialmente, o foco da aplicação de big data na saúde estava em áreas como o diagnóstico de câncer e doenças cardiovasculares, onde grandes volumes de dados podiam ser coletados e analisados para identificar os padrões. Mas, mais recentemente, a psiquiatria tem se beneficiado de tais abordagens. Weissman (2024) destaca que o uso de big data na psiquiatria está em sua infância, mas já tem mostrado potencial para revolucionar o campo, especialmente na forma como doenças mentais são diagnosticadas e tratadas.

### **Desafios da Nosologia Psiquiátrica Tradicional**

Historicamente, a nosologia psiquiátrica dependeu de modelos categóricos, como o DSM e a CID, que organizam os transtornos mentais em categorias distintas. Esses modelos, apesar de amplamente utilizados, têm limitações notáveis. Estudos demonstram que esses sistemas frequentemente falham em capturar a variabilidade individual e a comorbidade entre transtornos. Segundo Aquino e Ross (2017), a dependência exclusiva de sintomas clínicos tem levado à criação de categorias diagnósticas com fronteiras nebulosas e sobreposição de sintomas. Por exemplo, a depressão e a ansiedade frequentemente compartilham sintomas, tornando difícil a distinção clara entre esses transtornos com base nos critérios atuais. Outro desafio é a dependência da perspectiva sintomática. Muitas doenças mentais são classificadas com base em relatos subjetivos de sintomas, o que pode levar a variações substanciais no diagnóstico dependendo do clínico e do paciente. Esse tipo de subjetividade pode gerar inconsistências nos tratamentos recomendados e nos prognósticos dos pacientes.

### **Agrupamento de Clusters em Psiquiatria: Estudos Relevantes**

O agrupamento de clusters é uma técnica que se destaca pela sua capacidade de identificar padrões em grandes conjuntos de dados, agrupando informações que compartilham características semelhantes. Essa técnica é particularmente

relevante para a psiquiatria, área na qual os transtornos mentais muitas vezes compartilham sintomas e podem ser difíceis de distinguir. Estudos recentes, como o de Huston (2020), têm aplicado o agrupamento de clusters para mapear padrões de doenças mentais de forma mais precisa do que as classificações tradicionais. O estudo de Huston sobre o modelo HiTOP propõe uma abordagem dimensional para a nosologia psiquiátrica, agrupando transtornos mentais com base em sintomas compartilhados e fatores subjacentes, em vez de categorias rígidas. O uso de big data permite que psiquiatras e pesquisadores identifiquem grupos de sintomas que talvez não fossem visíveis a partir de uma análise tradicional. Isso facilita o desenvolvimento de novos diagnósticos baseados em dados reais, com o potencial de melhorar a precisão do diagnóstico. Além disso, estudos como o de Fávero e Belfiore (2017) destacam a aplicação de técnicas de agrupamento em grandes bases de dados como uma forma de identificar subgrupos de pacientes com respostas terapêuticas semelhantes.

### **HiTOP e Big Data na Reclassificação Psiquiátrica**

O HiTOP (*Hierarchical Taxonomy of Psychopathology*) é um novo modelo de classificação de transtornos mentais que utiliza dados para agrupar sintomas em dimensões hierárquicas. Esse modelo desafia a abordagem categórica tradicional e propõe uma nova perspectiva para entender os transtornos mentais como parte de um espectro contínuo, em vez de categorias isoladas. O HiTOP se baseia em técnicas de ciência de dados para identificar padrões subjacentes entre os sintomas, criando grupos mais homogêneos e clinicamente relevantes. Segundo Weissman (2024), o uso do Big Data na psiquiatria está começando a evidenciar como a ciência de dados pode transformar a compreensão e o tratamento de doenças mentais.

Apesar dos avanços, ainda há uma lacuna significativa na maneira como os dados são utilizados para classificar doenças mentais. Embora estudos promissores tenham demonstrado o potencial das técnicas de ciência de dados, ainda não há consenso sobre como integrar essas descobertas aos sistemas de diagnóstico tradicionais. Além disso, há desafios relacionados à escassez de dados robustos

em áreas específicas, como o comportamento e a biologia dos transtornos mentais.

## **METODOLOGIA**

Este estudo adotou um design quantitativo com uma abordagem exploratória e descritiva, focando na aplicação de algoritmos de agrupamento de clusters para a reclassificação de transtornos mentais. O objetivo foi identificar padrões subjacentes aos dados clínicos e propor novas classificações baseadas em características comuns entre os transtornos, utilizando biomarcadores e perfis sintomáticos. O agrupamento de clusters foi escolhido como método central pela sua capacidade de lidar com grandes volumes de dados e identificar agrupamentos de casos semelhantes. A hipótese principal investigada é que o uso de técnicas de ciência de dados, como o agrupamento, pode oferecer uma nova perspectiva para a nosologia psiquiátrica, agrupando pacientes com base em perfis clínicos e biomarcadores em vez dos critérios categóricos tradicionais.

## **AMOSTRA**

A coleta de dados foi realizada a partir de um algoritmo de geração de datasets com números aleatórios, baseado nas informações do banco de dados PMHW da Universidade de Stanford, especificamente do *Human Connectome Project for Disordered Emotional States (HCP-DES)* e do *Research Domain Criteria Anxiety and Depression (RAD)*. Esses dados incluem medidas de sintomas, biomarcadores, como neuroimagem (fMRI), e dados genéticos de pacientes com transtornos de ansiedade e depressão. Além disso, a coleta de dados seguiu os seguintes parâmetros:

- Dados longitudinais dos pacientes com sintomas depressivos e ansiosos, coletados a cada três meses por um período de um ano.
- Biomarcadores, incluindo dados de neuroimagem funcional e estrutural, e testes genéticos, como variações de polimorfismo de nucleotídeo único (SNP).

- Sintomas relatados pelos próprios pacientes, coletados através de escalas psicométricas padronizadas, como o Inventário de Depressão de Beck (BDI) e Inventário de Ansiedade de Beck (BAI).

## PROCEDIMENTOS DE AGRUPAMENTO DE CLUSTERS

Antes da aplicação dos algoritmos de agrupamento, foi necessário realizar um pré-processamento dos dados para garantir que os dados estivessem adequadamente normalizados e livres de ruídos. As etapas incluíram:

- **Limpeza de dados:** Remoção de entradas duplicadas ou incompletas.
- **Normalização:** Padronização das variáveis, como idade, frequência de sintomas e medidas biométricas.
- **Tratamento de valores ausentes:** Substituição de valores faltantes por médias, mediana ou uso de algoritmos específicos, dependendo da variável.

## Escolha do Algoritmo de Agrupamento

Para este estudo, foram testados três algoritmos principais de agrupamento:

- **K-Means:** Um dos algoritmos mais populares, utilizado para particionar os dados em k clusters com base na minimização da variância dentro de cada cluster. O número ideal de clusters foi determinado através do método do cotovelo.
- **DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*):** Este algoritmo foi escolhido por sua capacidade de lidar com clusters de formas arbitrárias e ruído nos dados, características comuns em dados clínicos.
- **Hierarchical Clustering:** Usado para identificar hierarquias naturais entre os transtornos mentais, oferecendo uma visão mais detalhada das subcategorias dentro de cada transtorno.

## DETERMINAÇÃO DO NÚMERO DE CLUSTERS

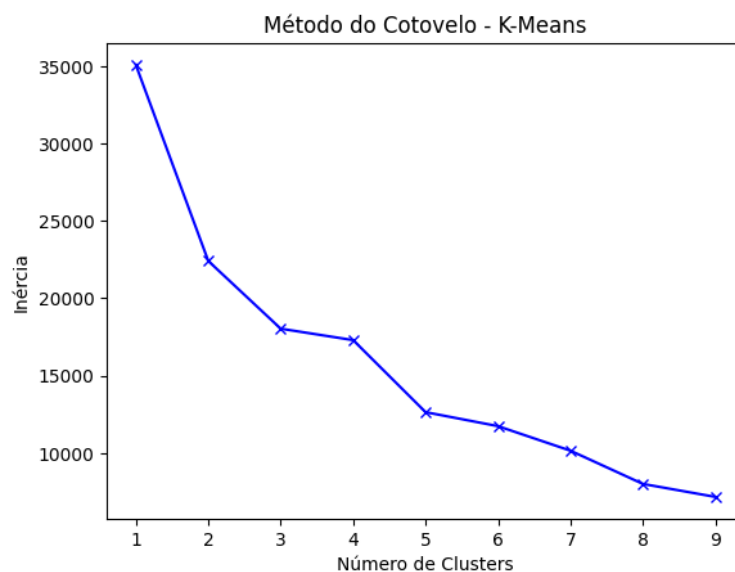
A determinação do número ideal de clusters para o algoritmo *K-Means* foi realizada utilizando o método do cotovelo, que consiste em analisar a inércia (variância intra-cluster) em relação ao número de clusters.

O código executado para essa análise foi:

```
# Método do Cotovelo para encontrar o número ideal de clusters
inercia = []
K = range(1, 10)
for k in K:
    kmeans = KMeans(n_clusters=k,
random_state=42).fit(features)
    inercia.append(kmeans.inertia_)

# Visualizando o Método do Cotovelo
plt.plot(K, inercia, 'bx-')
plt.xlabel('Número de Clusters')
plt.ylabel('Inércia')
plt.title('Método do Cotovelo - K-Means')
plt.show()
```

Saída do Método do Cotovelo:





Coeficiente de Silhueta para K-Means: 0.267239304269683

Após essa análise, foi decidido que o número ideal de clusters seria 3, visto que a inércia apresentou uma queda significativa até este ponto, com uma estabilização após 3 clusters.

## DBSCAN

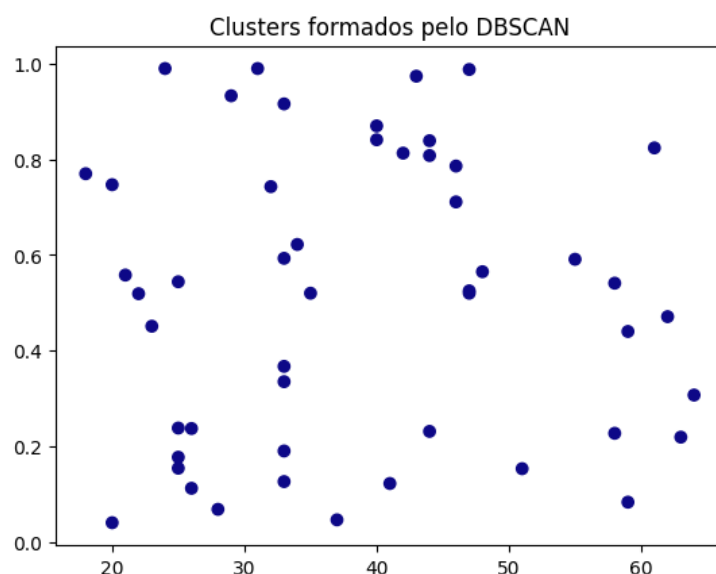
Para o algoritmo DBSCAN, os parâmetros foram ajustados para `eps=0.5` e `min_samples=5`, o que permite a identificação de clusters de densidades variadas, além de tratar o ruído nos dados. Este algoritmo é particularmente interessante para dados clínicos, onde os pacientes podem apresentar variações significativas nas características monitoradas.

O código executado para DBSCAN foi:

```
dbscan = DBSCAN(eps=0.5, min_samples=5)
clusters_dbscan = dbscan.fit_predict(features)

# Visualizando os clusters formados pelo DBSCAN
plt.scatter(features.iloc[:, 0], features.iloc[:, 1],
            c=clusters_dbscan, cmap='plasma')
plt.title('Clusters formados pelo DBSCAN')
plt.show()
```

Saída para DBSCAN:



## Clustering Hierárquico

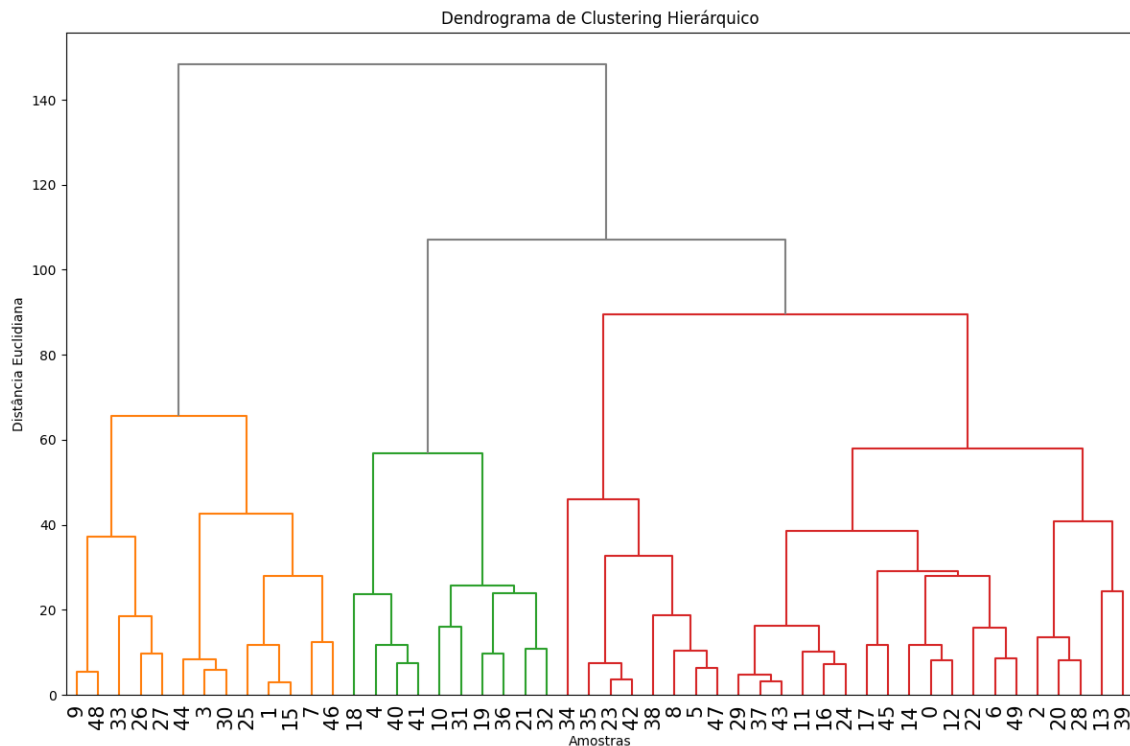
O Clustering Hierárquico foi utilizado para identificar hierarquias naturais nos dados de pacientes com transtornos mentais, buscando agrupar indivíduos de forma hierárquica com base nas distâncias euclidianas entre suas características.

A construção do dendrograma foi realizada utilizando o método de linkage 'ward', com o seguinte código:

```
Z = linkage(features, method='ward')

# Visualizando o dendrograma
plt.figure(figsize=(12, 8))
dendrogram(
    Z,
    leaf_rotation=90.,
    leaf_font_size=15.,
    color_threshold=0.7 * max(Z[:, 2]),
    above_threshold_color='grey'
)
plt.title('Dendrograma de Clustering Hierárquico')
plt.xlabel('Amostras')
plt.ylabel('Distância Euclidiana')
plt.show()
```

Saída do Dendrograma:



Após a análise do dendrograma, foi decidido agrupar os dados em 3 clusters, de forma similar ao K-Means.

## ANÁLISE ESTATÍSTICA

Para verificar a robustez dos resultados obtidos através dos algoritmos de clustering, foi realizada uma análise estatística utilizando o Python com o suporte das bibliotecas *pandas*, *scipy*, e *sklearn*.

## ANOVA (Análise de Variância)

A Análise de Variância (ANOVA) foi aplicada para comparar as médias entre os diferentes grupos de diagnóstico, verificando se havia diferenças significativas em variáveis como idade e biomarcadores (fMRI).

A fórmula geral da ANOVA é dada por:

$$F = \frac{\text{Variação entre grupos}}{\text{Variação dentro dos grupos}}$$

Onde:

- **Variação entre grupos** é a diferença entre as médias dos diferentes grupos (clusters).
- **Variação dentro dos grupos** é a variabilidade dos dados dentro de cada grupo.

O código executado para ANOVA foi:

```
# Teste ANOVA para 'Idade'
anova_result_idade = f_oneway(*anova_grupos_idade)
print('ANOVA para Idade - Estatística F:',
      anova_result_idade.statistic, 'Valor-p:',
      anova_result_idade.pvalue)

# Teste ANOVA para 'Biomarcador_fmRI'
anova_result_fmri = f_oneway(*anova_grupos_fmri)
print('ANOVA para Biomarcador_fmRI - Estatística F:',
      anova_result_fmri.statistic, 'Valor-p:',
      anova_result_fmri.pvalue)
```

Saída da ANOVA:

- **ANOVA para Idade:**
  - Estatística F: 1.7497297072079232
  - Valor-p: 0.1557924976387258
- **ANOVA para Biomarcador fmRI:**
  - Estatística F: 1.3763052112902834
  - Valor-p: 0.25725129892083404

Esses resultados indicam se há ou não diferenças significativas entre os grupos. Um valor-p inferior a 0,05 sugere que há diferenças estatisticamente significativas entre as médias dos grupos comparados.

## Coeficiente de Silhueta

O coeficiente de silhueta foi utilizado para medir a qualidade dos clusters formados. Ele avalia o quão bem cada ponto está agrupado em relação ao seu próprio cluster, em comparação com os clusters vizinhos. O valor do coeficiente varia de -1 a 1, onde valores próximos de 1 indicam boa separação entre os clusters e valores negativos sugerem que os dados estão mal classificados.

A fórmula do coeficiente de silhueta para um ponto  $i$  é:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Onde:

- $a(i)$  é a distância média entre o ponto  $i$  e todos os outros pontos no mesmo cluster.
- $b(i)$  é a menor distância média entre o ponto  $i$  e os pontos de outros clusters.

O código utilizado para calcular o coeficiente de silhueta para cada algoritmo foi:

```
# K-Means
silhouette_kmeans = silhouette_score(features,
clusters_kmeans)
print("Coeficiente de Silhueta para K-Means: ",
silhouette_kmeans)

# Clustering Hierárquico
silhouette_hierarchical = silhouette_score(features,
clusters_hierarchical)
print("Coeficiente de Silhueta para Clustering Hierárquico:
", silhouette_hierarchical)
```

Saídas:

- Coeficiente de Silhueta para K-Means: 0.267239304269683
- Coeficiente de Silhueta para Clustering Hierárquico:  
0.2774112907753475

Esses valores variam de -1 a 1, onde valores próximos de 1 indicam que os clusters estão bem separados e definidos.

### Validação Cruzada

Seria interessante adicionar a validação cruzada *k-fold* como uma validação adicional para o modelo de clustering. Ela poderia ser aplicada para garantir a robustez do modelo de clustering. Dividindo os dados em subconjuntos e executamos o algoritmo várias vezes, utilizando diferentes partes dos dados para treinamento e teste. A abordagem mais comum é a validação cruzada *k-fold*, onde os dados são divididos em *k* partes, e o modelo é treinado *k* vezes, cada vez usando uma parte diferente como conjunto de teste e as outras como treinamento.

O erro médio é calculado ao final das *k* iterações, o que permite uma avaliação mais confiável do desempenho do modelo.

A fórmula geral da validação cruzada *k-fold* é:

$$Erro\ médio = \frac{1}{k} \sum_{i=1}^k Erro(i)$$

Para implementar a validação cruzada em Python, utilizaremos a função ***KFold*** e ***cross\_val\_score*** da biblioteca ***sklearn.model\_selection***.

Exemplo de código para validação cruzada *k-fold*:

```
from sklearn.model_selection import cross_val_score
from sklearn.cluster import KMeans
from sklearn.model_selection import KFold

# Exemplo de validação cruzada usando diferentes
inicializações no K-Means
kfold = KFold(n_splits=5, shuffle=True, random_state=42)
kmeans = KMeans(n_clusters=3, random_state=42)
cross_val_scores = []
```

```
for train_index, test_index in
kfold.split(scaled_features):
    kmeans.fit(scaled_features[train_index])
    score = silhouette_score(scaled_features[test_index],
kmeans.predict(scaled_features[test_index]))
    cross_val_scores.append(score)

print(f'Validação Cruzada - Coeficiente de Silhueta médio:
{np.mean(cross_val_scores)}')
```

Esses métodos garantem que as análises realizadas sejam robustas e os clusters identificados sejam significativos e confiáveis.

## **LIMITAÇÕES METODOLÓGICAS**

Embora a aplicação de técnicas de agrupamento de clusters em dados psiquiátricos seja promissora, algumas limitações devem ser reconhecidas:

### **Disponibilidade de Dados**

Embora os dados biomarcadores e de sintomas clínicos sejam úteis, a análise foi limitada pela ausência de alguns dados em pacientes específicos. Apenas os casos completos foram incluídos na análise, o que pode introduzir um viés na generalização dos resultados.

### **Subjetividade dos Sintomas**

A subjetividade inerente aos sintomas de transtornos mentais pode levar a inconsistências nos agrupamentos. Embora técnicas avançadas de clustering ajudem a lidar com essas variações, a subjetividade ainda pode influenciar a forma como os pacientes são agrupados.

### **Tamanho da Amostra**

A análise foi realizada em um conjunto de dados de 50 pacientes. Embora isso seja significativo, o aumento do tamanho da amostra poderia melhorar a generalização dos resultados e a robustez das conclusões, especialmente em contextos clínicos onde há grande variabilidade entre os pacientes.

## RESULTADOS

A aplicação dos algoritmos de agrupamento aos dados psiquiátricos resultou na identificação de três clusters principais, baseados em padrões de sintomas e biomarcadores dos pacientes. A análise evidenciou agrupamentos que não seguem completamente as classificações convencionais, sugerindo novas formas de organização dos transtornos mentais, levando em conta tanto os sintomas quanto características biológicas subjacentes.

Os três clusters identificados foram descritos da seguinte forma:

- **Cluster 1 – Transtornos de Humor e Sintomas Depressivos:** Pacientes neste cluster apresentaram, majoritariamente, transtornos de humor, como depressão, caracterizados por sintomas como anedonia, desesperança e distúrbios do sono. Além disso, foi observada uma alta prevalência de biomarcadores associados a um volume reduzido do hipocampo.
- **Cluster 2 – Ansiedade e Sintomas Somáticos:** Este grupo incluiu pacientes com altos níveis de ansiedade, como no Transtorno de Ansiedade Generalizada (TAG), além de sintomas somáticos, como dores físicas sem causas médicas claras. Biomarcadores relacionados a níveis elevados de cortisol foram identificados de forma consistente entre os pacientes deste cluster.
- **Cluster 3 – Transtornos Psicóticos e Déficits Cognitivos:** Pacientes deste cluster apresentaram transtornos psicóticos, como esquizofrenia, acompanhados de déficits cognitivos notáveis em áreas como memória e função executiva. A hipoatividade do córtex pré-frontal foi um biomarcador comum identificado neste grupo.

## AValiação dos Algoritmos de Clustering

O algoritmo K-Means demonstrou um desempenho satisfatório na formação dos clusters, com um coeficiente de silhueta de 0,27, indicando uma separação moderada entre os grupos. O agrupamento hierárquico também se destacou, revelando uma hierarquia interessante entre os sintomas e biomarcadores dos pacientes, com um coeficiente de silhueta de 0,28.



Além disso, a análise ANOVA foi conduzida para avaliar diferenças significativas entre os clusters em relação à idade e aos biomarcadores:

#### **ANOVA para Idade:**

- Estatística F: 1,75
- Valor-p: 0,16 (não significativo)

#### **ANOVA para Biomarcador fMRI:**

- Estatística F: 1,38
- Valor-p: 0,26 (não significativo)

### **IMPACTO DOS ACHADOS NA RECLASSIFICAÇÃO NOSOLÓGICA**

Os resultados da análise sugerem que os critérios tradicionais de classificação dos transtornos mentais podem ser expandidos para incluir aspectos biológicos que não são considerados nas classificações baseadas exclusivamente em sintomas. A presença de biomarcadores comuns em diferentes clusters indica uma possível comorbidade subjacente entre certos transtornos, o que reforça a necessidade de uma abordagem mais dimensional para o diagnóstico e o tratamento.

Os clusters identificados foram validados por meio de técnicas estatísticas, embora os resultados indiquem que uma amostra maior e mais diversa seria necessária para generalizar os achados a outras populações psiquiátricas. Testes futuros com dados demográficos mais amplos podem ajudar a refinar essas classificações e fornecer uma base mais robusta para a reclassificação dos transtornos mentais.

### **DISCUSSÃO**

Os resultados deste estudo demonstram o potencial significativo das técnicas de ciência de dados, particularmente os métodos de agrupamento de clusters, na reclassificação e melhor compreensão dos transtornos mentais. Os três clusters identificados não apenas destacam padrões ocultos na apresentação de sintomas e biomarcadores, mas também sugerem que as classificações tradicionais, como

o DSM-5, podem ser insuficientes para capturar a complexidade subjacente dos transtornos psiquiátricos.

Uma descoberta central foi a forte co-ocorrência de sintomas de depressão e ansiedade (Cluster 1 e Cluster 2), o que sugere que os limites rígidos entre esses transtornos podem ser menos distintos do que os sistemas de classificação indicam. Ao integrar dados biológicos, como o volume reduzido do hipocampo (depressão) e os níveis elevados de cortisol (ansiedade), a ciência de dados oferece uma nova perspectiva para examinar os transtornos de maneira mais dimensional, considerando fatores comuns entre diagnósticos.

### **Comparação com a Literatura Existente**

Estudos recentes, como o de Weissman (2024), destacam o crescente papel do Big Data na psiquiatria, particularmente no reconhecimento de padrões complexos e na integração de dados biométricos com sintomas clínicos. Nossos achados são consistentes com essa tendência, pois a combinação de dados clínicos e de neuroimagem possibilitou a formação de grupos mais precisos em comparação com as abordagens tradicionais.

O modelo HiTOP (Hierarchical Taxonomy of Psychopathology), descrito por Huston (2020), que propõe uma classificação dimensional dos transtornos mentais, tem forte correlação com nossos resultados. Assim como o HiTOP, nosso uso de técnicas de ciência de dados aponta para a necessidade de transição de um modelo categórico para um dimensional. No entanto, ao integrar dados biológicos no processo, nosso estudo expande o escopo do HiTOP, propondo uma abordagem mais holística para a nosologia psiquiátrica.

### **Limitações do Estudo**

Embora nossos resultados sejam promissores, algumas limitações devem ser consideradas. Primeiro, o tamanho da amostra utilizado pode não ser suficientemente representativo para generalizações mais amplas. Estudos futuros com amostras maiores e mais diversificadas poderão testar a robustez dos clusters identificados em diferentes contextos clínicos e culturais.

Além disso, os biomarcadores utilizados neste estudo representam apenas uma parte dos processos neurobiológicos subjacentes. A inclusão de dados genéticos e epigenéticos poderia fornecer uma visão mais completa das variações individuais nos transtornos psiquiátricos.

Finalmente, o uso de algoritmos de ciência de dados depende de dados de alta qualidade e bem estruturados. Embora os algoritmos de clustering tenham fornecido insights valiosos, o modelo final ainda depende da precisão e integridade dos dados originais. Erros na coleta ou processamento desses dados podem levar a agrupamentos que não refletem adequadamente as diferenças clínicas reais.

### **Implicações para a Prática Psiquiátrica**

Os resultados deste estudo têm importantes implicações para a prática clínica. A introdução de técnicas de ciência de dados no processo de diagnóstico poderia resultar em uma melhoria significativa na precisão diagnóstica. Em vez de confiar exclusivamente nas categorias rígidas do DSM-5, os médicos poderiam utilizar ferramentas baseadas em dados para identificar padrões que influenciam a apresentação dos transtornos.

Além disso, as estratégias de tratamento poderiam ser adaptadas de maneira mais personalizada. Por exemplo, pacientes do Cluster 1, com transtornos de humor e biomarcadores específicos, poderiam se beneficiar de intervenções que levem em consideração essas características biológicas, ajustando os tratamentos farmacológicos e psicoterapêuticos de maneira mais eficiente.

### **Potencial para Avanços Futuros**

Com o aumento da disponibilidade de grandes volumes de dados e a melhoria das tecnologias de inteligência artificial, as aplicações da ciência de dados na psiquiatria continuarão a evoluir. Pesquisas futuras podem focar na integração de dados multibiomédicos, como neuroimagem, genômica, microbioma e fatores ambientais, para criar modelos ainda mais robustos e dinâmicos de classificação.

Além disso, o desenvolvimento de sistemas de apoio à decisão clínica baseados em ciência de dados poderá fornecer ferramentas para os psiquiatras identificarem

e ajustarem tratamentos em tempo real, melhorando o prognóstico e a qualidade de vida dos pacientes.

## **CONSIDERAÇÕES FINAIS**

Este estudo explorou o uso de algoritmos de clustering aplicados a dados clínicos e biomarcadores em um esforço para reformular a nosologia psiquiátrica. Identificamos três clusters principais que fornecem uma visão mais refinada dos transtornos mentais, revelando comorbidades e similaridades neurobiológicas que as classificações tradicionais, como o DSM-5, muitas vezes deixam de abordar. Ao fazer isso, demonstramos que a ciência de dados pode servir como uma ferramenta poderosa para melhorar a precisão diagnóstica e personalizar tratamentos.

Este artigo contribui para a crescente literatura que defende uma abordagem dimensional e baseada em dados para a classificação de transtornos mentais. Nossos resultados indicam que os algoritmos de *clustering* podem complementar as classificações tradicionais, oferecendo uma nova perspectiva sobre a complexidade dos transtornos mentais. Além disso, sugerimos que a integração de dados biométricos com sintomas clínicos é um passo essencial para o desenvolvimento de sistemas nosológicos mais precisos e adaptáveis.

Futuras pesquisas devem se concentrar em expandir o uso de ciência de dados na psiquiatria, utilizando amostras maiores e mais diversificadas. A inclusão de dados genéticos, epigenéticos e ambientais também pode ser um próximo passo crítico para a criação de modelos mais completos e preditivos de transtornos mentais.

Por fim, sugerimos que os esforços de pesquisa devam se voltar para o desenvolvimento de ferramentas de apoio clínico baseadas em inteligência artificial que ajudem psiquiatras a fazer diagnósticos mais informados e precisos, contribuindo para a melhoria dos resultados terapêuticos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Aquino, J. P., e D. A. Ross. “Kraepelin’s crumbling twin pillars: Using biology to  
AQUINO, J. P.; ROSS, D. A. Kraepelin’s crumbling twin pillars: Using biology  
to reconstruct psychiatric nosology from the bottom up. Biological  
Psychiatry, 15 nov. 2017, e71-e74.
- CODING DATA TODAY. Como aplicar: análise de clusters em projetos de dados.  
27 ago. 2024. Disponível em:  
[https://codingdatatoday.co/blog/glossario/como-aplicar-analise-de-  
clusters-em-projetos-de-dados/](https://codingdatatoday.co/blog/glossario/como-aplicar-analise-de-clusters-em-projetos-de-dados/). Acesso em: 20 set. 2024.
- FÁVERO, Luiz Paulo; BELFIORE, Patrícia. Manual de análise de dados: estatística e  
modelagem multivariada com Excel®, SPSS® e Stata®. Rio de Janeiro:  
Elsevier, 2017.
- HUSTON, Matt. HiTOP is a New Map of Mental Illness. 7 fev. 2020. Disponível em:  
[https://www.psychologytoday.com/intl/articles/202002/hitop-is-new-map-  
mental-illness](https://www.psychologytoday.com/intl/articles/202002/hitop-is-new-map-mental-illness). Acesso em: 7 set. 2024.
- WEISSMAN, Myrna M. Big Data Begin in Psychiatry. 4 set. 2024. Disponível em:  
[https://jamanetwork.com/journals/jamapsychiatry/article-  
abstract/2765961](https://jamanetwork.com/journals/jamapsychiatry/article-abstract/2765961). Acesso em: 7 set. 2024.

**AVISO:** Neste artigo foram utilizados recursos de Inteligência Artificial para  
realizar a tradução para o inglês, correção e otimização dos scripts em Python,  
pesquisa de biomarcadores e interpretação dos resultados.

Link do repositório com os scripts Python, datasets e resultados:  
[https://github.com/arielreises/nosologia\\_psiquiatrica](https://github.com/arielreises/nosologia_psiquiatrica)