**Acesso à Saúde Mental no Estado de São Paulo: Uma Análise Exploratória Baseada em Dados do DATASUS**

**Larissa Yumi Tinem1, Leo Tsuchida Hoçoya1,2**

1Faculdade de Computação e Informática (FCI)  
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

2Sistemas de Informação – Faculdade de Computação e Informática (FCI) – Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{10409094, 10332699}@mackenzista.com.br

***Resumo.*** *Este trabalho apresenta uma análise exploratória do acesso à saúde mental no estado de São Paulo, utilizando dados do DATASUS, complementados por informações do IBGE. O estudo enfoca a cobertura de Centros de Atenção Psicossocial (CAPS), Atenção Básica de Saúde e indicadores socioeconômicos regionais. A partir de técnicas de classificação descritiva com Inteligência Artificial, como clustering e árvores de decisão, as regiões do país serão categorizadas quanto ao déficit de atendimento, permitindo identificar padrões, desigualdades e lacunas. Os resultados fornecerão subsídios para o planejamento de políticas públicas mais eficazes e a priorização de recursos na atenção psicossocial.*

***Palavras-chave:*** *Saúde Mental, Acesso à Saúde, CAPS, Classificação Descritiva, Inteligência Artificial, Clustering*

***Abstract.*** *This study presents an exploratory analysis of access to mental health services in São Paulo state, using data from DATASUS, complemented by information from the IBGE. The study focuses on the coverage of Psychosocial Care Centers (CAPS), Primary Health Care, and regional socioeconomic indicators. Using descriptive classification techniques with Artificial Intelligence, such as clustering and decision trees, regions of the country will be categorized according to their service coverage deficits, allowing the identification of patterns, inequalities, and gaps. The results will provide insights to support more effective public policy planning and resource allocation in psychosocial care.*

***Keywords:*** *Mental Health, Access to Health, CAPS, Descriptive Classification, Artificial Intelligence, Clustering.*

1. **Introdução**

Conforme o artigo 196 da Constituição Federal de 1988, ‘a saúde é direito de todos e dever do Estado’ (Brasil, 1988). Entretanto, a efetivação desse direito enfrenta alguns desafios, como barreiras geográficas, má distribuição das unidades de saúde e precariedade dos serviços oferecidos, principalmente em regiões mais afastadas dos grandes centros urbanos como regiões rurais ou periféricas, dificultando o acesso dessa população à cuidados básicos e emergenciais (Rocha et al., 2021).

Já segundo Dimenstein et al. (2021), a desigualdade ao acesso à saúde mental persiste mesmo após a implementação da Reforma Psiquiátrica e Rede de Atenção Psicossocial (RAPS), limitando o acesso equitativo aos serviços especializados e concentrando os recursos em regiões urbanas mais centrais. Essa desigualdade regional evidencia a necessidade de análise detalhada dos padrões de cobertura e da identificação de lacunas no acesso a serviços de saúde mental em todo o país.

A Política Nacional de Saúde Mental, apoiada na lei nº 10.216/2001, busca consolidar um modelo de atenção à saúde mental aberto e de base comunitária, promovendo o convívio com a família e a comunidade em vez do isolamento institucional, como predominava nos antigos manicômios.

Os Centros de Atenção Psicossocial (CAPS) têm como objetivo acolher pessoas com transtornos mentais, bem como indivíduos com transtornos causados pelo uso de álcool e outras drogas, promovendo sua integração social e familiar, apoiando o desenvolvimento da autonomia e oferecendo atendimento médico e psicológico. A principal característica desses serviços é inseri-los em um ambiente social e cultural concreto, chamado de “território”, que corresponde ao espaço da cidade onde ocorre o cotidiano dos usuários e de suas famílias. Dessa forma, os CAPS representam a estratégia central da reforma psiquiátrica no Brasil.

Estudos recentes indicam que a cobertura dos CAPS e a integração com a Atenção Básica são fatores determinantes para reduzir internações psiquiátricas e ampliar o acesso a tratamento adequados. Miliauskas et al. (2019), ao analisarem as regiões metropolitanas do Rio de Janeiro e de São Paulo, observaram que a ampliação desses serviços está associada à diminuição de hospitalizações, evidenciando a importância de políticas públicas que fortaleçam a rede de atenção psicossocial.

Diante desse cenário, este trabalho tem como objetivo realizar uma análise exploratória do acesso à saúde mental no estado de São Paulo, utilizando dados disponíveis no DATASUS, complementados por informações do IBGE, como quantidade de CAPS, leitos, população por município e indicadores socioeconômicos. A partir desses dados estruturados, emergem duas questões centrais para a presente pesquisa:

* A distribuição de leitos e internações psiquiátricas, além da quantidade de CAPS, no estado de São Paulo é desigual entre regiões?
* Pacientes residentes em municípios de pequeno porte precisam se deslocar para centros maiores a fim de obter atendimento hospitalar em saúde mental?

Responder a essas questões é fundamental não apenas para compreender os padrões atuais de acesso, mas também para subsidiar políticas públicas que promovam maior equidade no cuidado em saúde mental no estado.

Para alcançar esses objetivos, será adotada uma abordagem de categorização descritiva das regiões do estado de São Paulo, com base em indicadores de cobertura em saúde mental. Os municípios serão agrupados em diferentes níveis de acesso (alto, médio e baixo), definidos a partir do número de internações psiquiátricas por 100 mil habitantes. Essa categorização será complementada pelo cruzamento com variáveis demográficas e estruturais, como porte populacional, presença de Centros de Atenção Psicossocial (CAPS), disponibilidade de leitos e indicadores socioeconômicos.

Essa estratégia permitirá identificar padrões e agrupamentos territoriais que evidenciem desigualdades no acesso aos serviços, distinguindo regiões com alta concentração de recursos daquelas com déficit de atenção psicossocial. Além disso, a análise possibilitará a identificação de clusters de desigualdade, em que coexistem municípios saturados de demanda e outros com cobertura insuficiente, fornecendo subsídios concretos para o planejamento de políticas públicas e a alocação mais equitativa de investimentos em saúde mental.

**2. Aspectos Éticos e Responsabilidade**

O uso de Inteligência Artificial (IA) em análises de saúde pública envolve questões éticas importantes, especialmente quando se trabalha com dados sensíveis sobre indivíduos e comunidades. Neste estudo, os dados utilizados são provenientes de fontes oficiais, garantindo anonimização e proteção à privacidade, evitando a identificação de indivíduos. Ressalta-se que o respeito à confidencialidade e à segurança das informações é um requisito ético fundamental, alinhado às normas legais e regulatórias vigentes.

Além disso, o uso de algoritmos de IA, como *clustering* e árvores de decisão, deve ser feito com cuidado, garantindo que os resultados sejam claros, confiáveis e não prejudiquem grupos específicos. Os pesquisadores têm a responsabilidade de usar os insights obtidos para orientar políticas públicas e decisões estratégicas de forma ética, justa e socialmente responsável, sem reforçar desigualdades ou estigmatizar pessoas vulneráveis.

Além disso, é importante comunicar os resultados de maneira clara e acessível, evitando simplificações que possam gerar mal-entendidos sobre o acesso à saúde mental nas diferentes regiões. A ética no desenvolvimento da solução envolve não apenas proteger os dados, mas também usar as informações de forma consciente para promover equidade, inclusão e melhorar a atenção psicossocial no país.

**3. Dataset, Análise Exploratória e Preparação dos Dados**

**3.1. Base de Dados**

A base de dados utilizada neste estudo foi composta por duas fontes oficiais e complementares.

A primeira refere-se ao Sistema de Informações Hospitalares do SUS (SIH-SUS), disponibilizado pelo [DATASUS](https://datasus.saude.gov.br/), que registra as internações hospitalares financiadas pelo SUS em todo o país. Para este trabalho, os dados foram obtidos por meio da biblioteca *Pysus*, em ambiente *Google Colab*, possibilitando a automação do processo de download, leitura e conversão dos arquivos para o formato .parquet, o que facilita a manipulação em análises computacionais.

A segunda fonte corresponde às estimativas populacionais do Brasil e do estado de São Paulo, fornecidas pelo [Instituto Brasileiro de Geografia e Estatística (IBGE)](https://www.ibge.gov.br/) por meio de sua *API oficial*. Os dados foram consultados utilizando a função FetchData, que retorna os resultados em formato JSON. Posteriormente, os dados foram convertidos para o formato .parquet, garantindo compatibilidade com os arquivos do SIH-SUS e permitindo a integração entre as diferentes bases.

Por se tratar de um volume massivo de informações, uma cópia dos dados processados foi armazenada permanentemente no *Google Drive*. Para visualizar o código de download dos dados, veja o arquivo-fonte *setup.ipynb* no repositório do *Github*. O link para acesso ao Google Drive também está presente no Github.

**3.2. Análise Exploratória e Preparação dos Dados**

Nesta primeira etapa, para a preparação dos dados e análise exploratória, utilizou-se como escopo uma amostra do conjunto de dados do SIH-SUS e do IBGE, especificamente para o estado de São Paulo no ano de 2015. O foco inicial foi uma análise superficial do conjunto de dados para entender sua estrutura.

Inicialmente, uma análise superficial foi realizada para compreender a estrutura e o conteúdo do *dataset*. Para isso, as funções *df.head()* e *df.tail()* foram utilizadas para visualizar, respectivamente, as 5 primeiras e últimas linhas do *dataset*. A disposição das colunas foi verificada com a listagem completa dos seus nomes. Essa etapa inicial permitiu identificar a presença de valores ausentes representados por *strings* vazias (‘ ‘). Para garantir a precisão de análises futuras e facilitar o entendimento dos dados, essas *strings* foram convertidas para o valor NaN.

O processo de preparação avançou com a filtragem dos dados, garantindo que o escopo de análise se restringisse ao tema. Foram selecionadas apenas as internações relacionadas a transtornos mentais e comportamentais, conforme a classificação do CID-10 (códigos que variam de F00 a F99). Em seguida, filtrou-se também os registros referentes ao Estado de São Paulo, utilizando o código 35. Uma validação inicial das colunas QT\_DIARIAS e DIAS\_PERM foi realizada, confirmando sua consistência.

A base de dados foi enriquecida com a junção das informações de população do IBGE. Para garantir que os dados do IBGE e do SIH-SUS pudessem ser combinados corretamente, foi criada uma coluna “D1C\_6\_DIGITOS”. Nela, o código original de 7 dígitos foi reduzido para 6, para ficar compatível com o padrão usado no Datasus.

**3.3. Análise Exploratória: Primeiros Resultados**

A partir dos dados preparados, a análise exploratória inicial focou na distribuição territorial das internações psiquiátricas. Foi calculada a taxa de internações por 100 mil habitantes (internações\_por\_100k) para cada município, uma métrica fundamental para comparar o acesso à saúde mental entre cidades com diferentes tamanhos populacionais.

Com base nessa taxa, os municípios foram classificados em três categorias:

* **Baixo Acesso (0 a 50 internações por 100k habitantes):** Representa municípios com uma taxa muito baixa de internações, o que pode sugerir uma suboferta de serviços ou a necessidade de deslocamento dos pacientes.
* **Médio Acesso (50 a 150 internações por 100k habitantes):** Indica uma taxa intermediária de internações.
* **Alto Acesso (150 ou mais internações por 100k habitantes):** Caracteriza os municípios que concentram um grande número de internações, sugerindo que funcionam como centros de referência que atendem a sua própria população e a de cidades vizinhas.

Essa categorização incial servirá como base para aprofundar o estudo sobre a desigualdade territorial na distribuição dos serviços de saúde mental.

**4. Referencial Teórico**

Este trabalho integra conceitos de direito à saúde pública, análise de dados e Inteligência Artificial (IA) para propor uma solução tecnológica ao problema da desigualdade no acesso aos serviços de saúde mental no Brasil.

**4.1. Direito e Acesso à Saúde Mental**

Segundo o artigo 196 da Constituição Federal de 1988, a saúde é um direito de todos e dever do Estado, devendo ser garantida por meio de políticas sociais e econômicas (BRASIL, 19880. No caso da Saúde Mental, a Lei nº 10.216/2001 estabeleceu a Política Nacional de Saúde Mental, que promove a atenção comunitária, o convívio familiar e a integração social.

Apesar dessas políticas, desigualdades regionais ainda persistem, afetando o acesso aos serviços especializados de saúde mental, como os Centros de Atenção Psicossocial (CAPS), leitos psiquiátricos, entre outros. Estudos indicam que a cobertura desses serviços é determinante para reduzir internações psiquiátricas e ampliar o acesso ao tratamento adequado (Miliauskas et al., 2019).

**4.2. Impacto do CAPS e Integração coma Atenção Básica**

Diversos estudos têm destacado a relevância dos CAPS como fator determinante para a redução de internações psiquiátricas e ampliação do acesso a cuidados em saúde mental. Miliauskas et al. (2019), ao investigarem as regiões metropolitanas de São Paulo e Rio de Janeiro, observaram que a expansão desses serviços contribuiu para a diminuição das hospitalizações, evidenciando a importância de políticas públicas que fortaleçam a rede de atenção psicossocial.

Além disso, a integração dos CAPS com a Atenção Básica é considerada fundamental para consolidar uma rede de cuidado mais resolutiva e equitativa, especialmente em regiões carentes de infraestrutura. Assim, compreender a distribuição desses serviços e sua cobertura populacional é um passo essencial para identificar desigualdades e subsidiar ações estratégicas em saúde mental.

**4.3. Análise Exploratória de Dados e Inteligência Artificial**

A análise exploratória de dados (AED) é um processo fundamental em qualquer estudo estatístico ou computacional, pois permite compreender a qualidade e as características dos dados antes da aplicação de técnicas mais complexas. Segundo Wickham e Grolemund (2017), a AED é um processo criativo, guiado por perguntas, que busca revelar padrões, verificar inconsistências e gerar hipóteses a partir da observação. A cada nova questão formulada, novos aspectos do conjunto de dados são revelados, aumentando a chance de descobertas relevantes.

Indicadores como número de CAPS por 100 mil habitantes, taxa de internações psiquiátricas e dados socioeconômicos podem ser combinados para identificar regiões com cobertura insuficiente e subsidiar políticas públicas mais efetivas (Dimenstein et al., 2021).

Além de EAD, técnicas de Inteligência Artificial (IA) podem ser aplicadas para analisar dados complexos de saúde pública, identificando padrões mais sutis e relações não lineares entre as variáveis. Por exemplo, clusterização pode ser usada para agrupar municípios com características semelhantes de cobertura e vulnerabilidade enquanto Random Forest pode ser usada para entender quais fatores socioeconômicos ou demográficos mais influenciam a desigualdade de acesso.

Assim, a integração da EDA com técnicas de IA permite uma análise mais completa e estratégica, oferecendo subsídios para planejamento e tomada de decisão em políticas públicas de saúde mental, especialmente em contextos de desigualdade regional.

**5. Metodologia**

Este é um estudo exploratório que visa analisar o acesso à saúde mental no Brasil por meio de dados estruturados provenientes de fontes oficiais, como o DATASUS, com informações sobre quantidade de CAPS, leitos psiquiátricos e internações em saúde mental por município, e o IBGE com dados populacionais, demográficos e indicadores socioeconômicos. Os dados serão organizados em um conjunto único, permitindo a integração das informações sobre a cobertura de serviços de saúde mental, com tratamento de inconsistências, valores ausentes e anonimização se necessário.

A análise exploratória envolverá a verificação de estatísticas descritivas, distribuição geográfica dos CAPS e leitos por município, além da correlação entre indicadores socioeconômicos e a disponibilidade de serviços de saúde mental. Essa etapa permitirá compreender o comportamento das variáveis, identificar padrões preliminares e preparar os dados para a aplicação de algoritmos, incluindo normalização, padronização e criação de categorias quando necessário.

Para identificar padrões e desigualdades no acesso à saúde mental, será empregada uma abordagem de classificação descritiva, utilizando técnicas de *clustering*, como o *K-means* e *Random Forest*. O *clustering* permitirá agrupar municípios ou regiões com características semelhantes em termos de número de CAPS, leitos, população e indicadores socioeconômicos, enquanto o *Random Forest* possibilitará compreender quais fatores influenciam a categorização das regiões, fornecendo interpretabilidade e insights relevantes para a formulação de políticas públicas.

Os dados utilizados são provenientes do Sistema de Informações Hospitalares (SIH-SUS), do Sistema de Informações Ambulatoriais de Saúde (SIA-SUS) e Instituto Brasileiro de Geografia e Estatística (IBGE).

Todo o processamento e análise dos dados será realizado no ambiente do *Google Colab* com *Python*, utilizando bibliotecas como Pysus para extrair os dados das bases públicas do Datasus, *Pandas* e *NumPy* para manipulação, *Matplotlib* e *Seaborn* para visualização, e *Scikit-learn* para implementação dos algoritmos de clustering e classificação. Espera-se que a aplicação dessa metodologia permita classificar regiões do país quanto ao déficit de cobertura em saúde mental, identificar padrões e agrupamentos relevantes, e fornecer subsídios para gestores públicos na priorização de investimentos e no planejamento estratégico da Rede de Atenção Psicossocial (RAPS), contribuindo para reduzir desigualdades regionais no acesso aos serviços especializados.

**6. Resultados Esperados**

Espera-se que este estudo permita identificar padrões e desigualdades no acesso à saúde mental em diferentes regiões do estado de São Paulo, destacando municípios ou áreas com maior déficit de cobertura de serviços, como CAPS e leitos psiquiátricos. A aplicação de técnicas de classificação descritiva, como *clustering* e árvores de decisão, deverá possibilitar a categorização das regiões em grupos com alto, médio ou baixo déficit de atendimento, fornecendo uma visão clara de onde os recursos estão concentrados e onde existem lacunas significativas.

Além disso, os resultados deverão evidenciar a relação entre indicadores socioeconômicos e a disponibilidade de serviços de saúde mental, permitindo compreender como fatores como renda, IDH e vulnerabilidade social influenciam o acesso da população aos serviços especializados. Espera-se que os agrupamentos e padrões identificados possam fornecer subsídios concretos para gestores públicos e formuladores de políticas, auxiliando na priorização de investimentos e no planejamento estratégico da Rede de Atenção Psicossocial (RAPS).

Por fim, o estudo pretende gerar informações visualmente interpretáveis, por meio de gráficos e mapas de distribuição, que facilitem a compreensão das desigualdades regionais e contribuam para a formulação de políticas públicas mais equitativas e eficazes no atendimento à saúde mental.

**8. Referências bibliográficas**

BRASIL. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Presidência da República, [2024]. Disponível em: https://www.gov.br/conselhonacional-de-saude/pt-br/acesso-a-informacao/legislacao/outrasnormativas/constituicaofederal.pdf/view.

BRASIL. *Lei nº 10.216, de 6 de abril de 2001*. Dispõe sobre a proteção e os direitos das pessoas portadoras de transtornos mentais e redireciona o modelo assistencial em saúde mental. Diário Oficial da União: seção 1, Brasília, DF, 6 abr. 2001. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/leis_2001/l10216.htm>.

MILIAUSKAS, C. R.; FAUS, D. P.; JUNKES, L.; RODRIGUES, R. B.; JUNGER, W. *Associação entre internações psiquiátricas, cobertura de CAPS e atenção básica em regiões metropolitanas do RJ e SP, Brasil*. Ciência & Saúde Coletiva, v. 24, n. 5, p. 1935-1944, maio 2019.

Rocha HA, Reis IA, Santos MAC, Melo APS, Cherchiglia ML. *Internações psiquiátricas pelo Sistema Único de Saúde no Brasil ocorridas entre 2000 e 2014*. Rev Saude Publica. 2021;55:14. <https://doi.org/10.11606/s15188787.2021055002155>

DIMENSTEIN, M. et al. *Equidade e acesso aos cuidados em saúde mental em três estados nordestinos*. Ciência & Saúde Coletiva, v. 26, n. 5, p. 1727–1738, maio 2021.

WICKHAM, H.; GROLEMUND, G. *R for Data Science: import, tidy, transform, visualize, and model data*. 1. ed. Sebastopol: O’Reilly Media, 2017. Disponível em: <https://r4ds.had.co.nz/exploratory-data-analysis.html>.