

Compreensão e preparação de dados aplicados na estratificação de risco em ansiedade

por Evaldo de Oliveira da Silva, Marcello Peixoto Bax, Auxiliatrice Caneschi Badaró, Raissa Barros Dominato, Carla Maria Guimarães de Assis, Manuela Mayworm Jens, Eduarda Novais Falcão, Rubens Amorim Souto Neto, João Vitor Carvalho, João Gabriel Garcia Paiva e Samuel Henrique Reis do Nascimento*

A pesquisa busca compreender dados estruturados identificados a partir estratificação de risco em ansiedade, juntamente com dados textuais (não estruturados) de tratamentos psicoterápicos a fim de prepará-los para análise qualitativa e quantitativa.



Fonte: Banco de Imagens Wix (2022)

Transtornos mentais estão entre as principais causas de incapacidade no mundo. Estudos do Global Burden Disease (GBD) apresentam que milhões de pessoas no mundo sofrem de depressão, ansiedade e cometem suicídios (WHITEFORD; FERRARI; DEGENHARDT, 2016). De acordo com a Mayo Foundation for Medical Education and Research (MAYO CLINIC, 2022) algumas das causas da ansiedade são: trauma, estresse devido a uma doença, acúmulo de estresse, personalidade, outros transtornos de saúde mental, parentes próximos com transtorno de ansiedade, drogas ou álcool.



De acordo com os números publicados pelo GDB, aproximadamente 300 milhões de pessoas possuem algum tipo de transtorno de ansiedade (DATTANI; RITCHIE; ROSER, 2018). A melhoria da qualidade de dados para estudos na área da saúde mental tem sido um desafio. Porém, o objetivo de enriquecer estudos que visam melhorar a assistência à saúde mental se faz relevante. Dados sobre parâmetros epidemiológicos, psicopatologias, faixas etárias, estimativas globais e nacionais são frequentemente usados em estudos em saúde mental (DATTANI; RITCHIE; ROSER, 2018). Dados utilizados para análise da condição mental dos pacientes envolvem tanto dados estruturados, tais como instrumentos de avaliação e questionários para estratificação de risco, quanto dados não-estruturados (evolução dos tratamentos com base em técnicas da psicologia) (MELLO; 2015, PAULA, 2019). O governo brasileiro por meio do Ministério da Saúde orienta a utilização de instrumentos que visam direcionar a assistência à saúde (incluindo saúde mental), a fim de revisar metas e resultados esperados em diferentes intervenções, conforme as necessidades dos pacientes (BRASIL, 2014).

Técnicas são utilizadas por profissionais em saúde mental para descrever os casos e tratamentos psicoterápicos. A Terapia Cognitivo-Comportamental (TCC) tem sido acolhida em diferentes campos que envolvam trabalhos terapêuticos, isso por apresentar um conjunto vasto de estudos e bases científicas, definindo-a como uma prática baseada em evidências em psicologia. Entre suas particularidades, considera-se o uso de protocolos de atendimentos e manuais que guiam a atuação prática a partir do desenho do caso clínico e possíveis diagnósticos. Ademais, tendo em vista que os transtornos de ansiedade se mostram cada vez mais presentes na vida da população, torna-se notória a necessidade crescente de estudo e tratamento destes no campo da Psicologia e de grande relevância para a pesquisa desse transtorno (MELO; LOURENÇO, 2020).

Os tipos de dados mencionados, tanto da estratificação de riscos, quanto dos tratamentos psicoterápicos, podem servir para impulsionar estudos científicos na área da psicologia, onde o conhecimento sobre eles pode ser reutilizado e

compartilhado em diferentes pesquisas na melhoria da aplicação dos tratamentos. Assim, entende-se a importância de compreender e preparar os diferentes dados produzidos. A estratificação de riscos em saúde mental pode ser utilizada como instrumento para orientar os tratamentos psicoterápicos na avaliação dos sinais e sintomas do paciente. No entanto, os dados na forma que estão estruturados, podem dificultar as análises da eficácia dos tratamentos, se os riscos foram mitigados ou quais fatores que influenciam no tratamento. Dados quantitativos, oriundos de instrumentos de avaliação do paciente (estratificação de riscos) e qualitativos (casos clínicos e evolução dos tratamentos) necessitam estar compreendidos e preparados para permitir as análises.

A e-Science refere-se à aplicação de métodos de obtenção de resultados científicos por meio da utilização de computação, e grande volume de dados (BERTHOLD *et al.*, 2010; SCHRÖER; KRUSE; GÓMEZ, 2021). Possui um processo suportado pelas fases iniciais de compreensão e preparação de dados. A compreensão de dados visa entender se existem dados suficientes para responder a diferentes questões de pesquisa (BERTHOLD *et al.*, 2010). Deve-se estabelecer o entendimento entre dados existentes, em relação aos dados necessários para desenvolvimento da pesquisa. Por meio desta fase que o cientista de dados obtém insights sobre os dados que podem ser utilizados em etapas posteriores (BERTHOLD *et al.*, 2010; SCHRÖER; KRUSE; GÓMEZ, 2021). Inicia-se a compreensão com a coleta inicial de dados e que prossegue com atividades para se familiarizar com os dados, identificar problemas de qualidade e detectar subconjuntos de interesse para pesquisa científica a fim de formar hipóteses (DANUBIANU; PENTIUC; TOBOLCEA, 2010). De acordo com Berthold *et al.* (2010) algumas questões norteiam a compreensão de dados:

- ✓ Que tipo de atributos temos e como são seus domínios?
- ✓ O domínio de um atributo é estável ao longo do tempo ou muda?
- ✓ É possível navegar pelas diferentes colunas de dados a fim de visualizá-los?
- ✓ Os dados podem ser correlacionados para extrair métricas?
- ✓ Existem valores discrepantes(outliers) que diferem dos outros dados?

- ✓ Os valores ausentes impactam no entendimento do conjunto de dados?

As questões acima podem ser respondidas a partir da execução dos seguintes métodos: compreensão de dados; qualidade de dados; visualização de dados e identificação de valores discrepantes. O uso de ontologias tem sido apresentado como mecanismo para auxiliar na compreensão de dados.

Ontologias são estruturas usadas para representar o conhecimento de um determinado domínio de aplicação (GRUBER; 1993, GUARINO, 1998). Alguns trabalhos utilizam ontologias como mecanismos para compreensão de dados, atribuindo-lhes significados (semântica). Com base em uma discussão sobre a hierarquia de compreensão de dados, Zhou *et al.* (2010) analisam a integração de informações sob uma perspectiva semântica. A gestão de dados suportada por ontologias oferece um caminho promissor para resolver problemas de compreensão, localização, reutilização e integração de dados e a qualidade das informações geradas (FUJIMOTO; CANEDO, 2018). Na fase de compreensão de dados, uma forma de tornar explícito o conhecimento da ontologia, é o mapeamento de dados para ontologia (SVÁTEK; RAUCH; RALBOVSKÝ, 2005). Em Brisson e Collard (2008) os autores modelam o conhecimento de domínio durante as etapas de compreensão de domínio e compreensão de dados para construir um sistema de informação orientado a ontologia (ODIS, An Ontology Driven Information System). De acordo com os autores, a compreensão dos dados significa a seleção e descrição dos dados de origem para capturar sua semântica e confiabilidade por meio de ontologias.

De acordo com Paton (2019) os cientistas de dados gastam tempos excessivos preparando dados. A partir da compreensão dos dados, a preparação de dados abrange a seleção, limpeza, mapeamento e integração de conjuntos de dados existentes para uma organização de dados adequada para modelagem e análise (BERTHOLD *et al.*, 2010; PATON, 2019). Dados devem estar preparados

adequadamente para suportar as técnicas de modelagem evitando análises enviesadas (PYLE, 1999).

Diferentes informações podem ser extraídas da estratificação e das evoluções dos tratamentos, com o objetivo de fundamentar as decisões sobre como avaliar os sinais e sintomas do paciente e quais dados são usados para atribuir o nível de risco. A ausência da definição e organização conceitual dos diferentes dados pode gerar interpretações ambíguas sobre a estratificação de risco e, como consequência, a elaboração equivocada da assistência à saúde do paciente.

Os dados coletados pela aplicação de questionários podem ser usados para apoiar as equipes multiprofissionais a responder questionamentos com base em diferentes tipos de estudos, a fim de propor tratamentos que possam melhorar a saúde mental do paciente. No entanto, a estrutura dos dados gerados no formato tabular carece de modelagem conceitual que explicita suas propriedades e relações, e não favorece a integração de diferentes conjuntos de dados que podem ser formados a partir não somente da própria estratificação, mas também de conjuntos de dados externos.

A partir do risco estratificado, são estabelecidas intervenções por meio de sessões de com o uso da TCC. Entende-se que o rastreamento do transtorno mental de ansiedade não deve utilizar somente instrumentos, tal como a estratificação de risco, para geração de dados quantitativos e métricas do nível de risco da saúde do paciente. Uma estratégia importante para rastreamento na atenção primária, é considerar os dados das evoluções (anotações). As evoluções podem constituir um conjunto de dados que servem como base para rastrear se os tratamentos estão sendo efetivos com objetivo de orientar a melhoria do estado de saúde dos pacientes, e, conseqüentemente, mitigar o risco estratificado. Os dados produzidos na estratificação são considerados como dados quantitativos, pois geram uma nota (final score) para o paciente com base em um conjunto de perguntas e a pontuação atribuída a cada uma. A análise qualitativa do caso clínico pode ser realizada usando os textos das evoluções, a fim de gerar dados qualitativos. Ambos os tipos de dados



sendo compreendidos, preparados e integrados em uma mesma base de dados abrem caminho para analisar diferentes variáveis para estudos científicos ou até mesmo suportar a descoberta de novos padrões de tratamentos psicoterápicos que obtiveram sucesso na mitigação dos riscos em ansiedade para diferentes grupos de pacientes.

Este texto refere-se ao projeto de pesquisa de Doutorado intitulado “Processo sistemático fundamentado em modelagem ontológica aplicado à estratificação de risco em saúde mental”, realizado no Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais (PPGGOC). O projeto tem como objetivo compreender dados estruturados identificados a partir estratificação de risco em ansiedade, juntamente com dados textuais (não estruturados) de tratamentos psicoterápicos a fim de prepará-los para análise qualitativa e quantitativa. A pesquisa aborda o uso de ontologias para anotar semanticamente os dados, e de tecnologias que permitam gerar grafos de conhecimento. Como resultado, espera-se que especialistas de domínio possam navegar pelos grafos para realizar análise quali-quantitativa e inferir novos conhecimentos sobre a eficiência dos tratamentos na mitigação dos riscos em ansiedade.

Este projeto é realizado em parceria com o Centro Universitário UniAcademia situado em Juiz de Fora, Minas Gerais. O UniAcademia dispõe de infraestrutura física e pode fornecer apoio técnico administrativo para execução do projeto. Desde 1972, o UniAcademia possui uma Clínica Escola de Psicologia com estrutura física e suporte de professores do curso de graduação em Psicologia que coordenam atividades de atendimento gratuito a dezenas de pacientes de Juiz de Fora e região. Alunos dos cursos de graduação em Psicologia e Engenharia de Software participam da pesquisa em estágio avançado.

Referências



BRISSON, Laurent; COLLARD, Martine. An ontology driven data mining process. **International Conference on Enterprise Information Systems**, Barcelona, jun. 2008. p. 54-61. Disponível em: <https://hal.archives-ouvertes.fr/ird-00842979>. Acesso em: 25 nov. 2022.

BERTHOLD, M. R.; BORGELT, C.; HÖPPNER, F.; KLAWONN, F. **Guide to intelligent data analysis: how to intelligently make sense of real data**. London: Springer, 2010. Disponível em: <https://link.springer.com/book/10.1007/978-1-84882-260-3>. Acesso em: 25 nov. 2022.

BRASIL. Ministério da Saúde. **Portaria Nº 3.088, de 23 de dezembro de 2011**. Institui a Rede de Atenção Psicossocial para pessoas com sofrimento ou transtorno mental e com necessidades decorrentes do uso de crack, álcool e outras drogas, no âmbito do Sistema Único de Saúde (SUS). Disponível em: http://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt3088_23_12_2011_rep.html. Acesso em 14 de abr de 2021.

BRASIL. Ministério da Saúde. Departamento de Atenção Básica. **Cadernos de Atenção Básica: Núcleo de Apoio à Saúde da Família - Volume 1: ferramentas para a gestão e para o trabalho cotidiano**. Brasília: Ministério da Saúde; 2014. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/nucleo_apoio_saude_familia_cab39.pdf. Acesso em 14 de abr de 2021.

DATTANI S., HANNAH, R.; ROSE, M. **Mental Health**. Published online at OurWorldInData.org, 2021. Disponível em: <https://ourworldindata.org/mental-health>. Acesso em 14 de abr de 2021.

DANUBIANU, Mirela; PENTIUC, Ștefan Gheorghe; TOBOLCEA, Iolanda. **CRISP-DM Model Applied for Knowledge Discovery in Speech Disorders Therapy Area**. 2010. Disponível em: <https://fiesc.usv.ro/wp-content/uploads/sites/17/2021/01/Danubianu1.pdf>. Acesso em 02 de junho de 2022.

FUJIMOTO, Márcia Takenaka; CANEDO, Edna Dias. Ontology based data management: a study in a Brazilian Federal Agency. *In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT*, 17., 2018, Krems, Austria. **Proceeding**...Krems: Springer International, 2018. p. 144-154. <https://hal.inria.fr/hal-01961528/document>. Acesso em 02 de junho de 2022.

GUARINO, Nicola (Ed.). Formal ontology in information systems. **Proceedings of the first international conference (FOIS'98)**, June 6-8, Trento, Italy. IOS press, 1998. Disponível em: <https://dl.acm.org/doi/10.5555/521669>. Acesso em 02 de junho de 2022.



GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. 2, p. 199-220, 1993. Disponível em: <https://doi.org/10.1006/knac.1993.1008>. Acesso em 02 de junho de 2022.

MAYO CLINIC. **Anxiety disorders** - Symptoms and causes. Disponível em: <https://www.mayoclinic.org/diseases-conditions/anxiety/symptoms-causes/syc-20350961>. Acesso em 14 de jul de 2022.

MELO, A. L. C.; LOURENÇO, L. M.; Terapia cognitivo-comportamental no tratamento de um caso de transtorno de ansiedade generalizada: um relato de caso. **Psicologia.pt**, 1º jun. 2020. Disponível em: <http://www.psicologia.pt/artigos/textos/A1408.pdf>. Acesso em 02 de junho de 2022.

MELLO, Veronica de Pádua. **Caminhos da educação em saúde na atenção básica**: proposta de reorganização do grupo do parque. 2015. 90p. Dissertação (Mestrado Profissional em Enfermagem na Atenção Primária em Saúde no Sistema Único de Saúde) - Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://teses.usp.br/teses/disponiveis/7/7144/tde-10052017-110135/fr.php>. Acesso em 02 de junho de 2022.

PAULA, George Luiz Costa de. **Classificação de risco em saúde mental**: implicações clínicas, éticas e sócio-políticas. 2019.

PATON, Norman. Automating data preparation: can we? should we? must we?. In: 21st International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data. 2019. Disponível em: <https://ceur-ws.org/Vol-2324/Paper00-InvTalk2-NPaton.pdf>. Acesso em 02 de junho de 2022.

PYLE, Dorian. **Data preparation for data mining**. San Francisco: Morgan Kaufmann, 1999.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A systematic literature review on applying CRISP-DM process model. **Procedia Computer Science**, 181, 526-534. 2021. Disponível em: <https://doi.org/10.1016/j.procs.2021.01.199>. Acesso em 02 de junho de 2022.

SIQUELLI, S. A.; HAYASHI, M. C. P. I. Ética em pesquisa de educação: uma leitura a partir da resolução 196/96 com expectativas da resolução 466/12. **Revista História & Perspectivas**, Uberlândia, v. 52, 2015. Disponível em: <https://seer.ufu.br/index.php/historiaperspectivas/article/view/30961>. Acesso em 02 de junho de 2022.

SVÁTEK, Vojtěch; RAUCH, Jan; RALBOVSKÝ, Martin. Ontology-enhanced association mining. In: ACKERMANN, Markus (Ed.). **Semantics, Web and mining**, Berlin: Heidelberg, 2005. p. 163-179. Disponível em:



https://link.springer.com/chapter/10.1007/11908678_11. Acesso em 02 de junho de 2022.

WHITEFORD, Harvey; FERRARI, Alize; DEGENHARDT, Louisa. Global burden of disease studies: implications for mental and substance use disorders. **Health Affairs**, v. 35, n. 6, p. 1114-1120, 2016. Disponível em:

<https://www.healthaffairs.org/doi/10.1377/hlthaff.2016.0082>. Acesso em 02 de junho de 2022.

ZHOU, Jingtao *et al.* A survey of semantic enterprise information integration. In: The 3rd International Conference on Information Sciences and Interaction Sciences. IEEE, 2010. p. 234-239. Disponível em: <https://doi.org/10.1109/ICICIS.2010.5534744>. Acesso em 02 de junho de 2022.

***Dados biográficos dos autores**

Evaldo de Oliveira da Silva é docente dos cursos de Engenharia de Software e Sistemas de Informação do Centro Universitário UniAcademia – UniAcademia. E-mail: evaldosilva@ufmg.br

Marcello Peixoto Bax é docente do curso de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais. E-mail: bax.ufmg@gmail.com

Auxiliatrice Caneschi Badaró é docente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: auxiliatricebadaro@uniacademia.edu.br

Raissa Barros Dominato é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: raissabdominato@gmail.com

Carla Maria Guimarães de Assis é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: carlinhamariadeassis@hotmail.com

Manuela Mayworm Jens é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: manuelajens@hotmail.com

Eduarda Novais Falcão é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: eduardanfalcão@gmail.com

Rubens Amorim Souto Neto é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: rubensasn@gmail.com

João Vitor Carvalho é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: joaovitorcarvalhov2@outlook.com



João Gabriel Garcia Paiva é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: joaogarciapaiva@hotmail.com

Samuel Henrique Reis do Nascimento é discente do curso de Psicologia do Centro Universitário UniAcademia – UniAcademia. E-mail: samuelhr.900021713@uniacademia.edu.br

Como citar

SILVA, Evaldo de Oliveira da *et al.* Compreensão e preparo de dados aplicados na estratificação de risco em ansiedade para análise quanti. **Ciência da Informação Express**, Lavras, v. 3, 28 nov. 2022.