

Reconhecimento de Padrões em Atendimentos do SUS no RS

Carina R. S. Dorneles^{1,2}, Karin Becker¹, João L. D. Comba¹, Carla M.D.S Freitas¹

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Av. Bento Gonçalves, 9500 – 91.501-970 – Porto Alegre – RS – Brazil

²Secretaria da Saúde – Av. Borges de Medeiros, 1501 – Porto Alegre – RS – Brazil

{crsdorneles, karin.becker, comba, carla}@inf.ufrgs.br

Abstract. *The scarcity of financial and technological resources poses significant challenges for managers in optimizing the efficient allocation of public health resources. Also, the large volume of data makes resource management and monitoring more complex. Aiming to contribute to technological innovation in public health, we propose using artificial intelligence techniques for pattern recognition and anomaly detection in the SUS ASSISTIR program.*

Resumo. *A escassez de recursos financeiros e tecnológicos desafia os gestores na otimização da gestão dos recursos da saúde pública. O grande volume de dados torna complexo gerir e monitorar esses recursos. Contribuindo para a inovação tecnológica na saúde pública no Rio Grande do Sul (RS, Brasil), este trabalho propõe o uso de técnicas de inteligência artificial para reconhecer padrões e detectar anomalias nos atendimentos SUS do programa ASSISTIR.*

1. Introdução

A gestão de recursos da saúde é um problema complexo e influenciado por diversos fatores. Na gestão do Sistema Único de Saúde (SUS), a liberação dos escassos recursos financeiros para contratação de serviços de saúde e de recursos humanos não é eficiente, em função da defasagem tecnológica das ferramentas de gestão e monitoramento.

A Secretaria Estadual da Saúde do Rio Grande do Sul (SES/RS) é responsável pela gestão, avaliação e monitoramento dos serviços de saúde, dentre outras atribuições de regulação do SUS. Administrativamente, o SUS no RS está distribuído em 7 macrorregiões, subdivididas em 30 regiões de saúde. Para melhorar a qualidade dos atendimentos de saúde à população, foi criado o programa de incentivo hospitalar ASSISTIR (Portaria nº 688/2024), contemplando ambulatórios que atendem diferentes especialidades.

O monitoramento dos dados de prestação de serviços, além de evidenciar se os serviços contratados estão sendo realizados, subsidia a formulação de novas políticas públicas e investimentos. A análise manual de dados é um processo ineficiente e oneroso que se potencializa quando se trata de grandes volumes de dados gerados por fontes heterogêneas, o que é o caso dos dados da saúde gerados pelo Datasus. Apesar desses dados estarem disponíveis para análise, a unificação das bases de dados para extração de informações customizadas ainda é um processo difícil e demorado, o que abre espaço para a adoção de métodos de inteligência artificial (IA) [Mellouli et al. 2024].

Como uma prova de conceito, no presente estudo, utilizamos técnicas de aprendizado de máquina visando identificar de forma consistente padrões e possíveis anomalias nos dados de atendimentos de quatro especialidades elencadas como prioritárias pela

SES/RS, a saber, Traumatologia/Ortopedia, Oftalmologia, Cirurgia Geral e Urologia. Buscamos descobrir: as características de desempenho das especialidades prioritárias do ASSISTIR (Q1); se há alguma macrorregião ou região de saúde do RS sem cobertura de alguma das especialidades prioritárias do ASSISTIR (Q2); e se há algum padrão que possa indicar possível anomalia nos atendimentos (Q3).

2. Trabalhos Relacionados

Estudo recente [Pasin 2023] analisou políticas públicas da COVID-19 nos países mais populosos, utilizando *K-Prototype* e métricas como o método Elbow. Gao et al. (2003) também aplicaram *K-Prototype* para identificar padrões de necessidades de saúde da população idosa, destacando limitações na definição do número de clusters e no ajuste dos pesos de atributos. Em outro estudo recente, Du e Yu (2023) propuseram um modelo de detecção de anomalias em seguros médicos com *Isolation Forest* (IF) [Liu et al. 2008] e diferentes algoritmos de agrupamento como *K-Means*, DBSCAN [Ester et al. 1996] e *Hierarchical Clustering* [Nielsen 2016]. O método HDBSCAN foi sugerido como alternativa mais robusta ao DBSCAN para lidar com ruídos [Campello et al. 2013]. Por fim, outros autores exploraram a distribuição de recursos hospitalares no Peru com *K-Means*, LDA e DBSCAN, evidenciando desigualdades regionais [Aguirre et al. 2024].

Com base nessas experiências, este estudo propõe o uso do *K-Prototype* para dados mistos e uma abordagem híbrida de detecção de anomalias com IF e HDBSCAN, para o monitoramento dos atendimentos de especialidades prioritárias no RS.

3. Materiais e Métodos

Foram realizadas reuniões com especialistas da SES/RS visando entender as necessidades de gestão e monitoramento e as regras da normativa vigente. Optou-se por considerar as especialidades prioritárias do programa ASSISTIR que possuíam registro de produção SUS do financiamento de Média e Alta Complexidade (MAC) de 2022 a 2024.

A análise exploratória dos dados permitiu verificar a distribuição dos atendimentos registrados nas especialidades médicas e observar desbalanceamento no número de atendimentos, destacando possíveis ruídos. O *dataset* possuía 41.538 observações, com 10 atributos, sem dados nulos. A técnica de agrupamento *K-prototype* foi utilizada para o reconhecimento de padrões de atendimentos por especialidade, o que demandou o agrupamento com variáveis mistas. A preparação dos dados para execução do agrupamento incluiu: a seleção dos 3 atributos principais (especialidade, porte do hospital e número de atendimentos), a normalização da variável numérica utilizando a função *MinMaxScaler* da biblioteca Sklearn e a determinação do número ideal de partições para o agrupamento, através do método *Elbow*, o qual sugeriu duas a três partições, sendo 3 o número adotado. O agrupamento foi avaliado com a métrica *Silhouette*, que resultou no coeficiente de 0,79 e também avaliado com o índice de *Davies-Bouldin*, que resultou em 0,44. Adotamos uma abordagem híbrida para detecção de anomalias, combinando os algoritmos IF e HDBSCAN, utilizando os hiperparâmetros padrão de cada algoritmo. Na aplicação do IF, os registros com escores menores que -0,3 apresentaram redução na densidade dos dados (visualizados através de histograma) e foram considerados como anômalos, correspondendo a 0,6% dos registros analisados. Com a aplicação do HDBSCAN, constatamos 14% de registros anômalos. Combinando os resultados positivos para anomalias obtidas em ambas as técnicas, o resultado final foi de 0,17% de registros anômalos.

Para a exibição dos resultados foi desenvolvido um *dashboard* com Streamlit.

4. Resultados

O agrupamento foi realizado considerando atendimentos por especialidade e o porte dos estabelecimentos e obtemos o número ideal de 3 grupos, os quais denominamos Grupo 1, Grupo 2 e Grupo 3. O *dashboard* analítico permitiu responder nossas questões norteadoras.

Q1: Quais as características de desempenho das especialidades prioritárias do ASSISTIR? Ao explorar as características dos grupos formados, sem aplicação de outros filtros, foi possível observar que os grupos apresentam padrões temporais de atendimento distintos. O Grupo 1 não teve registro de atendimentos no ano de 2022 em nenhuma das especialidades prioritárias, mas em 2023, teve uma concentração de atendimentos em Oftalmologia. Já em 2024, houve uma redução dos atendimentos nessa especialidade, com um número maior de atendimentos em Traumatismo/Ortopedia. Para o Grupo 2, observamos que todas as especialidades prioritárias foram atendidas entre 2022 e 2024 em todas as macrorregiões. No entanto, os dados revelaram um declínio nos atendimentos de Oftalmologia, com uma redução atípica nessa especialidade no ano 2024. Além disso, enquanto as demais especialidades registraram o maior número de atendimentos em 2023, destaca-se uma queda significativa nos atendimentos de Traumatismo/Ortopedia em 2024. Já no Grupo 3, foram localizados registros somente a partir do ano de 2024, os quais se referiam exclusivamente às consultas de Oftalmologia.

Q2: Há alguma macrorregião ou região de saúde do RS sem cobertura de alguma das especialidades prioritárias do ASSISTIR? Em geral, por possuírem ambulatórios enquadrados nos três grupos, todas as macrorregiões de saúde possuem atendimentos nas quatro especialidades prioritárias. No entanto, ao observar a cobertura de ambulatórios por região de saúde, identificamos que 50% das regiões não possuem algum ambulatório prioritário habilitado ou não registraram atendimentos. A ausência de ambulatórios nessas regiões pode ser vista como um sinalizador para orientar a abertura de novos serviços de saúde para a população considerando o aporte de novos recursos.

Q3: Há algum padrão que possa indicar possível anomalia? O padrão anômalo identificado foi o número de atendimentos acima da média. Além disso, na especialidade de Cirurgia Geral foram identificados dois ambulatórios, nas macrorregiões Sul e Metropolitana, que realizam somente consultas, um ambulatório de Traumatismo/Ortopedia, na macrorregião Metropolitana que realizou somente cirurgias e dois ambulatórios de Urologia na macrorregião Norte que realizam somente cirurgias. Esse comportamento não é comum, pois, uma vez habilitado, cada estabelecimento se compromete a realizar consultas e cirurgias, não com apenas um dos tipos de serviço. Também foram observados outros padrões anômalos, talvez decorrentes da falta de registros nos sistemas ou pela falta de recursos em anos anteriores, que pode ter proporcionado, de forma gradual, novos investimentos a partir de 2024.

5. Conclusões

Este trabalho concentrou-se nos ambulatórios de especialidades prioritárias do programa ASSISTIR e envolveu a consolidação de múltiplas fontes de dados em um único *dataset*. Além disso, foram implementadas funções de pré-processamento para aprimorar a qualidade dos dados, permitindo uma análise mais precisa dos padrões de atendimento por especialidade, grupos de estabelecimentos e volume de atendimentos.

Os resultados obtidos demonstraram que algumas especialidades, em determinadas macrorregiões, possuem números de atendimentos muito acima do esperado, en-

quanto outras, muito abaixo das demais, o que pode estar relacionado a erros de registro ou necessidades de saúde locais. Contudo, cabe aos gestores avaliarem o número de registros para esclarecer se ocorreram problemas nesses registros, desequilíbrio na distribuição de recursos ou algum outro evento atípico que tenha ocasionado o aumento, redução ou não execução de atendimentos. A busca de padrões visou identificar as necessidades de serviços de saúde com base nos registros dos atendimentos SUS. Por outro lado, a detecção de anomalias, tal como a aqui proposta, tem o papel de sinalizar registros que precisam de supervisão e serve de subsídio para a equipe técnica da SES/RS no entendimento e validação dos padrões encontrados.

O *dashboard* analítico desenvolvido permitiu uma análise ágil e facilitou a interpretação dos dados, mas ainda deverá ser avaliado quanto à usabilidade. Após essa avaliação e eventuais melhoramentos, o presente trabalho tem potencial para se tornar uma ferramenta estratégica para análise e monitoramento dos atendimentos de saúde, contribuindo para a gestão otimizada dos recursos do SUS no RS.

Agradecimentos

Agradecemos as contribuições dos especialistas da Secretaria da Saúde do RS. O presente trabalho é apoiado pela CAPES (Finance Code 001) e pela FAPERGS (RITE CIARS, processo 22/2551-0000390-7).

Referências

- Aguirre, F., Gonzales, H., Zegarra, A., and Espezúa, S. (2024). Retrospective analysis of hospitalization patterns in peru: Insights on inpatient resource utilization. In *2024 XXXI International Conference on Electronics, Electrical Engineering and Computing*, pages 1–7. IEEE.
- Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In Pei, J., Tseng, V. S., Cao, L., Motoda, H., and Xu, G., editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin. Springer.
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231.
- Liu, F. T., Ting, K. M., and Zhou, Z. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE.
- Mellouli, S., Janssen, M., and Ojo, A. (2024). Introduction to the issue on artificial intelligence in the public sector: Risks and benefits of AI for governments. *Digit. Gov. Res. Pract.*, 5(1):1–6.
- Nielsen, F. (2016). Hierarchical clustering. In *Introduction to HPC with MPI for Data Science*, pages 195–211. Springer International Publishing, Cham.
- Pasin, O. e Gonenc, S. (2023). An investigation into epidemiological situations of COVID-19 with fuzzy k-means and k-prototype clustering methods. *Sci. Rep.*, 13(1):6255.