

# Uma abordagem de descoberta de conhecimento para suporte à gestão municipal de saúde

Lilian F. Petroski<sup>1</sup>, Samarone J. Cardoso<sup>1</sup> and Marcelo Teixeira<sup>1</sup>

Universidade Tecnológica Federal do Paraná - UTFPR

`lilianpetroski@hotmail.com`, `samarone.cardoso@gmail.com`, `mtex@utfpr.edu.br`

**Abstract.** No âmbito da administração pública municipal, existe uma ampla gama de ferramentas de computação para subsidiar as operações transacionais, no entanto, há uma carência de sistemas dedicados a subsidiar o processo de tomada de decisão. Uma das formas de extrair conhecimento a partir de dados é por meio do processo *Knowledge Discovery in Databases* (KDD), que transforma dados brutos em informações úteis, válidas e interpretáveis por humanos. Este trabalho explora o processo KDD sobre a base de dados do sistema E-saúde, com o propósito de detectar o perfil de atendimento médico nas unidades municipais de saúde de Curitiba, para subsidiar os gestores na tomada de decisão. Foi utilizada, para a etapa de mineração, a ferramenta WEKA combinada com o algoritmo Apriori, o que permite identificar regras de associação a partir de relações não aparentes na base de dados. Ao todo, foram mineradas 34 regras, de relevância estatística, que trazem informações sobre a distribuição de recursos humanos no atendimento primário à saúde, organização dos atendimentos em cada unidade, bem como quais especialidades médicas seriam mais adequadas para cada unidade de acordo com a faixa etária dos pacientes e perfil de atendimento para proporcionar mais qualidade no atendimento à população.

**Keywords:** KDD, mineração de dados, gestão municipal

## 1 Introdução

As ferramentas voltadas às políticas públicas que permitem a população e aos gestores avaliarem a eficiência e eficácia dos serviços são cada vez mais relevantes para a sociedade. Atualmente, em âmbito municipal, percebe-se um grande volume de dados sendo gerados e armazenados em sistemas de gerenciamentos de banco de dados e, com esse crescimento, reduz-se as possibilidades das pessoas entendê-los sem ferramentas adequadas. Alguns sistemas são utilizados como tabuladores de informação e observa-se uma carência de sistemas dedicados a fornecer informações estratégicas à gestão. Uma das áreas mais afetadas por isso é a da saúde, que atua em milhares de atendimentos clínicos, laboratoriais e farmacêuticos, com potencial extraordinário para tomada de decisão. Em vista disso, soluções computacionais são essenciais para automatizar a pesquisa de padrões e a descoberta de informações baseadas em dados [15].

No escopo do aprendizado de máquina, uma forma para extrair conhecimento de bases de dados é por meio de um processo conhecido como *Knowledge Discovery in Databases* (KDD) [3, 8, 10]. O KDD se refere ao processo de descoberta de conhecimento em base de dados, sendo uma forma não usual de identificar padrões novos, válidos e compreensíveis. E ainda, permite a exploração de uma grande quantidade de dados [4].

Na literatura, [3, 7, 11] utilizam o KDD para identificar oportunidades para redução de custos em operadoras de plano de saúde; [5, 14] aplicam o KDD em uma base de dados de atendimentos municipais prestados pelo *Sistema Único de Saúde* (SUS) para analisar as possíveis causas e correlações de doenças; [9] exploram a triagem dos pacientes atendidos em *Unidades de Pronto Atendimento* (UPAs) do SUS; e [13] avaliam a efetividade dos atendimentos. A maioria dos trabalhos relacionados são voltados ao auxílio de profissionais em saúde, mas poucos se dedicam a fornecer suporte aos gestores públicos na destinação de recursos ou na criação de ações estratégicas para melhorar o atendimento à população.

Este trabalho explora o processo KDD sobre a base de dados do sistema E-saúde, relativo ao perfil de atendimento médico nas unidades municipais de saúde. Esse sistema registra atendimentos na rede de atenção à saúde do município de Curitiba - Paraná, que é composta por Unidades Básicas de Saúde, Unidades de Pronto Atendimento e Centros de Especialidades Médicas e Odontológicas. A base de dados é disponibilizada no portal de dados abertos do município [12] e foi utilizada a versão do primeiro trimestre de 2020, que possui aproximadamente 900 mil atendimentos.

O processo KDD conduzido visa extrair conhecimento relacionado aos atendimentos, especialização do médico, perfil dos pacientes e zoneamento. Para tal, ele utiliza o pipeline do python na etapa de pré-processamento dos dados, o PostgreSQL para armazenamento, e a ferramenta *Waikato Environment for Knowledge Analysis* (WEKA) para a etapa de mineração, com aplicação do algoritmo Apriori, que obtém regras de associação em conjuntos de dados com relações não aparentes. Esse algoritmo trabalha com um grande número de atributos do tipo nominal, sendo adequado para a base explorada, que possui a maioria dos atributos desse tipo. Por fim, realiza-se a interpretação dos resultados obtidos.

## 2 Aplicação do KDD na Base de dados do E-saúde

A gestão pública de saúde, em particular a gestão municipal, planeja, executa, controla e avalia ações dos serviços de saúde. Cabe à esta pasta, em geral, a maior fatia de um orçamento municipal, haja visto que os atendimentos primários aos cidadãos são de incumbência do município [2].

### 2.1 Base de dados do E-saúde

O sistema E-saúde é utilizado para o registro dos atendimentos da rede de atenção à saúde de Curitiba. A base de dados é disponibilizada no portal de

dados abertos, em formato separado por ponto e vírgula (.csv), e para este trabalho foram selecionados os dados do perfil de atendimento médico nas unidades municipais de saúde do primeiro trimestre de 2020, que contém 898.849 linhas e tamanho de 360 MB, com 42 atributos, contendo informações de atendimentos, dados dos pacientes, de internação e da farmácia Curitibana [12].

Dentre os 42 atributos da base em primeira análise, os mais importantes são os que contém informações dos atendimentos e dos pacientes. Em relação ao atendimento os atributos são: *tipo de unidade*, *descrição do procedimento*, *descrição do CBO*, *descrição do CID*, *solicitação de exames* e *desencadeou internamento*, que em conjunto informam características e padrões nos atendimentos. Já quanto aos pacientes os atributos são: *sexo*, *bairro* e *data de nascimento*, que contém aspectos importantes do perfil dos pacientes atendidos.

## 2.2 Preparação ou limpeza dos dados

O pré-processamento, ou limpeza, dos dados é uma das etapas mais importantes do KDD. Nela, são realizadas análises preliminares e tratamentos de dados inconsistentes, inválidos ou redundantes, facilitando a etapa de processamento [1, 14]. Na exploração inicial dos dados do E-saúde, utilizou-se o *python* para realizar análises preliminares, verificar o tipo de dado de cada atributo e procurar a existências de dados nulos ou inválidos.

Morfologicamente, dos 42 atributos da tabela, 30 são do tipo texto, 3 são tipo data e 9 tem tipo numérico e são aderentes aos tipos que constam no dicionário de dados. Foram identificados 17 atributos com ocorrência de valores nulos. Para tratamento dos atributos *descrição do CID* e *código do CID* as 165 linhas com valores nulos foram removidas, por se tratar de poucas instâncias em relação ao total da base.

Para os atributos *Tratamento no Domicílio*, *Abastecimento*, *Tipo de Habitação*, *Destino Lixo*, *Fezes/Urina*, *Cômodos*, *Em Caso de Doença*, *Grupo Comunitário*, *Meio de Comunicação* e *Meio de Transporte*, os valores nulos foram substituídos pela moda, que é o valor que mais aparece na coluna, o que evita distorções ou inconsistências de processamento.

Já os atributos referente à internações, que são *Data do Internamento*, *Estabelecimento Solicitante*, *Estabelecimento Destino*, *CID do Internamento* e *Área de Atuação*, só são preenchidos quando ocorre internamento e, por isso, apresentam uma grande quantidade de linhas nulas correspondente a mais de 90% do total da base. Por isso, esses atributos foram descartados.

Após a limpeza dos dados, esses foram carregados em um banco de dados *PostgresSQL*, o que demorou cerca de 15 minutos, utilizando uma máquina virtual com Ubuntu 20.04 e com 6 GB de memória RAM.

## 2.3 Transformação dos dados

Na etapa de transformação dos dados, deve-se observar qual o tipo de dado mais adequado ao algoritmo a ser aplicado na etapa de mineração. Os atributos *origem do usuário* e *residente*, que tinham como dado o código utilizado

no banco de dados original, foram substituídos pela descrição correspondente, para possibilitar a aplicação do algoritmo Apriori para obtenção das regras de associação.

Como o município de Curitiba subdivide seu território em distritos sanitários, foi criado um novo atributo para este dado a partir do bairro do paciente. Também foi criado o atributo de idade, calculado a partir da data de nascimento. A partir da idade foi criado o atributo de faixa etária. Os novos atributos foram incluídos no *Postgres* por meio de comandos em SQL, executados diretamente no *pgAdmin4*.

#### 2.4 Mineração de dados da Base do E-Saúde e interpretação dos resultados

No WEKA, versão 3.8.5, para importação dos dados a serem minerados, foi utilizada a opção de conexão com o banco de dados. Realizada consulta importando os atributos selecionados para aplicação do algoritmo Apriori, que trabalha com tipo de dado nominal. Por isso, não foram utilizados os atributos com dados do tipo data e numérico.

A aplicação do Apriori utilizando 29 atributos, com os parâmetros de confiança mínima de 90% e de suporte mínimo em 10% e número de regras até 150, resultou em 150 regras após percorrer os dados por 2 ciclos [6]. Com o resultado obtido verifica-se a recorrência de regras utilizando os atributos de *nacionalidade*, *município* e *origem\_do\_usuario*. Porém, como existe uma baixa variação de valores nesses atributos, eles não geram conhecimento relevante.

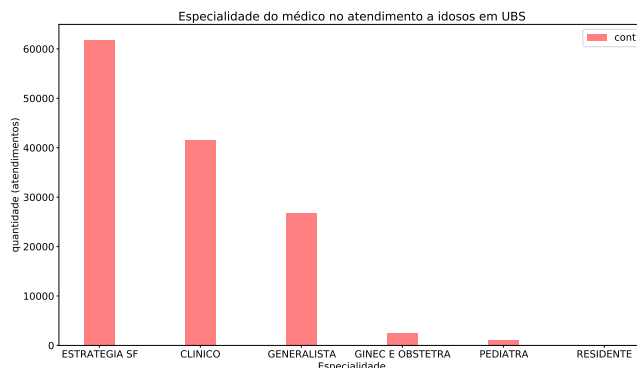
Os atributos de *código\_do\_usuario* e *código\_do\_profissional* apresentam uma grande quantidade de registros exclusivos e, dessa forma, pouco contribuiriam para as regras de associação, sendo então removidos do processo de mineração.

Foram realizados mais 2 processamentos de mineração [6] utilizando os mesmos parâmetros, porém, removendo atributos que resultaram em regras insatisfatórias. Esse tratamento foi realizado dada a grande variedade de atributos na base explorada. E finalmente, a aplicação do Apriori com os 6 atributos remanescentes referentes a dados de atendimento e do paciente, sendo eles *faixa\_etária*, *descrição\_do\_CID*, *tipo\_de\_unidade*, *descrição\_do\_procedimento*, *descrição\_do\_CBO* e *ds*. Foram utilizados os parâmetros de confiança mínima de 80% e de suporte mínimo de 10%, e um número de regras igual a 100. Foram encontradas 34 regras após percorrer os dados por 18 ciclos [6].

Dentre as regras mais representativas, a regra 7 evidencia que 133.635 consultas médicas em atenção primária para faixa etária de idosos ocorre em unidade básica com confiança de 100%. Partindo dessa regra podemos verificar qual a especialidade do médico que efetuou os atendimentos, conforme Fig. 1.

Diante da combinação da regra 7 com a *descrição\_do\_CBO*, o gestor poderia disponibilizar nas unidades de saúde médicos com especialidade geriátrica, trazendo mais qualidade ao atendimento da população idosa. No gráfico, ainda fica evidente que ocorrem atendimentos a idosos por médico com especialidade em pediatria.

Fig. 1. Gráfico especialidade do médico no atendimento a idosos

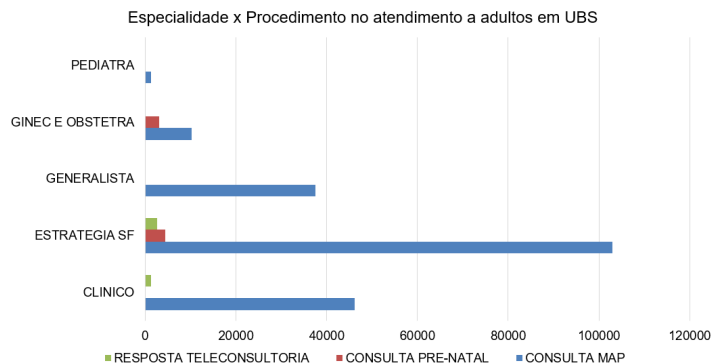


Na regra 8, constata-se que 112.243 atendimentos prestados por médico da estratégia da família para a faixa etária adulto ocorrem em unidades básicas com confiança de 100%. Em complemento, a regra 11 demonstra que 103066 atendimentos são para consulta médica de atenção primária, também com confiança de 100%. As regras 21, 22 e 23 também correlacionam atendimentos prestados por médico da estratégia da família para a faixa etária adulto, que ocorrem em unidades básicas no procedimento de consulta médica de atenção primária, mas essas com confiança de 92%.

Com base nas regras 8, 11, 21, 22 e 23, a Fig. 2 mostra um paralelo entre *descrição\_do\_CBO* versus *descrição\_do\_procedimento*. Nota-se que há atendimentos a adultos por outras especialidades e com outros procedimentos. Para melhor visualização no gráfico foram considerados especialidades com mais de 1000 atendimentos. Um ponto que chama atenção é que há médicos com especialidade de estratégia da família atendendo consulta pré-natal. Com base nessa informação, o gestor pode acompanhar se os médicos estão atuando na sua especialidade e se há necessidade de realocação de pessoal ou das atividades executadas.

Com base nas regras analisadas, verificou-se que a aplicação do KDD, com o algoritmo Apriori na etapa de mineração de dados, pode ser uma ferramenta interessante de suporte à gestão municipal de saúde. O KDD foi aplicado apenas nos dados relativos ao primeiro trimestre de 2020. Embora, com os atributos usados, tenham sido identificadas informações úteis e relevantes para a gestão, estima-se que outras tantas informações ainda sejam passíveis de descobertas via novas formas de exploração da base. Recomenda-se, ainda, que as regras obtidas sejam analisadas pelos gestores de saúde municipal e especialistas, quanto a sua qualidade, aplicabilidade e correspondência com os processos reais, a fim de que de fato possam subsidiar tomadas de decisões.

**Fig. 2.** Gráfico especialidade x procedimento no atendimento a adultos nas unidades básicas de saúde



## 2.5 CONCLUSÃO

A esfera municipal é responsável pela execução da maior parcela do orçamento destinado à saúde no Brasil. Assim, é imprescindível uma boa gestão da rede de atenção a saúde, para eficiência e eficácia na destinação de recursos e atendimento de qualidade aos cidadãos. Este trabalho contribui nesse sentido, ao possibilitar a extração de informações a partir de dados de atendimentos a pacientes.

Após percorrer as primeiras etapas do KDD, lapidando a base de dados explorada para a mineração de dados, foram utilizadas várias combinações de atributos. A partir disso, obteve-se o conjunto de seis atributos que são relevantes para os gestores avaliarem a distribuição de recursos, fazerem escolhas direcionadas as demandas buscando supri-las, além de proporcionar informações sobre o perfil do público ao qual unidade presta seus serviços.

Na interpretação dos resultados, foram avaliadas as regras mais significativas do ponto de vista de gestão. Nessa etapa, fica evidente que a abordagem encontra regras no conjunto de dados explorado, obtendo conhecimento útil que deve passar por análise humana para estabelecer a qualidade e aplicabilidade das informações. Tais regras podem ser o ponto de partida para outras análises, por exemplo, a agregação de outros atributos à regra obtida, aprofundando a obtenção de conhecimento e revelando novas informações que não estavam evidentes na regra original. Também é possível descobrir exceções às regras mineradas, que pudessem revelar informações significativas para a gestão.

## Bibliography

- [1] Camilo CO, Silva JCd (2009) Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG) 1(1):1–29
- [2] CGU (2021) Órgãos que mais aplicam recursos na área de saúde e maiores favorecidos. URL [shorturl.at/blrCT](http://shorturl.at/blrCT)
- [3] Escobar L, Dallagassa M, Sokoloski W, Tsunoda D, Carvalho D (2019) Descoberta de padrões para identificação de casos de alto custo em operadoras de planos de saúde. *Revista Stricto Sensu* 4:01–21
- [4] Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI magazine* 17(3):37–37
- [5] Feuser RJ (2017) Mineração de dados com regras de associação aplicada em dados de unidade de saúde de pronto atendimento. RIUT
- [6] de Fátima Petroski L (2021) Resultado da aplicação do apriori. URL <https://github.com/LilianPetroski/Apriori>
- [7] Gregory G, Pretto F (2016) Mineração de dados para descoberta de conhecimento em dados de promoção à saúde. *Destques Acadêmicos* 8(4)
- [8] Hirota K, Pedrycz W (1999) Fuzzy computing for data mining. *Proceedings of the IEEE* 87(9):1575–1600
- [9] Maciel T, Seus V, Machado K, Borges E (2015) Mineração de dados em triagem de risco de saúde. *Revista Brasileira de Computação Aplicada* 7(2):26–40
- [10] Mohamed EB, Ltifi H, Ayed MB (2013) Using visualization techniques in knowledge discovery process for decision making. In: 13th International Conference on Hybrid Intelligent Systems (HIS 2013), IEEE, pp 93–98
- [11] Paula LCd, Prado EFd (2012) Aplicação do processo de kdd em uma gestora de planos de saúde. *Revista da Iniciação científica da Libertas* 2:100–109
- [12] PMC (2021) Dados abertos prefeitura municipal de curitiba. URL [shorturl.at/kBEFG](http://shorturl.at/kBEFG)
- [13] Santos WH (2018) Estudo da base de dados abertos e-saúde da prefeitura de curitiba usando técnicas de mineração de dados. Master’s thesis, Universidade Tecnológica Federal do Paraná
- [14] da Silva KAM (2019) Análise de perfis de doenças com base em técnicas de descoberta de conhecimento em bases de dados. B.S. thesis, Universidade Tecnológica Federal do Paraná
- [15] Witten IH, Frank E, Hall MA (2011) *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd edn. Morgan Kaufmann, Boston