

Análise da Resolutividade das Unidades Básicas de Saúde de Um Distrito Sanitário de Curitiba Por Meio de Técnicas de Mineração de Dados

William Hamilton dos Santos ¹, Gilson Yukio Sato²

^{1 e 2} Programa de Pós-Graduação em Engenharia Biomédica (PPGEB)

Universidade Tecnológica Federal do Paraná (UTFPR)

Av. Sete de Setembro, 3165 80230-901 Curitiba - PR - Brasil

williamead@gmail.com, sato@utfpr.edu.br

Abstract. *Data related to the medical care provided to patients in the public health system of Curitiba feed the E-Health computer system. We applied the data mining in the public database of the Municipal Health Department, derived from E-Health, to study attributes that determine the resolution given by physicians working in Basic Health Units, specifically regarding to the number of exams requested and referrals for specialized care. For this, the decision tree classification algorithms and rule generators were applied to records generated in one of the sanitary districts of the city, returning results that allow units managers to monitor and determine the training and recycling needs of the professionals working in their units.*

Resumo. *Dados relacionados à assistência médica prestada a pacientes no sistema público de saúde de Curitiba alimentam o sistema de informática E-Saúde. Aplicamos a mineração de dados na base de dados pública da Secretaria Municipal de Saúde, derivada do E-Saúde, para estudar atributos que determinam a resolutividade dada por médicos atuantes em Unidades Básicas de Saúde, especificamente quanto à quantidade de exames solicitados e encaminhamentos para atendimento especializado. Para isso, os algoritmos de classificação de árvore de decisão e geradores de regras foram aplicados nos registros gerados em um dos distritos sanitários da cidade, retornando resultados que permitem que gestores das unidades monitorem e determinem as necessidades de treinamento e reciclagem dos profissionais que trabalham em suas unidades.*

1. Introdução

Considerando a saúde como um direito de todos e um dever do Estado, é função do Sistema Único de Saúde (SUS) [Brasil 2018], dispor de condições para a promoção, e recuperação da saúde do indivíduo e da comunidade, ressaltando os princípios da universalidade, integralidade e resolutividade.

Assim o acesso da população à rede de atendimento deveria se dar por meio dos serviços de nível primário da rede de atenção na Unidade Básica de Saúde (UBS) da sua

localidade, pois ela é reconhecida como porta de entrada do usuário [BRASIL 2011] ao sistema do SUS como um todo.

Desta forma, o funcionamento adequado da rede de saúde em nível primário (UBS), possibilitaria a resolutividade da maioria dos problemas, deixando os centros de especialidades e hospitais, executarem seus objetivos principais, aumentando a satisfação dos usuários, bem como proporcionando melhor uso dos recursos existentes para o sistema como um todo.

A rede de serviços do SUS de Curitiba atendeu a população de 1.848.943 habitantes [Saúde 2018a], em 2013, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE). Conta com 141 serviços próprios, dentre os quais estão 109 Unidades Básicas de Saúde e 9 Unidades de Pronto Atendimento, além de outros 23 equipamentos especializados [Saúde 2018b].

O objetivo deste trabalho é aplicar algoritmos de mineração na base de dados pública derivada do sistema E-Saúde para encontrar atributos que indiquem motivos que possam atestar a resolutividade dos atendimentos médicos das unidades (UBS), bem como avaliar a precisão dos atributos encontrados.

2. Referencial Teórico

2.1. Mineração de Dados

Com a computação sempre presente, principalmente por causa da World Wide Web, uma quantidade enorme de dados vem sendo produzida, coletada e armazenada de diversas maneiras em escalas cada vez maiores [Gabardo 2015].

Estas demandas necessitam de enfoques e ferramentas para ajudar os seres humanos na extração de informações úteis (conhecimento) desses volumes de rápido crescimento de dados computacionais. Teorias que cobrem esse assunto pertencem ao campo da Descoberta de Conhecimento [Oliveira Júnior 2015] em Bases de Dados, ou KDD.

O KDD [Fayyad et al. 1996] é um processo não trivial que busca a identificação de padrões, em dados, que sejam válidos, inéditos, potencialmente úteis e compreensíveis. Trata do processo de descoberta de um conhecimento que tenha utilidade, como um todo. Muitas vezes, o termo “Mineração de Dados” é usado de forma ampla, como sinônimo de KDD. Entende-se como mineração de dados a aplicação de algoritmos específicos para extrair padrões de dados.

2.2. Mineração de Dados Médicos

A mineração de dados aplicada à dados médicos é um processo desafiador [Ferreira 2015], por exemplo, muitas vezes, operadores preenchem os sistemas de informática de forma incompleta ou falha, tendendo a produzir bases de dados esparsas, com dados altamente variados, demandando o uso de diferentes técnicas e ferramentas para que as bases possam ser exploradas de maneira que produzam algum resultado eficaz.

Existem ainda restrições éticas, legais e sociais relativas à privacidade e a validação clínica dos achados. Assim, na maioria das vezes, os dados que realmente

chegam a ser analisadas são um subconjunto dos dados originais. Contudo, cada dia mais, a mineração de dados vem ganhando destaque na área da saúde. Auxiliando as organizações de saúde a tomar decisões de gerenciamento em relação aos utilizadores e seus colaboradores [Ferreira 2015], como os profissionais de enfermagem e médicos.

2.3. Classificadores Utilizados

Classificação pode ser entendida como a distribuição sistemática em diversas categorias segundo analogias e características comuns [Librelotto e Mozzaquatro 2013]. Os algoritmos classificadores utilizados neste trabalho foram escolhidos em função das características dos dados de saída de cada um, por isso foram empregados um classificador de árvore de decisão e um classificador baseado em regras.

Uma árvore de decisão é um conjunto de condições dispostas em uma estrutura hierárquica. É um modelo em que um exemplo é classificado seguindo o caminho de condições satisfeitas a partir da raiz até atingir uma folha, que vai corresponder a um rótulo de classe [Romero et al. 2008], por isso as árvores de decisão são consideradas modelos simples e de fácil compreensão.

Já em um classificador baseado em regras, o modelo aprendido é exibido por meio de um conjunto de regras do tipo “se-então”. As regras são uma boa maneira de representar informação ou pedaços de conhecimento [Han et al. 2011].

3. Materiais e Métodos

3.1. Portal de Dados Abertos

A Prefeitura de Curitiba disponibiliza, por meio de seu portal na internet [Portal 2018], documentos, informações e dados governamentais de domínio público para a livre utilização, garantindo acesso a dados primários que possam ser reutilizados produzindo novas informações e aplicações para a sociedade.

Para esse trabalho foi selecionada a base que contém os dados disponibilizados referentes ao perfil de atendimento dos profissionais médicos das Unidades Básicas de Saúde. Essa base de dados constitui-se em um arquivo no formato .csv que pode ser baixado diretamente do portal de dados abertos.

Por possuir dados de toda rede de atendimento referente o primeiro trimestre de 2017, o arquivo obtido possui tamanho de 280 Mb, totalizando 730.738 registros. Junto com ele, na mesma página é disponibilizado outro arquivo que é o dicionário de dados da base.

Este dicionário contém os campos e a descrição de cada um, bem como outros detalhes técnicos. Na base de dados `medicos.csv`, são disponibilizados 37 campos para cada instância (ou registro).

3.2. Banco de Dados Para a Armazenagem dos Registros

Foi utilizado o banco de dados MySQL [MySQL 2018] para a criação de uma tabela chamada MEDICO contendo os 37 atributos originais da base de dados, de acordo com a descrição do dicionário de dados. Em seguida foram criados mais 10 novos atributos na tabela MEDICO e populados com valores calculados ou inferidos dos campos originais. O passo seguinte foi fazer uma seleção dos atributos que tem ligação com o

tema do trabalho, reduzindo a dimensionalidade para um total de 26, conforme mostrados na tabela 1. Os atributos em caixa baixa são atributos originais da base, já os em caixa alta foram criados.

Tabela 1: Atributos utilizados nas análises dos dados

Código	Nome do Atributo da Tabela MEDICO	Descrição
1	DIA_DA_SEMANA	Dia da Semana do Atendimento
2	IDADE_FAIXA	Faixa da Idade do Paciente
3	sexo	Sexo do Paciente
4	descr_unidade	Descrição da Unidade de Atendimento
5	descr_procedimento	Descrição do Procedimento Realizado
6	descr_CBO	Descrição da Ocupação do Profissional
7	cod_CID	Código do Diagnóstico
8	CAP_CID	Capítulo do CID do Diagnóstico
9	solicit_exam	Indica se ocorreu solicitação de Exames
10	FARM_PRESCR	Prescrição de medicamentos na Farmácia Curitiba
11	FARM_DISPEN	Dispensação de medicamentos na Farmácia Curitiba
12	FARM_NAO_PADRON	Medicamento Não Padronizado
13	enc_atend_especia	Indica se houve encaminhamento para atendimento de Especialista
14	desencadeou_interna	Indica se desencadeou Internamento
15	tratam_domic	Tipo de Tratamento de Água no domicílio
16	abastecimento_agua	Tipo de Abastecimento de Água no domicílio
17	energia_eletrica	Indica se há energia elétrica no domicílio
18	tipo_habitacao	Tipo de habitação no domicílio
19	destino_Lixo	Destino do lixo no domicílio
20	fezes_urina	Destino das fezes/urina no domicílio
21	COMODOS_FAIXA	Faixa de Cômodos da casa do paciente
22	em_caso_doenca	Serviços procurados em caso de doença
23	grupo_comunitario	Grupo Comunitário em que o paciente participa
24	meio_comunicacao	Meios de Comunicação utilizados no domicílio
25	meio_transporte	Meios de Transporte utilizados no domicílio
26	ORIGEM_CURITIBA	Paciente mora em Curitiba

Os atributos adicionais foram criados para otimizar a aplicação dos algoritmos classificadores. Por exemplo, foi criado o atributo `CAP_CID`, que agrupou em 21 valores as 5000 ocorrências do atributo original `CID` melhorando a convergência dos resultados quando submetido aos algoritmos.

3.3. Ferramenta PHP Para Manipulação dos Dados

Para a manipulação dos dados foi desenvolvida uma ferramenta específica, implementada na linguagem de programação PHP. Essa ferramenta consistiu em uma interface web com três funcionalidades:

F1. Importar os dados do arquivo `medico.csv` para a tabela MEDICO do MySQL.

F2. Gerar estatísticas, de acompanhamento, para os dados importados, campo a campo.

F3. Gerar os arquivos com a extensão `.arff`, principal formato lido pelo Weka.

3.4. Ambiente Weka

Para as análises dos atributos, foi utilizado o Weka, *Waikato Environment for Knowledge Analysis*, reconhecido por ser referência em mineração de dados [Weka 2018]. Foi desenvolvido pela Universidade de Waikato, na Nova Zelândia, em linguagem JAVA. Tem como principal objetivo resolver problemas de mineração de dados baseado em aprendizagem de máquina, por meio de métodos de análise em conjuntos de dados, de uma forma rápida e flexível [Hall et al. 2009].

No Weka, o algoritmo indutor de árvore de decisão escolhido foi o J48 versão implementada na linguagem Java do algoritmo C4.5. Já o algoritmos gerador de regras de classificação utilizado foi o JRip, que é uma versão otimizada do algoritmo IREP.

4. Resultados e Discussão

O Weka foi alimentado por um arquivo `.arff` gerado com 19.926 instâncias correspondente aos dados das 14 Unidades Básicas de Saúde e da de Pronto Atendimento do distrito sanitário do bairro Boqueirão, correspondentes aos atendimentos do mês de março de 2017. Foram aplicados os algoritmos J4.8 e JRip para a classificação do atributo `solicit_exam`, pois dentre os atributos disponíveis na base de dados disponibilizada, é o que mais reflete a expectativa de atendimento do usuário evidenciando a resolutividade de uma Unidade Básica de Saúde.

Em ambos os classificadores foi usada a validação *cross-validation*, que consiste em dividir o conjunto total de dados em k subconjuntos. Total de $k=10$ no experimento realizado, sendo os subconjuntos mutuamente exclusivos e do mesmo tamanho. Usando 9 conjuntos, em cada ciclo, para treinar o modelo e o conjunto restante para teste do modelo obtido.

4.1. Classificador J.48

A árvore de decisão obtida, com tamanho de total de 39 e número de folhas igual a 32, pode ser vista no quadro 1. Do total de 26 atributos, o que o algoritmo J4.8 evidenciou como raiz, ou seja, como preponderante no momento de solicitar exames, foi o `enc_atend_especia`, com 4.484 ocorrências iguais a Sim dentro de um universo de 19.926 ocorrências. Seguido pelo atributo `descr_CBO` no segundo nível da árvore. Obtendo índice de 74.66% das instâncias classificadas corretamente.

Uma característica importante a se notar na árvore gerada, são os dois números exibidos ao lado de cada folha. O primeiro mostra o total de classificações realizadas, já o segundo o total de classificações realizadas incorretamente dentro do total realizado, dando assim uma ideia da eficácia da classificação realizada.

```

enc_atend_especia = Nao: Nao (15442/3473)
enc_atend_especia = Sim
|   descr_CBO = MEDICO_CLINICO
|   |   FARM_NAO_PADRON = NAO_PADRON_NAO
|   |   |   CAP_CID = 10:CapX: Sim (25/11)
|   |   |   CAP_CID = 11:CapXI: Sim (40/10)
|   |   |   CAP_CID = 12:CapXII: Sim (16/7)
|   |   |   CAP_CID = 13:CapXIII: Sim (237/84)
|   |   |   CAP_CID = 14:CapXIV: Sim (34/12)
|   |   |   CAP_CID = 18:CapXVIII: Sim (165/57)
|   |   |   CAP_CID = 19:CapXIX: Nao (4/0)
|   |   |   CAP_CID = 1:CapI: Nao (4/1)
|   |   |   CAP_CID = 21:CapXXI: Nao (571/254)
|   |   |   CAP_CID = 3:CapII: Sim (2/0)
|   |   |   CAP_CID = 4:CapIV
|   |   |   |   sexo = F: Sim (67/26)
|   |   |   |   sexo = M: Nao (32/13)
|   |   |   CAP_CID = 5:CapV
|   |   |   |   sexo = F: Sim (47/21)
|   |   |   |   sexo = M: Nao (28/11)
|   |   |   CAP_CID = 6:CapVI: Sim (6/3)
|   |   |   CAP_CID = 7:CapVII: Nao (1/0)
|   |   |   CAP_CID = 8:CapVIII: Nao (4/0)
|   |   |   CAP_CID = 9:CapIX: Sim (128/57)
|   |   |   CAP_CID = ?: Sim (0)
|   |   FARM_NAO_PADRON = NAO_PADRON_SIM: Sim (168/51)
|   descr_CBO = MEDICO_DA ESTRATEGIA DE SAUDE DA FAMILIA
|   |   IDADE_FAIXA = 1:Bebe: Nao (59/20)
|   |   IDADE_FAIXA = 2:Crianca: Nao (143/59)
|   |   IDADE_FAIXA = 3:Adolescente: Sim (84/38)
|   |   IDADE_FAIXA = 4:AdultoJovem: Sim (399/146)
|   |   IDADE_FAIXA = 5:Adulto: Sim (577/177)
|   |   IDADE_FAIXA = 6:Idoso: Sim (522/209)
|   descr_CBO = MEDICO_GENERALISTA: Sim (0)
|   descr_CBO = MEDICO_GINECOLOGISTA_E_OBSTETRA: Sim (834/103)
|   descr_CBO = MEDICO_PEDIATRA: Nao (281/91)
|   descr_CBO = MEDICO_PSIQUIATRA: Sim (0)
|   descr_CBO = MEDICO_RESIDENTE: Sim (0)

Number of Leaves : 32
Size of the tree : 39

```

Quadro 1: Árvore de decisão gerada com o classificador J4.8 do atributo Solicitação de Exames

4.2 Classificador JRIP

O resultado do Classificador JRip, foi a geração de 13 regras, com 12 indicando a solicitação de exames igual a Sim, totalizando de 3984 – 20,0 % das 19926 instâncias e 1 igual a Não, totalizando 15942 – 80,0% do total das instâncias. Na tabela 2, as regras obtidas foram colocadas em ordem de decrescente de ocorrências. Do total de 26 atributos, o que o algoritmo JRip evidenciou como preponderante no momento de solicitar exames, foi o atributo `enc_atend_especia`, com 3.572 ocorrências iguais a Sim.

O fato da saída gerada ser um texto com os atributos com seus respectivos valores conectados pelo operador lógico *and*, facilita a leitura e a respectiva interpretação da regra como um todo. Assim percebe-se que o atributo `enc_atend_especia` igual a Sim é prevalente nas regras encontradas.

Semelhante a árvore de decisão, na última coluna da tabela 2 são exibidos dois números, o primeiro com o total de classificações realizadas, e o segundo com o total de classificações realizadas incorretamente. O classificador JRip, de maneira global, obteve um índice de 74.80% das instâncias classificadas corretamente.

Tabela 2: Regras geradas com o classificador JRip do atributo Solicitação de Exames

Núm	Regras com solicit_exam=Sim	Ocorrências Total/Erros
1	(enc_atend_especia = Sim) and (descr_CBO = MEDICO_GINECOLOGISTA_E_OBSTETRA)	834/103
2	(enc_atend_especia = Sim) and (IDADE_FAIXA = 4:AdultoJovem)	610/259
3	(enc_atend_especia = Sim) and (descr_CBO = MEDICO_DA ESTRATEGIA DE SAUDE DA FAMILIA) and (IDADE_FAIXA = 5:Adulto)	514/165
4	(enc_atend_especia = Sim) and (IDADE_FAIXA = 5:Adulto)	504/240
5	(enc_atend_especia = Sim) and (IDADE_FAIXA = 6:Idoso) and (sexo = F) and (grupo_comunitario = NAO_PARTICIPA)	359/151
6	(descr_CBO = MEDICO_CLINICO) and (CAP_CID = 4:CapIV)	356/174
7	(enc_atend_especia = Sim) and (CAP_CID = 13:CapXIII)	355/77
8	(enc_atend_especia = Sim) and (descr_CBO = MEDICO_DA ESTRATEGIA DE SAUDE DA FAMILIA) and (IDADE_FAIXA = 6:Idoso)	266/115
9	(enc_atend_especia = Sim) and (sexo = F) and (CAP_CID = 18:CapXVIII) and (FARM_NAO_PADRON = NAO) and (meio_transporte = ONIBUS)	85/11
10	(enc_atend_especia = Sim) and (CAP_CID = 18:CapXVIII) and (DIA_DA_SEMANA = 5-Quinta)	45/9
11	(descr_CBO = MEDICO_CLINICO) and (CAP_CID = 9:CapIX) and (descr_unidade = UMS_MENONITAS)	36/12
12	(descr_CBO = MEDICO_CLINICO) and (CAP_CID = 9:CapIX) and (FARM_NAO_PADRON = Sim)	20/6
13	solicit_exam=Não	15942/3601

O percentual de instâncias classificadas corretamente, para os classificadores J4.8 e JRip, foi de 74.66% e 74.80 %, respectivamente, mostrando assim que o desempenho dos algoritmos foi semelhante, com ambos convergindo ao apontar o atributo `enc_atend_especia` como principal a ser considerado para quando se considera a solicitação de exames de um paciente em uma Unidade Básica de Saúde.

5. Conclusões

Utilizando um algoritmo classificador de árvore de decisão (J4.8) e um gerador de regras (JRip), foi possível evidenciar os atributos que contribuem para a solicitação de exames, nas Unidades Básicas de Saúde do distrito sanitário do bairro Boqueirão.

Ambos os algoritmos apresentaram boa precisão na classificação, com percentuais quase idênticos, próximo a 75%. A combinação dos dois também ajuda na interpretação dos resultados, unindo a simplicidade das árvores de decisão com o formato textual das regras, o que facilita a interpretação de um gestor, indicando ao mesmo como e onde atuar na gerência dos recursos humanos das unidades de saúde.

A ferramenta web criada, possibilitou a manipulação dos dados de forma automatizada, dispensando um maior conhecimento especializado em computação, fazendo o gestor focar no objetivo do estudo em questão, que foi o de evidenciar a resolutividade das Unidades Básicas de Saúde, podendo tomar decisões como, por exemplo, desenvolver ações de reciclagem dos profissionais médicos para possivelmente diminuir a solicitação de exames. Assim, podemos concluir que o objetivo deste trabalho, com auxílio de um arcabouço computacional, foi alcançado de maneira efetiva e satisfatória.

Referências

- Brasil (2018). Sistema Único de Saúde, Portal do Ministério da Saúde. Disponível em : <http://www.saude.curitiba.pr.gov.br/a-secretaria/historico-da-secretaria.html> Acessado em 28/03/2018.
- Brasil (2011). Ministério da Saúde. O SUS de A a Z : garantindo saúde nos municípios. Série F. Comunicação e Educação em Saúde 3. ed. Brasília. Disponível em: http://bvsmms.saude.gov.br/bvs/publicacoes/sus_az_garantindo_saude_municipios_3e_d_p1.pdf Acessado em 30/03/2018.
- Saúde (2018a). Histórico da Secretaria Municipal de Saúde de Curitiba. Disponível em: <http://www.saude.curitiba.pr.gov.br/a-secretaria/historico-da-secretaria.html> Acessado em 25/02/2018.
- Saúde (2018b). Distritos Sanitários da Secretaria Municipal de Saúde de Curitiba. Disponível em: <http://www.saude.curitiba.pr.gov.br/a-secretaria/equipe/distritos-sanitarios.html> Acessado em 25/02/2018.
- Gabardo, A. C, (2015). Análise de Redes Sociais – Uma Visão Computacional. Ed. Novatec. São Paulo.
- Oliveira Júnior, J. G. de (2015). Identificação de Padrões para a Análise da Evasão em Cursos de Graduação Usando Mineração de Dados Educacionais. Diss. UTFPR p-15.
- Fayyad, U. M., Gregory Piatetsky-Shapiro and Padhraic Smyth (1996). Advances in Knowledge Discovery and DM. American Association for A. I. p. 1–34.
- Ferreira, Eduardo Liboredo, (2015). Mineração de Dados Aplicada à dados médicos. Belo Horizonte. 68f. Dissertação UFMG, p-3, Belo Horizonte.
- Ferreira, Eduardo Liboredo; Rausch, Herbert; Campos, Sergio (2014). Medical Data Mining: A Case Study of a Paracoccidiodomycosis Patient's. IEEE 16th International Conference on e-Health Networking, Applications and Services.
- Librelotto, Solange Rubert; Mozzaquatro, Patricia Mariotto (2013). Análise dos Algoritmos de Mineração J48 e Apriori Aplicados na Detecção de Indicadores da Qualidade de Vida e Saúde. Rev Interdisciplinar de ensino Pesq. e Ext. V.1 n1.
- Romero, C., Ventura, S., Espejo, P. G., and Hervás, C. (2008). Data mining algorithms to classify students. The First International Conference on Educational DM, p 8–17.
- Han, J., Kamber, M., Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition
- Portal (2018). Portal de Dados Abertos da Prefeitura Municipal de Curitiba. Disponível em: <http://www.curitiba.pr.gov.br/dadosabertos/> Acessado em: 26/02/2018.
- MySQL (2018). Global Development Group. The world's most advanced open source database. Disponível em: <http://www.postgresql.org> Acessado em 20/02/2018.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The Weka data mining software: an update. SIGKDD newsletter, 11:10–18.
- Weka (2018). Data Mining with Open Source Machine Learning Software in Java. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/> Acessado em 05/02/2018.