

*Análise de Perfis de Doenças com Base em Técnicas de
Descoberta de Conhecimento em Bases de Dados*

Prof. Dr. André Pinz Borges
apborges@utfpr.edu.br

Autores: Kaique Augusto Morais da Silva, Rodrigo Feuser, Richardson Ribeiro,
Dalcimar Casanova, Marcelo Teixeira, André Pinz Borges

Agenda

- Introdução
- Objetivo
- KDD
- Trabalhos relacionados
- Desenvolvimento
- Resultados
- Conclusões



Introdução

- Trabalho desenvolvido em conjunto com UTFPR-PB
- Dados de exames, consultas são armazenados em *Pontruários Eletrônicos de Pacientes (PEP)*
- Investimentos são realizados no desenvolvimento de PEPs
- Vantagens dos PEP:
 - a agilidade no preenchimento dos documentos; segurança dos dados; atualização em tempo real e a portabilidade.

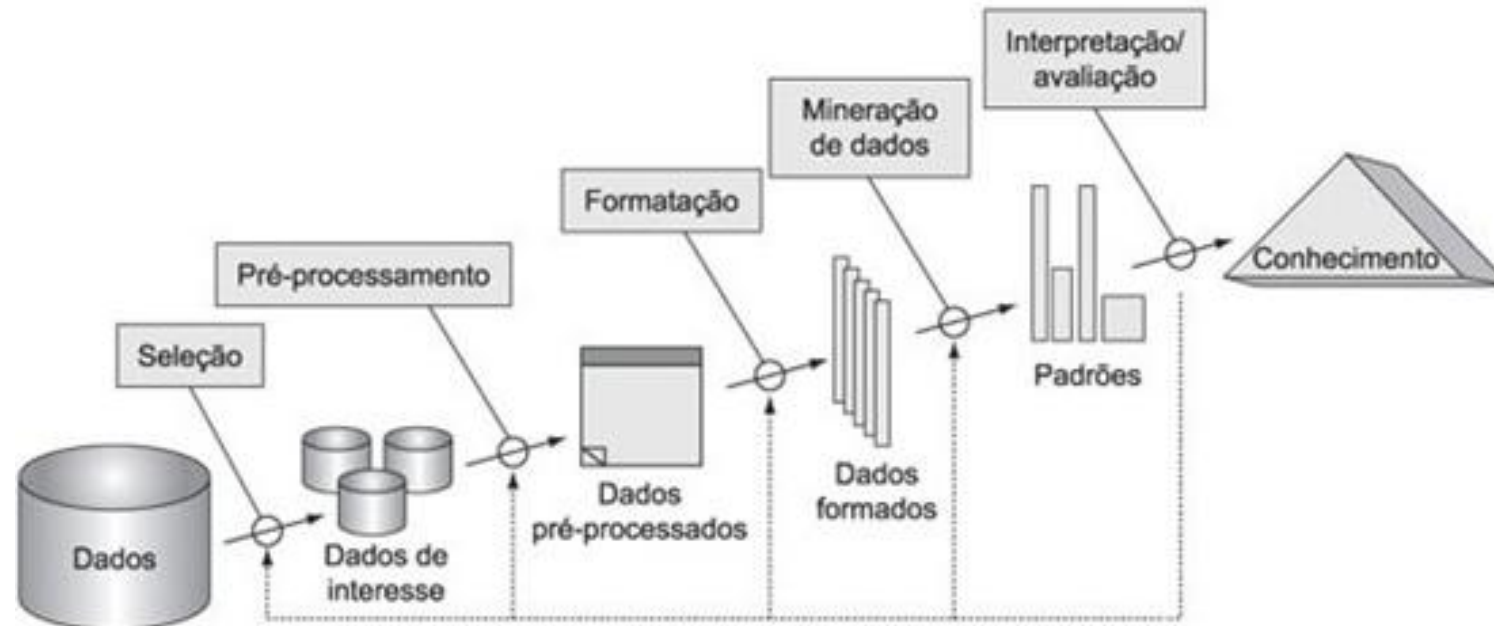
Objetivo

A partir dos dados de um sistema de PEP, descobrir conhecimento potencialmente útil à profissionais de saúde

Analisar Perfis de Doenças com Base em Técnicas de Descoberta de Conhecimento em Bases de Dados

KDD

- Knowledge Discovery in Databases (em português, Descoberta de Conhecimento em Bases de Dados)
 - Processo, não trivial, de identificar, em grandes conjuntos de dados, padrões que sejam válidos, novos, úteis e compreensíveis, buscando melhorar o entendimento de um problema ou um procedimento de tomada de decisão.



Etapas do KDD. [Witten 2011]

Trabalhos Relacionados

- Vilarinho R. (2017) utilizou algoritmos de mineração para obtenção de informações úteis relativas a casos de Dengue nos municípios brasileiros.
- Trindade (2012) aplicou o KDD para a identificação de padrões de comportamento das Hepatites Virais nas bases de dados do SINAN (Sistema de Informação de Agravos e Notificações)
- Martins e Lima (2014) apresentam um estudo com as vantagens e desvantagens de usar um PEP
- Feuser R. (2017), que aplicou os processos do KDD para PEP oriundo do SUS, utilizando algoritmo de associação a priori na etapa da mineração de dados, encontrando regras com elevado fator de confiança.

Desenvolvimento

- Origem dos dados:
 - projeto com a participação das secretárias de Saúde, Ciência e Tecnologia do município de Pato Branco, bem como a participação da Universidade Tecnológica Federal do Paraná (UTFPR) Campus Pato Branco.
- 43.879 pacientes
- 2.296.626 registros de atendimentos

Preprocessamento

1. identificação da faixa etária dos pacientes;
2. utilização de dados do censo (e-SUS) tais como:
 - bairro, altura, peso, frequência escolar, se frequenta benzedeira, se possui plano de saúde, se é fumante, se possui diabetes, se é gestante, se possui asma, se é alcoólatra, se já teve infarto, AVC ou derrame;
3. remoção de registros com dados ausentes ou com ruídos foram eliminados
 - Exemplo, pacientes que não possuem identificação de doenças ou que possuem apenas informações de exames;
4. remoção de dados discrepantes relacionados à altura e peso dos pacientes, utilizando para isto valores mínimos e máximos.
5. Enriquecimento com dados da Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID)

Análise do Grupo de CIDs

Tabela 1: Quantidade de ocorrências de Traumatismos

Tipo de Traumatismo	Qtde
Outras causas externas de traumatismos acidentais	16
Sequelas de traumatismos, intoxicações e de outras consequências das causas externas	5
Cabeça	18
Abdome, dorso, coluna lombar e da pelve	4
De localização não especificada do tronco, membro ou outra região do corpo	13
Cotovelo e do antebraço	11
Joelho e da perna	51
Ombro e do braço	7
Pescoço	3
Punho e da mão	49
Quadril e da coxa	7
Tórax	33
Tornozelo e do pé	40
Total Geral	257

Tabela 2: Quantidade de ocorrências de Neoplasias

Tipo de Neoplasia	Qtde
Melanoma e outras(os) neoplasias malignas da pele	2
Neoplasias (tumores) benignas(os)	65
Neoplasias (tumores) de comportamento incerto ou desconhecido	3
Neoplasias (tumores) in situ	20
Neoplasias (tumores) malignas(os)	3
Neoplasias (tumores) malignas(os), declaradas ou presumidas como primárias, dos tecidos linfático, hematopoético e tecidos correlatos	7
Neoplasias malignas da mama	2
Neoplasias malignas do aparelho respiratório e dos órgãos intratorácicos	7
Neoplasias malignas dos órgãos digestivos	4
Neoplasias malignas dos órgãos genitais masculinos	6
Total Geral	119

Resultados

- Weka 3.8
- Algoritmos :
 - C4.5 (algoritmo J48)
 - Bagging
 - Boosting (algoritmo AdaBoostM1)
 - Todos utilizando as configurações padrão do Weka.
- Validação cruzada 10-folds

Resultados

- Percentuais de acerto

	Neoplasias			Traumatismos		
	J48	Bagging	Boosting	J48	Bagging	Boosting
Verdadeiros positivos	84%	85%	87%	8%	7%	7%
Falso positivos	9%	10%	8%	0%	0%	0%
Precisão	70%	70%	87%	60%	61%	65%

Resultados

- Trecho das Regras da Árvore de Decisão dos Neoplasias

BAIRRO = PINHEIRINHO: Neoplasias [tumores] in situ (11.0/1.0)

BAIRRO = SANTA TEREZINHA: Neoplasias [tumores] in situ (4.0)

BAIRRO = ALVORADA

| Frequenta Escola = NAO: Neoplasias [tumores] benignas(os) (17.0/3.0)

| Frequenta Escola = SIM: Neoplasias malignas do aparelho respiratório e dos órgãos intratorácicos (7.0)

BAIRRO = SAO FRANCISCO: Neoplasias [tumores] benignas(os) (11.0)

BAIRRO = PLANALTO

| Faixa etaria = 40 A 60: Neoplasias [tumores] benignas(os) (4.0/1.0)

| Faixa etaria = 0 A 7: Neoplasias [tumores] in situ (0.0)

| Faixa etaria = 30 A 40: Neoplasias [tumores] in situ (0.0)

| Faixa etaria = ACIMA DE 60: Neoplasias [tumores] in situ (4.0/1.0)

| Faixa etaria = 18 A 30: Neoplasias [tumores] in situ (0.0)

BAIRRO = MORUMBI

| Faixa etaria = 40 A 60: Neoplasias malignas da mama (2.0)

| Faixa etaria = 0 A 7: Neoplasias [tumores] benignas(os) (0.0)

| Faixa etaria = 30 A 40: Neoplasias [tumores] benignas(os) (2.0)

| Faixa etaria = ACIMA DE 60: Neoplasias [tumores] benignas(os) (0.0)

| Faixa etaria = 18 A 30: Neoplasias [tumores] benignas(os) (0.0)

BAIRRO = VENEZA: Neoplasias [tumores] benignas(os) (2.0)

BAIRRO = VILA VERDE: Neoplasias [tumores] benignas(os) (2.0/1.0)

Resultados

- Trecho das Regras da Árvore de Decisão dos Traumatismos

BAIRRO = MORUMBI

```
| Faixa etaria = 0 A 7: Traumatismos do tornozelo e do pizço (1.0)
| Faixa etaria = 40 A 60
| | Sexo = MASCULINO: Traumatismos do joelho e da perna (3.0)
| | Sexo = FEMININO: Traumatismos do ombro e do braizço (2.0)
| Faixa etaria = ACIMA DE 60: Traumatismos do joelho e da perna (0.0)
| Faixa etaria = 30 A 40: Traumatismos do joelho e da perna (0.0)
| Faixa etaria = 18 A 30: Traumatismos do punho e da mizço (3.0)
| Faixa etaria = 8 A 17: Traumatismos do joelho e da perna (0.0)
```

BAIRRO = PINHEIRINHO

```
| Faixa etaria = 0 A 7: Traumatismos do joelho e da perna (0.0)
| Faixa etaria = 40 A 60
| | Hipertensao Arterial = NAO: Traumatismos do joelho e da perna (3
| | Hipertensao Arterial = SIM: Traumatismos do cotovelo e do antebra
| Faixa etaria = ACIMA DE 60: Traumatismos do joelho e da perna (3.0)
| Faixa etaria = 30 A 40: Traumatismos do joelho e da perna (0.0)
| Faixa etaria = 18 A 30: Traumatismos do tornozelo e do pizço (6.0/3.0)
| Faixa etaria = 8 A 17: Traumatismos do joelho e da perna (0.0)
```

Conclusões

- KDD é uma alternativa para obtenção de informações dos PEPs
- Dados de PEPs são difíceis de serem disponibilizados
 - Não há legislação sobre a disponibilidade de prontuários eletrônicos padronizados, ou centralizados
- Divisão em dois grupos de doenças (neoplasias e traumatismos) permitiu analisar os atributos mais relevantes na identificação destas doenças.
- Limitações:
 - Falta de um profissional da saúde para analisar os dados resultantes;
 - Grande número de prontuários com dados ausentes.

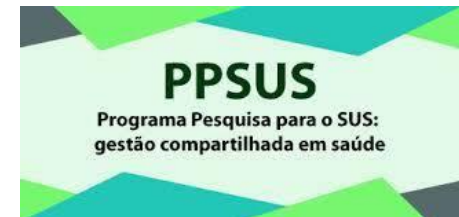
Trabalhos Futuros

- Validação dos resultados obtidos por profissionais da saúde para definir a veracidade das regras encontradas.
- Isso poderia expandir este trabalho para outros locais de atendimentos de emergência, como: hospitais, ambulatórios e ou clínicas médicas.
- Integração com dados do IBGE
- Comparação com dados de cidades com mesmas características,
- Aplicações de novos parâmetros e algoritmos.

Agradecimentos

- Apoio:
 - Decit/SCTIE/MS, por intermédio do CNPq, com apoio da Fundação Araucária e da SESA-PR.
 - Programa Pesquisa para o Sistema Único de Saúde: Gestão Compartilhada em Saúde - PPSUS.
- Docentes da UTFPR-PB

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ



Referências

Feuser, R. (2017) “Mineração de Dados com Regras de Associação Aplicada em Dados de Unidade de Saúde de Pronto Atendimento”. UTFPR, Pato Branco (PR).

Martins, C. and Lima, S.M. (2014) “Vantagens e desvantagens do prontuário eletrônico para instituição de saúde”. Revista de Administração em Saúde (RAS), v. 16, n. 63.

Trindade, C M et al. (2012) “Technology in health: knowledge discovery in public health databases: study of viral hepatitis in the state of Paraná”, Brazil. Iberoamerican Journal of Applied Computing, Ponta Grossa, v. 2, n. 2.

Vilarinho R.A. (2017) “Uso de Técnicas de Mineração de Dados para Classificação das Ocorrências de Casos de Dengue nos Municípios Brasileiros”. UFOP – Universidade Federal de Ouro Preto. Março.

Witten, I. H. (2011) “Data Mining Practical Machine Learning Tools and Techniques”. 3 ed.

Agradecimentos

