



APLICAÇÃO DE MINERAÇÃO DE DADOS PARA CARACTERIZAÇÃO DA EVOLUÇÃO DE PACIENTES COM DOR CRÔNICA NA COLUNA VERTEBRAL

CLEYTON DE C. MARIO¹, LUCIANA B. R. DOS SANTOS², CINARA GHEDINI HITTA³

¹Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP, Câmpus Caraguatatuba, cleytoncm@gmail.com

²Doutora em Computação Aplicada, Orientadora, Professora, IFSP, Campus Caraguatatuba, lurebelo@ifsp.edu.br

³Doutora em Engenharia Eletrônica e Computação, Colaboradora, Pesquisadora, ITA, cinara@ita.br

Área de conhecimento: Banco de Dados - 1.03.03.03-0

RESUMO: A dor na coluna vertebral é um dos grandes problemas da atualidade. Em um trabalho publicado em 2015 por colaboradores da *Global Burden of Disease Study*, no Brasil a dor lombar está ranqueada em primeiro lugar entre os distúrbios que causam mais anos de vida perdido e incapacidade. Já as dores na cervical estão ranqueadas em quarto lugar. Neste projeto, são analisadas bases de dados de uma clínica de fisioterapia existentes desde 2008 com 2000 pacientes. O objetivo principal é realizar um pré-processamento e análise dos dados, verificando qual técnica de Mineração de Dados é mais indicada para os tipos de dados existentes. A partir dos resultados obtidos no presente estudo, espera-se alcançar parte de um projeto maior, que é fazer uma caracterização da evolução dos pacientes com dores nas na coluna vertebral submetidos ao tratamento fisioterapêutico baseado em evidências científicas, além de traçar um perfil dos mesmos, o que torna possível a elaboração de um plano de ação mais eficiente.

PALAVRAS-CHAVE: coluna vertebral; dor lombar e cervical; dados legados; mineração de dados.

DATABASE RESEARCH AIMING UNDERSTANDING OF PATIENTS WITH CHRONIC PAIN OF DEVELOPMENTS IN SPINE

ABSTRACT: The back pain is one of the biggest problems nowadays. At a project published in 2015 by Global Burden of Disease Study employees, in Brazil the lumbar pain is ranked at the first place as the disturbs that results in more years of life lost and lack of capacity. The cervical pain is ranked at the fourth place. This project aims at analysing a physiotherapy's clinical database, that exists since 2008, containing 2000 patients. The main objective is to preprocess and analyze the data, verifying which Data Mining technique is best suited for the existing data types. The resulting database and the preprocessing automation will support a major research subject: analyzing and characterizing data of patients with back pains submitted to the physiotherapeutic treatment based on scientific evidences. In addition, plotting the patient's' profile supports the definitions of action plans and treatment more efficiently.

KEYWORDS: spine; low back and cervical pain; legacy data; data mining.

INTRODUÇÃO

“A dor lombar não especificada não é um problema autolimitante, mas é uma enfermidade recorrente e persistente.” (AXÉN & LEBOEUF-YDE, 2013). Pesquisadores do mundo inteiro se esforçam em melhor compreender aspectos relacionados à dor na coluna. Um dos formatos de ação é entender se há padrões de evolução do paciente e desta forma, perceber se pode-se gerar precocemente um possível prognóstico evolutivo previamente ao início do tratamento. “Entender o comportamento da dor lombar é importante para as clínicas e pesquisadores por fornecer informações necessárias, e potenciais benefícios de tratamento.” (DUNN, JORDAN & CROFT, 2006).

Uma possibilidade para a compreensão da evolução da dor é a abordagem de modelagem por meio da análise latente de classes. Essa linha tem sido cada vez mais reconhecida pela sua utilidade em identificação de subpopulações homogêneas dentro de um grupo heterogêneo, para a identificação de grupos significativos ou classes de indivíduos.

Assim, o presente projeto pretende caracterizar a evolução da dor em pacientes com dores na coluna vertebral submetidos ao tratamento fisioterapêutico baseado em evidências científicas, utilizando técnicas de mineração de dados. Para isso, serão analisadas as bases de dados da clínica que contém o histórico dos pacientes.

MATERIAL E MÉTODOS

O presente estudo tem o objetivo de identificar e definir quais técnicas de *Data Mining* serão as mais indicadas para o tratamento dos dados fornecidos pela clínica fisioterápica, além de definir quais dados são os mais relevantes para realização da análise de dados.

Os pacientes, dos quais os dados serão analisados, foram submetidos a um atendimento fisioterapêutico baseado em evidência composto por ações de dois a seis meses de tratamento intensivo, direcionadas para as necessidades individualizadas do paciente, pautadas em ações de orientações, terapia manual e exercícios específicos.

Para tal, a clínica fisioterápica forneceu uma amostragem de suas bases de dados em formato digital, assim transcrevendo os dados do prontuário manuscrito para uma planilha eletrônica sem qualquer tipo de tratamento.

De modo mais específico, como atividade para o desenvolvimento deste projeto de pesquisa, foi realizado o curso Python Fundamentos para Análise de Dados fornecido pela Data Science Academy para auxílio no desenvolvimento desta pesquisa; Aplicação das fases de Data Mining, especialmente a fase de pré-processamento, visando identificar quais dados possuem mais significância para a análise; e está previsto estudo para identificação de quais tipos de algoritmos podem trazer um melhor resultado ao projeto, esta última atividade em desenvolvimento.

RESULTADOS E DISCUSSÃO

Como a amostragem de dados fornecida pela clínica veio com ruídos, foi necessário fazer um tratamento dos dados e seleção de quais colunas deveriam ser utilizadas para a análise de dados, na Figura 1 é possível visualizar os dados que estão faltando em várias colunas. Usando uma estratégia simples de eliminar as colunas que estavam com a maioria dos campos em brancos, houve uma diferença de cerca de 400 colunas entre a planilha com os dados completos e a nova planilha tratada, que pode ser vista na Figura 2, já na Figura 3 é possível visualizar o tamanho das amostragem no padrão linhas por colunas, para tal foi utilizado comandos em Python (PYTHON) e a utilização da biblioteca Panda (PANDA).

	AG	AH	AI	AJ	AK	AL	
1	Dor secundária constante?	História - Data de início dos sintomas? (Estimativa em meses)	História - Trauma?	História - Se possui algum trauma	qual é o mecanismo da lesão?	História - Informações gerais	Mo
2				Sim			
3			12			Sim	
4				Não			
5	Lombar	Não		240			Sim
6				Não			
7	Cervical	Não		240			Sim
8				Sim			
9				Sim			
10		24			Sim		
11	Nega	Não	Cervical e lombar	Lombar	Não	Cervical	Não
12			Sempre				Sim
13	Não			jan/00			
14		120			Sim		
15		240			Sim		
16				Sim			
17				Sim			
18	Não	Torácica	Não		24		
19				4			Sim
20				Sim			
21				Sim			
22							
23							

Figura 1 - Amostragem sem tratamento

	A	B	C	D	E	F	G	H
1	reincidente	programa	sexo	profissao	areaDor	dorPrincipal	dorPrincipalConstante	evolucao
2	Não		3 feminino	Auxiliar ADM	Lombar irradiada abaixo do joelho	Lombar irradiada abaixo do joelho	Não	
3	Sim		2 feminino	Professor(a) aposentado	Lombar	Lombar	Sim	10 minutos de atraso
4	Sim		2 masculino	Professor(a) aposentado	Região Inguinal	Lombar	Não	Avaliação
5	Não		1 feminino	do lar	Cervical e lombar	Cervical	Não	término avaliação
6	Não		3 masculino	operador de produção	Cervical	Cervical	Sim	
7	Não		1 feminino	do lar	Cervical e lombar	Lombar	Não	término avaliação
8	Não		3 feminino	Professor(a) de dança	Lombar	Lombar	Não	inibição quadrado lombar
9	Não		2 masculino	médico	Sem dor	Sem dor	Não	maca tração lombar (80 kg 30% 10%
10	Não		1 masculino	engenheiro de segurança	Cervical e lombar	Lombar irradiada abaixo do joelho	Sim	FNJ
11	Sim		2 feminino	Professor(a) aposentado	Cervical e lombar	Lombar	Não	inibição
12	Sim		3 feminino	Professor(a) aposentado	Lombar	Lombar	Não	
13	Não		1 feminino	assistente social	Cervical	Cervical	Não	avaliação
14	Não		2 masculino	mecânico de manutenção aposentado	Lombar	Lombar	Não	estabilização
15	Não		1 masculino	físico	Lombar	Lombar	Não	manipulação lombar
16	Não		1 masculino	operador de máquina	Lombar	Lombar	Não	maca flexo distração
17	Sim		2 masculino	transportador escolar	Lombar	Lombar	Não	treino de mobilidade lombar
18	Não		1 feminino	cozinheira	Cervical	torácica e lombar	Cervical irradiada para os braços	Não
19	Não		1 feminino	aposentado(a)	Lombar	Lombar irradiada acima do joelho	Não	término avaliação. Inibição mm
20	Não		1 masculino	operador de produção	Cervical e lombar	Lombar	Não	término avaliação
21	Não		1 masculino	aposentado(a)	Lombar irradiada abaixo do joelho	Lombar irradiada abaixo do joelho	Não	inibição quadrado lombar e piriforme
22								
23								
24								
25								
26								

Figura 2 - Amostragem tratada

```
# Carregando os dados do arquivo em dataset
dados = pd.read_csv("FormularioPacientes.csv", sep = ";")
```

```
# Recuperando o formato do dataset (dimensões)
dados.shape
```

(20, 592)

```
# Carregando os dados do arquivo em dataset
dadosTratados = pd.read_csv("Evolucao.csv", sep = ";")
```

```
dadosTratados.shape
```

(20, 155)

Figura 3 - Comandos em Python

Após o tratamento dos dados foi possível aplicar fazer algumas análises que podem ser vistas na Figura 4 e Figura 5, onde foram aplicadas técnicas de cruzamento de dados entre profissão e área de dor e profissão e a dor principal. Essas informações foram entregues aos responsáveis da clínicas na espera de confirmação da relevância das informações encontradas.



Figura 4 - Relação profissão com dor principal

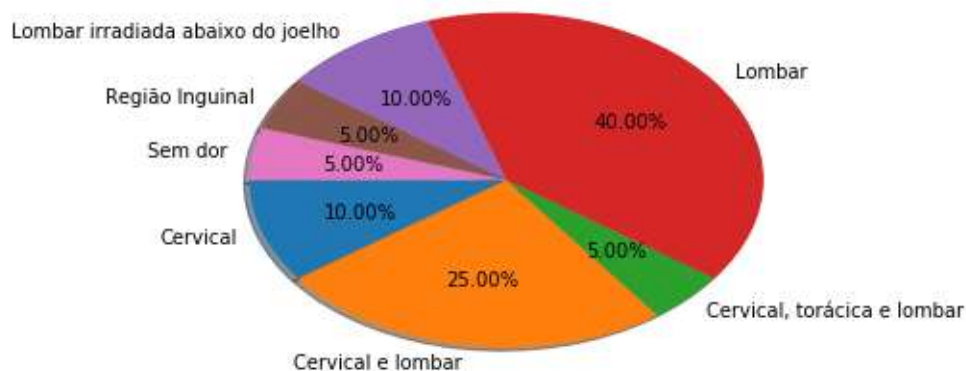


Figura 5 - Relação profissão com área da dor

CONCLUSÕES

Após a conclusão da etapa de pré-processamento da amostra obtida, planeja-se definir quais dados poderão ser úteis ou não para a análise de dados focando a indicação dos preditores clínicos, assim como também espera-se ter a definição de quais tipos de algoritmos serão utilizados na mineração.

Com a implantação completa do sistema, prevê-se o início da análise dos dados contidos na nova base de dados, para tal serão utilizadas técnicas de mineração de dados, objetivando o encontro de preditores clínicos que possam formar *clusters* de pacientes, com isso auxiliando no entendimento da evolução da dor e na melhor forma de tratá-la.

REFERÊNCIAS

IBEN AXÉN, CHARLOTTE LEBOEUF-YDE, Best Practice & Research Clinical Rheumatology, v. 27, Issue 5, p. 601-612 (October 2013) Trajectories of low back pain.

KETE M. DUNN, KELVIN JORDAN & PETER R. CROFT, American Journal of Epidemiology, v. 163, nº 8, p. 754-751 (2006) Characterizing the Course of Low Back Pain: A Latent Class Analysis.

PYTHON. DISPONÍVEL EM <https://www.python.org/about/apps/>. ACESSO EM: 04 DE JULHO DE 2017.

PANDA. DISPONÍVEL EM <http://pandas.pydata.org/>. ACESSO EM: 04 DE JULHO DE 2017.