

UNIVERSIDADE FEEVALE

WESLLEI FELIPE HECKLER

ANÁLISE PREDITIVA PARA IDENTIFICAÇÃO DE
TENDÊNCIAS SOBRE PACIENTES DO “PROJETO DE
EXTENSÃO REABILITAÇÃO PULMONAR” DA
UNIVERSIDADE FEEVALE

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo, março de 2018

WESLLEI FELIPE HECKLER

ANÁLISE PREDITIVA PARA IDENTIFICAÇÃO DE
TENDÊNCIAS SOBRE PACIENTES DO “PROJETO DE
EXTENSÃO REABILITAÇÃO PULMONAR” DA
UNIVERSIDADE FEEVALE

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientador: Juliano Varella de Carvalho

Novo Hamburgo, março de 2018

RESUMO

As doenças respiratórias atingem um dos principais sistemas do corpo humano e afetam grande parte da população brasileira. Nos casos mais graves, podem limitar as funcionalidades e a força muscular dos portadores, impactando em atividades cotidianas simples e, conseqüentemente, na qualidade de vida dos mesmos. Os Programas de Reabilitação Pulmonar auxiliam no tratamento dessas doenças. O “Projeto de Extensão Reabilitação Pulmonar” da Universidade Feevale é um programa de reabilitação pulmonar que atende pacientes da comunidade e visa melhorar a qualidade de vida de portadores de doenças respiratórias crônicas através do desenvolvimento de ações educativas e assistenciais. As informações sobre os pacientes e resultados do tratamento são armazenadas em uma base de dados. A quantidade de dados dificulta a análise dos resultados. Neste trabalho será desenvolvida uma ferramenta para aplicar técnicas de *machine learning* e análise preditiva na base de dados do “Projeto de Extensão Reabilitação Pulmonar”. O objetivo da ferramenta é identificar tendências em relação aos pacientes e extrair conhecimento sobre a base de dados para, com isso, contribuir na aplicação do tratamento de reabilitação pulmonar. Além disso, a ferramenta também disponibilizará visualizações para auxiliar na leitura dos dados e dos resultados por profissionais da área da saúde.

Palavras-chave: *Machine learning*. Análise preditiva. Extração de conhecimento. Doenças respiratórias crônicas. Programas de Reabilitação Pulmonar.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA	11
BIBLIOGRAFIA	12

MOTIVAÇÃO

Atualmente as doenças respiratórias atingem grande parte da população brasileira. Em 2015, o Ibope realizou uma pesquisa que revelou que 44% dos brasileiros apresentam sintomas de doenças respiratórias. A pesquisa também revelou que a maior incidência dos sintomas foi nos estados do Sul do país (VIDALE, 2015).

Segundo Duchiae (1992), a poluição do ar é um dos principais fatores que causam as doenças respiratórias, podendo tanto atingir pessoas saudáveis quanto agravar os sintomas de pessoas que já possuem uma doença respiratória. Outro fator bastante significativo é o tabagismo, que é o principal agente causador da doença pulmonar obstrutiva crônica (II CONSENSO BRASILEIRO SOBRE DOENÇA PULMONAR OBSTRUTIVA CRÔNICA, 2004).

Essas doenças atingem um dos principais sistemas do corpo humano e, por isso, suas consequências são graves. Em níveis mais críticos, podem afetar atividades cotidianas simples. Segundo Poulain et al. (2003, tradução nossa), pacientes com doença pulmonar obstrutiva crônica apresentam dificuldades na realização de exercícios, como por exemplo a caminhada, devido à obstrução das vias aéreas. Além disso, os pacientes portadores de doença pulmonar obstrutiva crônica também podem apresentar ansiedade e depressão, causando fadiga e dispnéia. Esses sintomas implicam significativamente na realização de atividades diárias e, conseqüentemente, na qualidade de vida dos pacientes (MAURER et al., 2008, tradução nossa).

As doenças respiratórias são divididas em dois grupos conforme prejudicam as funções pulmonares. As doenças respiratórias **restritivas** têm como característica a redução da capacidade pulmonar (RODRIGUES et al., 2002). Já as **obstrutivas** “são aquelas associadas a aumento da resistência nas vias aéreas, que se reflete funcionalmente por significativa redução nos fluxos expiratórios máximos” (FILHO, 1998). Também existe a doença respiratória mista, que é caracterizada pela junção dos sintomas das doenças restritivas e obstrutivas.

Um dos sintomas mais comuns das doenças respiratórias é a dispnéia, uma sensação de falta de ar. Esse sintoma pode variar de acordo com cada caso, podendo ser mais intenso em alguns pacientes. Em geral, esse sintoma é percebido quando o paciente realiza alguma atividade, mesmo que essa atividade não exija muito esforço. Além disso, a ansiedade e o medo também podem causar essa sensação, visto que a falta de ar também está ligada ao emocional do paciente (GILMAN; BANZETT, 2009, tradução nossa).

Os Programas de Reabilitação Pulmonar são bastante indicados para o tratamento e têm resultados muito positivos. Sobre esses programas, Nici et al. (2006, tradução nossa) comentam que:

A reabilitação pulmonar é uma intervenção baseada em evidências, multidisciplinar e abrangente para pacientes com doenças respiratórias crônicas que são sintomáticas e muitas vezes diminuem as atividades da vida diária. Integrada no tratamento individualizado do paciente, a reabilitação pulmonar é projetada para reduzir os sintomas, otimizar o estado funcional, aumentar a participação e reduzir os custos dos cuidados de saúde através da estabilização ou reversão das manifestações sistêmicas da doença.

O “Projeto de Extensão Reabilitação Pulmonar” da Universidade Feevale é um Programa de Reabilitação Pulmonar que atende pacientes da comunidade de ambos os sexos e com idade superior a 40 anos. O projeto visa desenvolver e implementar ações educativas e assistenciais que promovam a melhoria da qualidade de vida em portadores de doenças respiratórias crônicas, tais como asma, doença pulmonar obstrutiva crônica, fibrose pulmonar, bronquiectasia, entre outras. O tratamento dura três meses em média.

São realizadas entrevistas pré e pós-tratamento por meio de formulários para registro de informações. As entrevistas são aplicadas para coletar informações sobre o perfil do paciente como por exemplo gênero, idade, altura, peso, doença respiratória e período do tratamento. Além disso, visam coletar indicadores sobre a saúde como por exemplo o número de vezes que o paciente tossiu em determinado período, a existência de secreção ou pressão no peito e a ocorrência de dispneia. Também são coletados indicadores sobre a qualidade de vida do paciente. Entre esses indicadores estão o grau de limitação de atividades habituais diárias (caminhar, conversar, arrumar a cama, lavar a louça, entre outras), a qualidade do sono e o nível de disposição.

Juntamente a essas informações, é registrado o desempenho dos pacientes em testes de carga máxima, que consistem na aplicação de 10 exercícios durante o tratamento de reabilitação. Os indicadores de desempenho são utilizados para medir a evolução do paciente durante o tratamento. Essas informações são registradas desde 2002 em uma base de dados armazenada em uma planilha excel. Ao todo, a base de dados contém 542 linhas, que representam os pacientes, e 356 colunas, que representam os atributos. A quantidade de registros e atributos dificulta a análise e visualização dos resultados. Em contrapartida, é possível que a base de dados contenha informações importantes sobre relacionamentos entre pacientes e doenças respiratórias que podem ser estudadas por profissionais da saúde.

A quantidade de dados no mundo está aumentando e tende a continuar crescendo. Quanto maior essa quantidade, mais difícil se torna o entendimento das pessoas sobre esses dados, que podem conter informações potenciais. Desde que a vida humana começou, as pessoas buscam padrões sobre os dados, visando obter algum benefício. Caçadores buscam padrões sobre a migração de animais, agricultores buscam padrões sobre o crescimento das plantações e empreendedores buscam padrões de comportamento que podem ser transformados em negócios lucrativos. Padrões podem trazer explicações sobre os dados e auxiliar na predição de dados futuros (WITTEN; FRANK, 2005, tradução nossa).

Extrair conhecimento sobre bases de dados pode gerar informações úteis e benefícios para quem as analisa. Goldschmidt e Passos (2005) destacam que

A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação.

Segundo Mitchell (2006, tradução nossa), *Machine Learning* abrange as áreas de Ciência da Computação e Estatística e consiste na criação de algoritmos que aprendam e se adaptem automaticamente a partir da experiência adquirida sobre um determinado problema. Ademais, também consiste no estudo das inferências que podem ser realizadas sobre os dados analisados.

Técnicas de *Machine Learning* podem ser aplicadas na área da saúde de diversas formas. Alguns exemplos de aplicação são a classificação de imagens microscópicas de células, detecção de surtos de doenças e detecção de padrões anômalos de sintomas e sua distribuição geográfica (MITCHELL, 2006, tradução nossa).

A análise preditiva é uma subárea de *Machine Learning* e tem como objetivo prever comportamentos futuros a partir do aprendizado sobre dados do passado referentes a determinado problema. Conforme Alpaydin (2010, tradução nossa) destaca, “assim que tivermos uma regra que se ajusta aos dados anteriores, se o futuro for semelhante ao passado, podemos fazer previsões corretas para novas instâncias”.

Essa técnica tem diversas aplicações na área da saúde. Como por exemplo na análise de quais tratamentos serão mais eficientes para determinados futuros pacientes (MITCHELL, 2006, tradução nossa) e na busca por doadores de sangue (SILVA, 2018). Outro exemplo de aplicação é na predição da ação de medicamentos através da análise de suas propriedades químicas e de sua estrutura tridimensional, acelerando a descoberta de novos medicamentos e reduzindo seu custo (WITTEN; FRANK, 2005, tradução nossa).

Este trabalho, portanto, propõe o desenvolvimento de uma ferramenta para extração de conhecimento da base de dados do “Projeto de Extensão Reabilitação Pulmonar”. Através da aplicação de técnicas de *Machine Learning*, a ferramenta tem como objetivo identificar tendências sobre os perfis dos pacientes. Além disso, a ferramenta também visa facilitar a leitura dos resultados através da geração de visualizações sobre os dados.

OBJETIVOS

Objetivo Geral:

Aplicar técnicas de *Machine Learning* na base de dados do Projeto de Extensão "Reabilitação Pulmonar" da Universidade Feevale, a fim de identificar tendências em relação aos perfis dos pacientes.

Objetivos Específicos:

- Otimizar a aplicação do tratamento de reabilitação dos pacientes.
- Gerar visualizações para facilitar a leitura dos dados por profissionais da área da saúde.
- Identificar relações entre pacientes e doenças respiratórias.
- Extrair conhecimento da base de dados por meio da aplicação de análise preditiva.

METODOLOGIA

Neste trabalho será realizada uma pesquisa de natureza aplicada. Para tal, serão estudadas a linguagem de programação R e suas bibliotecas, bem como as principais técnicas de *machine learning* e análise preditiva. Após esse estudo, o conhecimento adquirido será aplicado no desenvolvimento de uma ferramenta para extração de conhecimento da base de dados do “Projeto de Extensão Reabilitação Pulmonar” da Universidade Feevale.

Em relação aos objetivos, será realizada uma pesquisa explicativa, pois serão aplicadas técnicas de *machine learning* para identificação de tendências na base de dados do projeto de reabilitação pulmonar. Com isso, o trabalho visa auxiliar a compreensão de profissionais da área da saúde sobre a aplicação do tratamento para reabilitação pulmonar.

Os procedimentos técnicos utilizados para o desenvolvimento do trabalho serão a pesquisa bibliográfica e a pesquisa experimental. Inicialmente, será realizado um levantamento do referencial teórico através da pesquisa em livros e trabalhos correlatos. Após isso, serão estudados algoritmos e técnicas de análise preditiva para aplicação no desenvolvimento da ferramenta. Como citado anteriormente, serão utilizados os dados coletados no projeto de reabilitação pulmonar para a aplicação de *machine learning*.

Após o levantamento do referencial teórico e pesquisa experimental, será iniciado o desenvolvimento da ferramenta que aplicará técnicas de análise preditiva sobre os dados, extraíndo conhecimento e identificando tendências sobre os pacientes. Ademais, também serão geradas visualizações para facilitar a leitura dos dados e análise dos resultados. Após isso, a ferramenta desenvolvida será avaliada por profissionais envolvidos no projeto de reabilitação pulmonar para garantir que está alinhada à real necessidade do projeto.

Por fim, do ponto de vista da abordagem, essa será qualitativa visto que objetivo do trabalho não é quantificar os dados estatisticamente, mas sim extrair informações que possam contribuir para a compreensão de profissionais da área da saúde sobre o tratamento. Assim sendo, a partir da aplicação de *machine learning*, espera-se identificar tendências sobre pacientes com doenças respiratórias crônicas e, desta forma, auxiliar no tratamento para reabilitação pulmonar.

Seguindo a metodologia acima, este trabalho compromete-se a alcançar os objetivos geral e específicos. Além disso, também se direciona para a resposta da seguinte questão de pesquisa: É possível identificar tendências sobre pacientes através da aplicação de técnicas de

análise preditiva na base de dados do “Projeto de Extensão Reabilitação Pulmonar” da Universidade Feevale?

CRONOGRAMA

Trabalho de Conclusão I

Etapas	Meses			
	Mar	Abr	Mai	Jun
Anteprojeto				
Pesquisa bibliográfica sobre doenças respiratórias crônicas, programas de reabilitação pulmonar e <i>machine learning</i>				
Estudar linguagem R e suas bibliotecas				
Estudar métodos de machine learning e análise preditiva				
Estudar técnicas para geração de visualizações				
Elaborar TC I				

Trabalho de Conclusão II

Etapas	Meses			
	Ago	Set	Out	Nov
Preparar os dados para análise				
Aplicar <i>machine learning</i>				
Analisar os conhecimentos extraídos				
Elaborar visualizações				
Disponibilizar ferramenta online				
Elaborar TC II				

BIBLIOGRAFIA

II CONSENSO BRASILEIRO SOBRE DOENÇA PULMONAR OBSTRUTIVA CRÔNICA - DPOC. Jornal Brasileiro de Pneumologia, São Paulo, v. 30, sup. 5, p. S1-S42, nov. 2004. Disponível em: <http://www.jornaldepneumologia.com.br/pdf/suple_124_40_dpoc_completo_finalimpresso.pdf>. Acesso em: 20 março 2018.

ALPAYDIN, Ethem. **Introduction to machine learning.** Massachusetts: The MIT Press, 2010. 2° ed.

DUCHIADE, Milena. **Poluição do ar e doenças respiratórias: uma revisão.** Caderno de Saúde Pública. 8(3): 311-330, julho/setembro de 1992. Disponível em: <<https://www.scielosp.org/pdf/csp/1992.v8n3/311-330/pt>>. Acesso em: 20 março 2018.

FILHO, João Terra. **Avaliação laboratorial da função pulmonar**, v. 31, p. 191-207. In: SIMPÓSIO DE DOENÇAS PULMONARES. Ribeirão Preto: USP, abr./jun. 1998. Disponível em: <http://revista.fmrp.usp.br/1998/vol31n2/avalicao_%20laboratoria_funcao_pulmonar.pdf>. Acesso em: 15 março 2018.

GILMAN, Sean A.; BANZETT, Robert. **Physiologic changes and clinical correlates of advanced dyspnea.** Current Opinion in Supportive & Palliative Care, Londres, v. 3, n. 2, p. 93-97, jun. 2009. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2834174/>>. Acesso em: 17 março 2018.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: Um Guia Prático.** Rio de Janeiro: Elsevier Editora Ltda., 2005.

MAURER, Janet; REBBAPRAGADA, Venkata; BORSON, Soo; GOLDSTEIN, Roger; KUNIK, Mark; YOHANNES, Abebaw; HANANIA, Nicola. **Anxiety and Depression in COPD.** Chest. 134(4): 43-56, outubro de 2008. Disponível em: <[http://journal.chestnet.org/article/S0012-3692\(08\)60341-4/pdf](http://journal.chestnet.org/article/S0012-3692(08)60341-4/pdf)>. Acesso em: 19 março 2018.

MITCHELL, Tom. **The Discipline of Machine Learning.** Pittsburgh: Carnegie Mellon University, julho de 2006. Disponível em: <<http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>>. Acesso em: 22 março 2018.

NICI, Linda; DONNER, Claudio; WOUTERS, Emiel; ZUWALLACK, Richard; AMBROSINO, Nicolino; BOURBEAU, Jean; CARONE, Mauro; CELLI, Bartolome; ENGELEN, Marielle; FAHY, Bonnie; GARVEY, Chris; GOLDSTEIN, Roger; GOSSELINK, Rik; LAREAU, Suzanne; MACINTYRE, Neil; MALTAIS, Francois; MORGAN, Mike; O'DONNELL, Denis; PREFALUT, Christian; REARDON, Jane; ROCHESTER, Carolyn; SCHOLS, Annemie; SINGH, Sally; TROOSTERS, Thierry. **American Thoracic Society/European Respiratory Society Statement on Pulmonary Rehabilitation.** American Journal of Respiratory and Critical Care Medicine. v. 173, n° 12, p. 1390-1413, junho de 2006. Disponível em: <<https://www.atsjournals.org/doi/full/10.1164/rccm.200508-1211ST#readcube-epdf>>. Acesso em: 20 março 2018.

POULAIN, Magali; DURAND, Fabienne; PALOMBA, Bernard; CEUGNIET, François; DESPLAN, Jacques; VARRAY, Alain; PRÉFAUT, Christian. **Six-minute walk testing is more sensitive than maximal incremental cycle testing for detecting oxygen desaturation in patients with COPD**. Chest. 123(5): 1401-1407, maio de 2003. Disponível em: <[http://journal.chestnet.org/article/S0012-3692\(15\)33669-2/pdf](http://journal.chestnet.org/article/S0012-3692(15)33669-2/pdf)>. Acesso em: 19 março 2018.

RODRIGUES, Joaquim C.; CARDIERI, Joselina M. A.; BUSSAMRA, Maria Helena C. de F.; NAKAIE, Cleyde M. A.; ALMEIDA, Marina B. de; FILHO, Luiz V. F da S.; ADDE, Fabíola V. **Provas de função pulmonar em crianças e adolescentes**. Jornal Brasileiro de Pneumologia, outubro de 2002. Disponível em: <<http://www.luzimarteixeira.com.br/wp-content/uploads/2010/07/provas-de-funcao-pulmonar-em-criancas-e-adolescentes.pdf>>. Acesso em: 15 março 2018.

SILVA, Fernando Henrique da. **Estudo e desenvolvimento de métodos para predição de doadores de sangue**. 2018. 80f. Dissertação (Mestrado em Modelagem e Otimização) - Universidade Federal de Goiás, Catalão, GO, 2018. Disponível em: <<http://repositorio.bc.ufg.br/tede/handle/tede/8169#preview-link0>>. Acesso em: 24 março 2018.

VIDALE, Giulia. **44% dos brasileiros sofrem com problemas respiratórios**. Revista Veja, São Paulo, agosto 2015. Saúde. Disponível em: <<https://veja.abril.com.br/saude/44-dos-brasileiros-sofrem-com-problemas-respiratorios/>>. Acesso em: 17 março 2018.

WITTEN, Ian H.; FRANK, Eibe. **Data Mining: practical machine learning tools and techniques**. Burlington: Elsevier Inc., 2005. 2° ed.