

UNIVERSIDADE FEEVALE

EDUARDO EISMANN SOUZA

DESCOBERTA DE CONHECIMENTO EM DADOS SOBRE  
COVID-19  
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo  
2020

EDUARDO EISMANN SOUZA

DESCOBERTA DE CONHECIMENTO EM DADOS SOBRE  
COVID-19

(Título Provisório)

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: Dra. Marta Rosecler Bez

Novo Hamburgo  
2020

## RESUMO

A partir do grande volume de infecções e mortes por Covid-19, que se mostrou presente no ano de 2020, é possível perceber a importância de se obter um conhecimento maior sobre o vírus, para auxiliar profissionais de saúde a traçar e prever ações para infecções causadas pelo mesmo. Também adquirir conhecimento e aumentar a percepção de características que possam estar relacionadas à essa e de outras ameaças ao sistema respiratório. Este trabalho tem como objetivo, desenvolver um *dashboard* que possibilite a visualização de dados sobre testes de COVID-19, realizados por três hospitais brasileiros. Para tanto, a metodologia utilizará uma abordagem qualitativa, empregando pesquisa bibliográfica e experimental. Nesse sentido, serão estudadas as características de doenças respiratórias, bem como a análise de exames diversos em conjunto com o pré-processamento dos dados coletados. As bases de dados serão unificadas para, então, com o auxílio da linguagem de programação R, promover a construção de filtros de seleção para pacientes e exames. Por fim, desenvolvendo o *dashboard* que ao ser concluído passará por validação de especialistas na área da saúde.

Palavras-chave: *Data Mining*. Descoberta de Conhecimento. Análise de Dados. Covid-19. *Dashboard*.

## SUMÁRIO

MOTIVAÇÃO .....	5
OBJETIVOS .....	8
METODOLOGIA .....	9
CRONOGRAMA .....	11
BIBLIOGRAFIA .....	12

## MOTIVAÇÃO

Por meio de uma pesquisa recém-divulgada pelo Ibope, foi mostrado que 44% dos brasileiros apresentam sintomas de doenças respiratórias, sendo na sua maioria, asma e bronquite crônica. O trabalho foi realizado com 2010 brasileiros com idade entre 18 e 65 anos, em todas as regiões do país. O estudo revelou que 65% da prevalência dos sintomas respiratórios são localizados nos estados do sul do país.

O pneumologista e professor na Universidade Federal de São Paulo (Unifesp), Clystenes Odyr Soares, atribui essa característica ao clima da região, que é caracterizada por invernos mais rigorosos e secos, em comparação com as demais regiões do Brasil. Soares explica que a incidência de temperaturas mais baixas e a pouca umidade relativa do ar, são vistos como um grande risco para o sistema respiratório (VIDALE, 2015).

Em estudo publicado por Iuliano *et al.* (2017, tradução nossa) para o site The Lancet, segundo estimativas realizadas pelo Centro de Controle e Prevenção de Doenças Respiratórias dos Estados Unidos (US-CDC) e da Organização Mundial de Saúde (OMS), é possível reforçar o argumento de Soares, onde associa-se mais de 450 mil mortes por ano no mundo ligadas a gripe sazonal.

Outra doença que também ataca o sistema respiratório, deixou marcado o ano de 2020 em todo o mundo, o Coronavírus (Covid-19). Registrado inicialmente como uma grave pneumonia, a doença que teve início na China se alastrou e contaminou pessoas nos cinco continentes (GONÇALVES, 2020). No dia 11 de março de 2020, a Organização Mundial de Saúde declarou em uma comitiva de imprensa, que classifica os casos de Covid-19 como uma pandemia. O vírus Covid-19, é o primeiro do tipo coronavírus a receber essa classificação (HEIN, 2020, tradução nossa).

Em notícia publicada por Farge (2020), para o site Agência Brasil, diferente de doenças respiratórias que apresentam tendências maiores em determinadas épocas do ano, segundo a Dra. Margaret Harris, porta-voz da Organização Mundial de Saúde, o Covid-19 não é caracterizado como uma doença respiratória sazonal, diferente da característica observada por Iuliano *et al.* (2017, tradução nossa) em outras doenças respiratórias.

Segundo estimativa feita pela British Broadcasting Corporation (BBC), entre 4 de janeiro e 15 de maio de 2020, houve cerca de 440 mil óbitos em todo o mundo por infecção de Covid-19, estimou-se também que, além dos dados oficiais, ainda há aproximadamente

130 mil mortes relacionadas, visto que muitos países divulgam apenas os dados de óbitos ocorridos dentro de hospitais (DALE; STYLIANOU, 2020).

De acordo com a notícia publicada em 31 de agosto de 2020, pelo jornal português Público, em relatório da OMS é descrito que o impacto do Covid-19 abalou os sistemas de saúde em todo o mundo, com severas interrupções em serviços essenciais de saúde em quase todos os países. A OMS classifica como serviços essenciais de saúde condições crônicas, saúde mental, reprodutiva, materna e infantil (PÚBLICO, 2020, tradução nossa).

O volume de dados gerados no mundo está aumentando, e com o passar do tempo tende a crescer cada vez mais. Quanto maior a quantidade de informação, mais difícil se torna o entendimento das pessoas sobre esses dados, que podem conter informações e referências potenciais no que diz respeito a descoberta de novas características em um ambiente. Desde o início da vida humana, as pessoas buscam padrões sobre os dados, de forma a obter benefícios. A descoberta de padrões em um conjunto de dados pode trazer explicações e auxiliar na predição de dados futuros (WITTEN; FRANK, 2005, tradução nossa).

A extração de conhecimento sobre bases de dados pode gerar informações extremamente úteis e proporcionar benefícios para quem as analisa. Goldschmidt e Passos (GOLDSCHMIDT; PASSOS, 2005, p. 7) 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.

Para fim de obter melhor visualização e clareza de informações, (SECAF, 2019), em matéria publicada no site Setting Consultoria, argumenta sobre a utilização da ferramenta de *dashboard* no meio empresarial, salientando que seu uso é um recurso extremamente útil para a tomada de decisão, proporcionando uma visão mais dinâmica e objetiva.

Por fim, este trabalho propõe o desenvolvimento de um *dashboard* onde, a partir de bases de dados de três hospitais brasileiros, fornecidos pela iniciativa "COVID-19 Data Sharing/BR", da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), fazer um comparativo de informações e características de exames realizados por pacientes, a fim de detectar a presença do Covid-19. O intuito principal do uso do *dashboard* é auxiliar profissionais da área da saúde a identificar padrões e semelhanças entres os casos que estejam com resultados positivos para Covid-19.

Este trabalho utilizará dados disponibilizados pelo repositório COVID-19 Data Sharing/BR, disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

## OBJETIVOS

Objetivo geral:

Desenvolver um *dashboard* que torne possível a visualização de dados sobre testes de COVID-19.

Objetivos específicos:

- Estudar a base de exames de COVID-19 de diversos laboratórios
- Unir as bases de dados em uma única base
- Explorar técnicas de mineração de dados
- Desenvolver o *dashboard*
- Submeter o *dashboard* à avaliação de especialista na área de saúde

## METODOLOGIA

A figura abaixo exibe os conceitos aplicados à metodologia desta pesquisa, em vermelho.

Figura 1 – Classificação da pesquisa



Fonte: Adaptado de Bez (2011).

Quanto à natureza do presente trabalho, é caracterizada como uma pesquisa aplicada, visto que será criado um *dashboard* para demonstrar características de dados recolhidos de três hospitais brasileiros.

A forma de abordagem será qualitativa, dado que os resultados serão interpretados e comparados com as saídas provenientes do *dashboard*, estas, por sua vez, serão submetidas à avaliação e interpretação de especialistas, onde não há necessidade de quantificar, mas sim identificar a qualidade das informações apresentadas.

O objetivo do estudo é exploratório, uma vez que é feita a mineração dos dados provenientes de bases nacionais e estes são analisados e comparados entre si, onde será possível expor alguns padrões e semelhanças no que se trata à ocorrência de pacientes infectados.

Sobre os procedimentos técnicos a serem utilizados, como o tema COVID-19 é relativamente novo e pouco se sabe, a pesquisa será baseada em artigos científicos referentes ao tema, mineração de dados em TCCs, artigos e livros, caracterizando-a como uma pesquisa bibliográfica. O presente trabalho também é definido como uma pesquisa experimental, dado que a proposta é o desenvolvimento de um *dashboard* onde se torna possível a busca por padrões em um grupo de pacientes infectados, envolvendo a análise de especialistas na área da saúde para qualificar os resultados, não sendo validado apenas por métodos teóricos.

O desenvolvimento do trabalho é dado pela ordem descrita abaixo:

- Estudo de características de doenças respiratórias;
- Análise de exames diversos;
- Pré-processamento dos dados coletados;
- Unificar as bases de dados;
- Aprofundamento em linguagem R para desenvolvimento da solução visual;
- Construção de filtros de seleção para pacientes e exames;
- Desenvolvimento do *dashboard*;
- Validação com especialistas na área de saúde.

Este trabalho utilizará dados disponibilizados pelo repositório COVID-19 Data Sharing/BR, disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

## CRONOGRAMA

### Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Escrita do anteprojeto				
Revisão do anteprojeto				
Entrega do anteprojeto				
Estudo focado em COVID-19				
Análise e pré-processamento de exames				
Análise de trabalhos correlatos				
Aprofundamento em linguagem R				
Desenvolvimento do <i>dashboard</i>				
Redação do TCC 1				
Revisão do TCC 1				
Entrega do TCC 1				

### Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
Desenvolvimento do <i>dashboard</i>				
Teste do protótipo com profissionais da área				
Avaliação dos resultados				
Redação do TCC 2				
Revisão do TCC 2				
Entrega do TCC 2				
Apresentação dos resultados à banca avaliadora				

## BIBLIOGRAFIA

BEZ, M. R. **Uso de tecnologia para apoiar a implantação de métodos ativos nos currículos de medicina.** Proposta de Tese. Centro Interdisciplinar de Novas Tecnologias na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2011.

DALE, Becky.; STYLIANOU, Nassos. **Coronavírus: como o 'excesso de mortes' pode revelar o verdadeiro número de vítimas da pandemia de covid-19.** BBC News Brasil, 18 junho 2020. Disponível em: <<https://www.bbc.com/portuguese/internacional-53092095>> Acesso em: 30 ago. 2020.

FARGE, Emma. **OMS diz que pandemia de covid-19 é "uma grande onda", não é sazonal.** Agência Brasil, Genebra, julho 2020. Disponível em: <<https://agenciabrasil.ebc.com.br/internacional/noticia/2020-07/oms-diz-que-pandemia-de-covid-19-e-uma-grande-onda-nao-e-sazonal>>. Acesso em: 26 agosto 2020.

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

GONÇALVES, Siumara. **Da pneumonia na China à pandemia, o caminho do coronavírus até o ES.** Jornal A Gazeta, Vitória, março 2020. Disponível em: <<https://www.agazeta.com.br/es/gv/da-pneumonia-na-china-a-pandemia-o-caminho-do-coronavirus-ate-o-es-0320>>. Acesso em: 26 agosto 2020.

HEIN, Alexandra. **WHO declares coronavirus global 'pandemic'.** Fox News, Nova Iorque, 11 março 2020. Disponível em: <<https://www.foxnews.com/health/who-declares-coronavirus-global-pandemic>>. Acesso em: 2 setembro 2020.

Iuliano et al. **Estimates of global seasonal influenza-associated respiratory mortality: a modelling study.** The Lancet, dezembro 2017. Disponível em: <[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)33293-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)33293-2/fulltext)> Acesso em: 5 ago. 2020.

PRODANOV, Cleber C.; FREITAS, Ernani César de. **Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2ª edição. Universidade Feevale, Novo Hamburgo, RS, 2013.

PÚBLICO. **Covid-19: A pandemia abalou sistemas de saúde de todo o mundo, diz a OMS**. Portugal, Lisboa: Coronavírus, 2020. Disponível em: <<https://www.publico.pt/2020/08/31/mundo/noticia/covid19-pandemia-abalou-sistemas-saude-mundo-oms-1929833>>. Acesso em: 2 setembro 2020.

SECAF, Vera Maria Stuart. **Dashboard para empresas: o que é e 4 exemplos para se inspirar**. Setting Consultoria, 3 dezembro 2019. Disponível em: <<https://www.setting.com.br/blog/gestao-empresarial/dashboard-para-empresas/>>. Acesso em: 3 setembro 2020.

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: 1 agosto 2020.

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