



ANÁLISE DE SINTOMAS E COMORBIDADES DE INFECTADOS PELO COVID-19 NO ESTADO DE SANTA CATARINA NO PRIMEIRO ANO DE PANDEMIA

SAMUEL DA SILVA FEITOSA; CRISTIANO MESQUITA GARCIA; JEAN MARCEL DE ALMEIDA ESPINOZA; VIVIAN DA SILVA CELESTINO REGINATO; GUSTAVO BOURDOT BACK

RESUMO

Em março de 2020 a Organização Mundial da Saúde (OMS) declarou a COVID-19, doença causada pelo novo Coronavírus (SARS-CoV-2), como pandemia global. Além do contexto de infecção, logo se notou que a associação de comorbidades elevava os riscos de mortalidade pela doença, e que os casos mais graves se correlacionavam à sintomas característicos. Diante deste cenário, tornou-se necessário monitorar e entender a doença e suas particularidades, de forma a orientar a estratégia mais adequada de saúde pública. Neste sentido o presente estudo tem por objetivo analisar os dados de infectados por COVID-19 no Estado de Santa Catarina, no primeiro ano de pandemia, e relacionar os casos confirmados à presença de comorbidades e/ou sintomas que pudessem conduzir a internação ou óbito. Trata-se de um estudo epidemiológico descritivo, retrospectivo, baseado na análise de dados disponibilizados pelo Governo do Estado de Santa Catarina e, no uso de técnicas de mineração de dados e de inteligência computacional para extração de informações e verificação de potenciais relações. Verificou-se que, isoladamente, os sintomas de tosse, febre e dor de garganta foram mais comuns nos infectados, sendo o primeiro frequente em mais de 40% dos casos. Em relação aos casos de internação e de óbitos, observou-se que os sintomas mais frequentemente observados foram a dispneia, a tosse e a associação entre dispneia e tosse. Quanto às comorbidades, constatou-se que doenças cardiovasculares crônicas, diabetes e obesidade foram as mais identificadas em contaminados. Isoladas ou associadas entre si, estas comorbidades também foram as mais observadas nos casos de internação e de óbitos. Por fim, percebeu-se que, apesar das limitações entre a amostra de dados e o desempenho das técnicas empregadas, a metodologia demonstra o potencial para obtenção de informações de apoio à gestão da saúde pública.

Palavras-chave: Inteligência Computacional, Aprendizagem de Máquina, Mineração de Dados; Coronavírus

1 INTRODUÇÃO

No fim de 2019 uma doença respiratória aguda se espalhou rapidamente ao redor do mundo, se evidenciando inicialmente na China. Devido a facilidade de transmissão, a severidade dos sintomas associados e outras complicações, a OMS declarou, em 11 de março de 2020, a COVID-19 como pandemia global. Todo este cenário levou a uma brusca mudança de hábitos, incluindo: maior frequência de lavagem de mãos, uso extensivo de álcool em gel e uso de máscaras (FEITOSA *et al.*, 2022).

Diversas comorbidades, como problemas renais, diabetes, pressão alta, etc., têm sido relacionadas a um maior índice de mortalidade por COVID-19 e de acordo com Feitosa *et al.* (2022), este fato impacta diretamente na importância do monitoramento e entendimento da doença, pois permite tanto à equipe multiprofissional realizar o tratamento mais indicado ao

paciente, considerando suas particularidades e vulnerabilidade, quanto aos gestores adotarem recomendações de saúde pública conforme os riscos associados a determinadas comorbidades.

Estudos também tem evidenciado a correlação entre os casos graves de COVID-19 (os que causam internação e óbitos) à apresentação de sintomas, como febre, dor de cabeça, falta de ar, cansaço, entre outros. A exemplo do que ocorre no mundo, em Santa Catarina foi verificado que mais de 82% dos pacientes que morreram devido ao Coronavírus no primeiro ano de pandemia, apresentaram, pelo menos, um sintoma (ESTADO DE SANTA CATARINA, 2021a).

Segundo Feitosa *et al.* (2022), Técnicas de Mineração de Dados (TMD) vêm sendo utilizadas em diversas áreas para extrair informações relevantes de conjuntos de dados. Essas técnicas normalmente se encontram dentro de processos maiores, denominados *Knowledge Discovery in Databases* (KDD), pois provém um *framework* para extrair padrões úteis e desconhecidos de conjuntos de dados que são utilizados em algoritmos de aprendizado de máquina e inteligência computacional (GOLDSCHMIDT *et al.*, 2015).

Neste sentido, este trabalho tem por objetivo analisar os dados de infectados pela COVID-19 no Estado de Santa Catarina, no primeiro ano de pandemia, utilizando técnicas de mineração de dados de forma a relacionar os casos confirmados à apresentação de comorbidades e/ou sintomas que pudessem levar a internação ou óbito.

2 MATERIAIS E MÉTODOS

Este é um estudo epidemiológico descritivo, retrospectivo, baseado na análise de dados anonimizados extraídos do portal de Dados Abertos do Estado de Santa Catarina. Destes dados foram filtrados os inseridos no período entre 01/03/2020 e 01/04/2021. O conjunto de dados é composto por 41 atributos, entre eles: *data_publicacao*, *recuperados*, *comorbidades*, *sintomas*, *data_inicio_sintomas*, *data_coleta*, *regional*, *idade*, *sexo*, entre outros ((ESTADO DE SANTA CATARINA, 2021b).

Após a obtenção e tratamento dos dados, análises utilizando mineração de dados e técnicas de inteligência computacional foram aplicadas, como forma de extrair informações úteis e verificar se existiam potenciais relações entre comorbidades, sintomas, internação e óbitos. Todo o processo foi realizado utilizando *Python* 3.8, em conjunto com as bibliotecas *Pandas*, *Apyori* e *Seaborn*. A plataforma de tratamento utilizada foi o *Google Colab*.

3 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados obtidos e as análises realizadas, que podem ser replicadas também para outros estados, considerando a disponibilidade de dados similares, com pequenas modificações na etapa de tratamento.

3.1 Análise de Sintomas

Iniciou-se a análise a partir da identificação estatística dos sintomas mais presentes nas pessoas diagnosticadas com a COVID-19, como *tosse*, *febre* e *dor de garganta*. Na Figura 1 o sintoma *tosse* chama a atenção por ser frequente em mais de 40% dos casos.

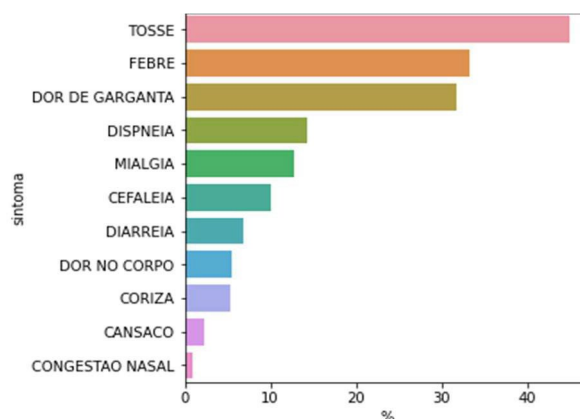


Figura 1 - Sintomas mais frequentes em pessoas contaminadas por COVID-19 em SC

Com base nos dados de sintomas, foi aplicado o algoritmo *Multi-layer Perceptron* (MLP) como uma forma de identificar possíveis interações com base nos sintomas apresentados. Devido ao desbalanceamento do conjunto de dados (95% de dados para não-internação e 5% para internação), o algoritmo não encontrou bons resultados capazes de refletir quais sintomas podem levar à internação. Dessa forma, optou-se ainda por utilizar a aplicação de técnicas de *undersampling* e a análise resultou em uma acurácia de 85%, porém com precisão de 20% para a classe minoritária (internação).

Foi analisada, também, a associação entre os sintomas apresentados pelos pacientes, utilizando a técnica *Apriori*, considerando três cenários distintos: a associação entre sintomas por si só, a relação entre sintomas com internação, e a relação entre os sintomas e os pacientes que morreram. A Figura 2 mostra as associações identificadas entre os sintomas apresentados pelos pacientes, onde as principais relações encontradas foram *tosse e dispneia*, presentes em 14% dos casos; *febre, tosse e dor de garganta*, presentes em cerca de 14% dos casos; e *febre e dispneia*, que apareceram juntas em cerca de 11% dos casos.

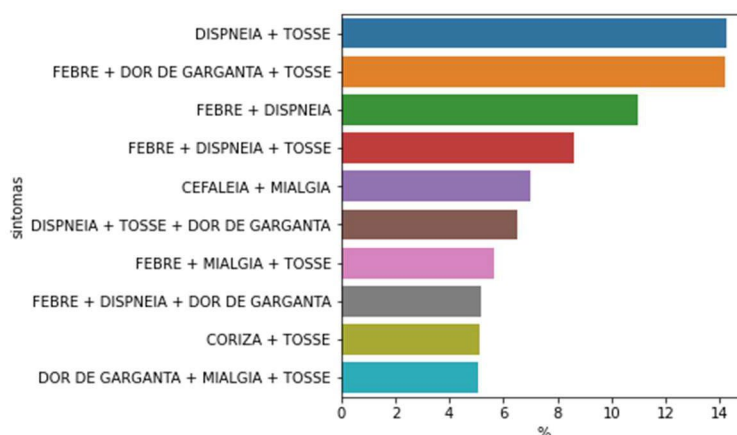


Figura 2 - Sintomas associados mais frequentes em contaminados por COVID-19 em SC Na Figura 3 é possível visualizar o resultado obtido pela aplicação do algoritmo *Apriori*, relacionando os sintomas dos pacientes e a internação. Nesse caso, identificou-se que *dispneia*, *tosse* e *febre* são mais frequentes em pacientes que precisaram de internação, correspondendo a cerca de 5,8%, 5,5% e 4,5%, respectivamente. É possível perceber que o conjunto formado é composto pelos sintomas mais frequentes e isso é um efeito do *Apriori*. Para essa análise, os parâmetros utilizados foram: suporte de 2% e confiança de 30%.

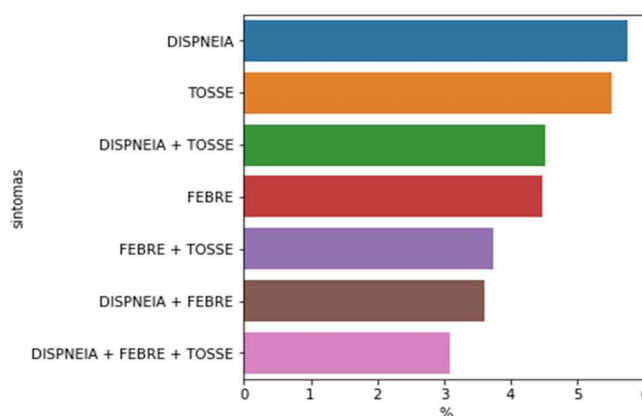


Figura 3 - Sintomas associados mais presentes em internados por COVID-19 em SC

Por fim, na Figura 4 é apresentada a análise que relaciona sintomas aos óbitos de pacientes, onde o algoritmo *Apriori* identificou os sintomas *dispneia*, *tosse* e a combinação de *dispneia* e *tosse* como mais relacionados aos pacientes que tiveram a sua vida interrompida pela doença.

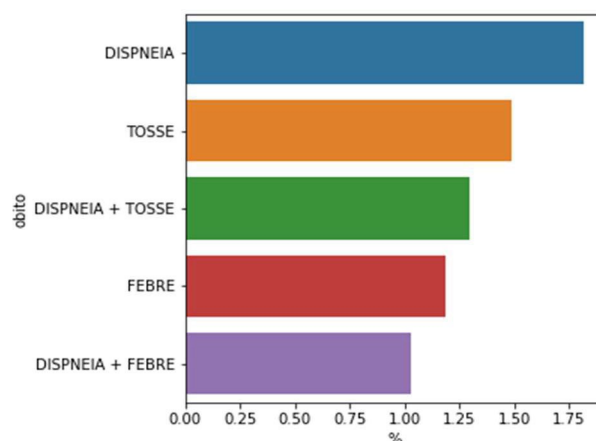


Figura 4 - Sintomas associados mais presentes em casos de óbito por COVID-19 em SC

3.2 Análise de Comorbidades

A análise estatística a respeito das comorbidades das pessoas com COVID-19 indica uma maior frequência de doença cardiovascular crônica, seguida por diabetes, e logo após, obesidade, conforme pode ser visto na Figura 5.

De forma similar ao que foi apresentado para os sintomas, foi realizada análise de comorbidades. A partir da aplicação do algoritmo classificador MLP, foi obtida uma acurácia geral de 77%. Entretanto, outras métricas como *recall* e *f1-score* apresentaram desempenho insatisfatório, considerando apenas dados das comorbidades, mesmo utilizando técnicas de *undersampling*. De forma similar ao ocorrido com os sintomas, é possível afirmar que apenas com os dados abertos fornecidos pelo estado, utilizando as informações de comorbidades não é possível prever com assertividade se os pacientes necessitarão de internação.

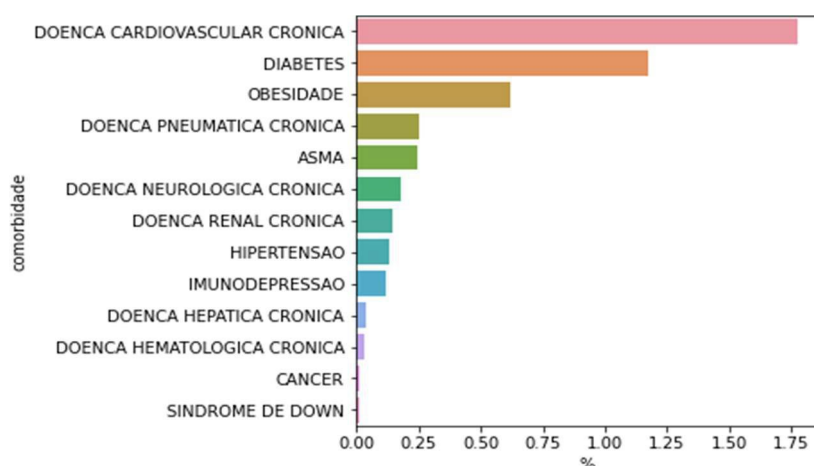


Figura 5 - Comorbidades mais frequentes em infectados por COVID-19 em SC

Com relação às comorbidades, também foram realizadas análises de associação entre as comorbidades dos pacientes, utilizando a técnica *Apriori*. Similarmente à seção anterior, foram considerados três cenários: a associação entre comorbidades por si só, a relação entre comorbidades e internação, e, por fim, a relação entre as comorbidades e os óbitos de pacientes. A Figura 6 apresenta um gráfico que identifica associações entre as comorbidades das pessoas diagnosticadas com a COVID-19, onde as principais relações encontradas foram *doença cardiovascular crônica* e *diabetes* e *obesidade*; e *doença pneumática crônica* e *diabetes*. Chama a atenção a frequência de *doença cardiovascular crônica* e *diabetes* e *obesidade*, aparecendo de forma simultânea em cerca de 4% dos casos. Os parâmetros utilizados foram: confiança mínima de 30%; suporte mínimo de 2%.

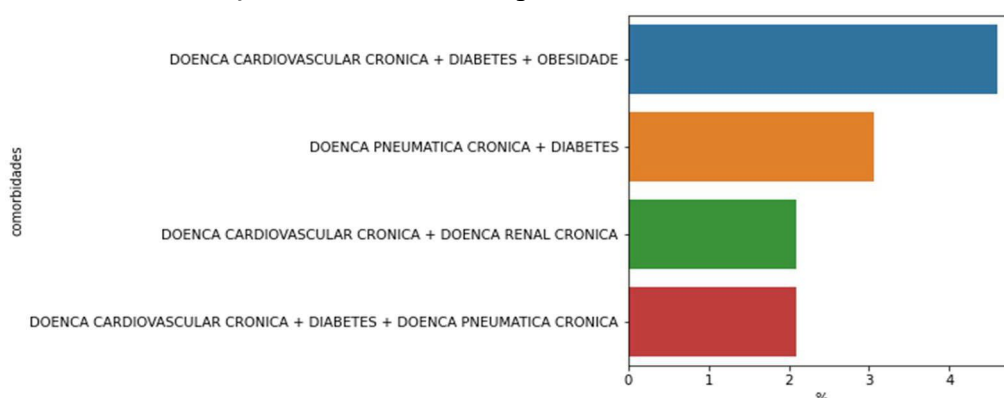


Figura 6 - Comorbidades mais frequentes em infectados por COVID-19 em SC

Na Figura 7 é possível visualizar o resultado obtido pela aplicação do algoritmo *Apriori* relacionando comorbidades com internação. Nesse caso, identificou-se que as comorbidades *doença cardiovascular crônica* e *diabetes* possuem mais possibilidades de levar um infectado à internação. Cerca de 35% dos internados possuíam *doença cardiovascular crônica*. Os parâmetros para essa análise foram: suporte mínimo: 6% e confiança mínima: 30%.

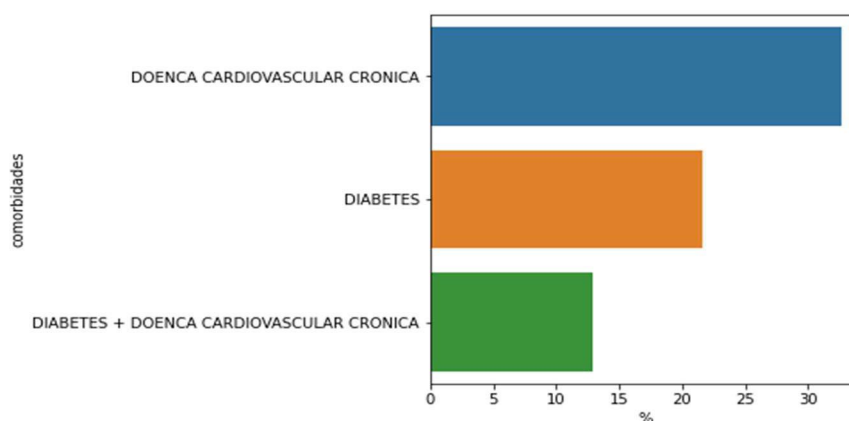


Figura 7 - Comorbidades mais frequentes em internados por COVID-19 em SC

Por fim, o algoritmo *Apriori* foi aplicado para encontrar relações entre as comorbidades e os óbitos em decorrência desse vírus. Conforme a Figura 8, as comorbidades mais associadas a óbitos são *doença cardiovascular crônica*, *diabetes* e *doença pneumática crônica*. Chama a atenção a ocorrência simultânea de *diabetes* e *doença cardiovascular crônica*, em mais de 8% dos casos. Parâmetros utilizados: confiança mínima de 30% e suporte mínimo de 3%.

A partir das análises realizadas sobre as comorbidades de pessoas infectadas pela COVID-19, é possível notar claramente que as comorbidades que mais tornam o humano vulnerável à COVID-19 são *doença cardiovascular crônica* e *diabetes*.

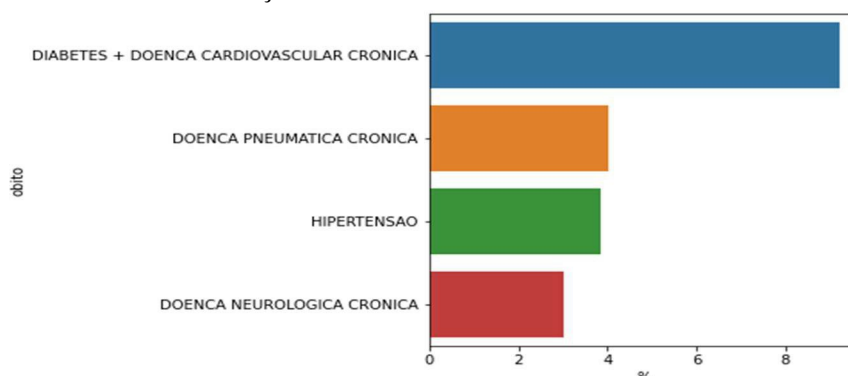


Figura 8 - Comorbidades mais frequentes em casos de óbito por COVID-19 em SC

4 CONCLUSÃO

Esse trabalho apresentou diversas análises sobre os dados abertos relativos à COVID-19 no estado de Santa Catarina no primeiro ano de pandemia, entre 04/11/2020 e 01/04/2021, utilizando técnicas de mineração de dados e inteligência computacional na identificação de relações entre sintomas e comorbidades de cidadãos infectados pela COVID-19. Como pôde ser percebido nos resultados, o uso da técnica *Apriori* apresenta uma abordagem interessante para destacar sintomas e comorbidades que mais se relacionam com internações e óbitos de pessoas. Essas informações são muito importantes no sentido de permitir às autoridades a criação de políticas específicas para o conjunto da população que se apresenta mais suscetível à COVID-19. Além disso, também foi possível perceber que, considerando apenas os dados disponibilizados pelo governo do estado, algoritmos de aprendizagem de máquina não conseguem prever com desempenho conveniente possíveis internações e óbitos. Isto indica que são necessárias maiores informações acerca de cada paciente, para que esta predição ocorra de forma mais eficiente.

Como trabalhos futuros, recomenda-se identificar outros fatores ou atributos que possam ser relevantes para serem incorporados aos conjuntos de dados utilizados para predição

de internação e óbitos a partir de algoritmos de aprendizagem de máquina.

Como pôde ser notado nas análises realizadas, apesar da limitação dos dados apresentados dos pacientes e sintomas, é possível realizar associações automaticamente a partir de inteligência computacional, trazendo valiosas informações para que as autoridades tomem medidas preventivas e mantenham a população informada.

REFERÊNCIAS

ESTADO DE SANTA CATARINA, Secretaria do Estado de Saúde. **Estado atualiza os dados sobre as variantes do Coronavírus**. 2021a. Disponível em <<https://www.saude.sc.gov.br/index.php/noticias-geral/12252-estado-atualiza-os-dados-sobre-as-variantes-do-coronavirus>>. Acesso em Maio de 2021.

ESTADO DE SANTA CATARINA, Secretaria do Estado de Saúde. **Metadados - COVID-19 - Casos confirmados**. 2021b. Disponível em <<http://dados.sc.gov.br/dataset/1ba1801e-aec0-4dba-ae2a-7732f0a0c9f7/resource/2128b7d4-7799-4306-ba47-4fcbb7c197f2/download/covid-19-casos-confirmados-dicionario-de-dados.pdf>>. Acesso em maio de 2021.

FEITOSA, S. S.; GARCIA, C. M.; ESPINOZA, J. M. A.; REGINATO, V. S. C. Análise de sintomas e comorbidades de infectados pela COVID-19 de Santa Catarina por meio de técnicas de mineração de dados. **Panorama da COVID-19 no Brasil**. Organizadores: Miguel Albuquerque; Tiago Gandra. Curitiba: CRV, 2022. ISBN: 978-65-251-2970-9. 656 p. P: 575-588.

GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data Mining**. Elsevier Brasil, 2015.