

Processamento de linguagem natural não supervisionado na identificação de pacientes suspeitos de infecção por COVID-19

Unsupervised natural language processing in the identification of patients with suspected COVID-19 infection

Procesamiento del lenguaje natural no supervisado para identificar a los pacientes sospechosos de infección por COVID-19

Rildo Pinto da Silva ¹
Juliana Tarossi Pollettini ¹
Antonio Pazin Filho ¹

doi: 10.1590/0102-311XPT243722

Resumo

Os pacientes com síndrome pós-COVID-19 se beneficiam de programas de promoção de saúde e sua rápida identificação é importante para a utilização custo efetiva desses programas. Técnicas tradicionais de identificação têm fraco desempenho, especialmente em pandemias. Portanto, foi realizado um estudo observacional descritivo utilizando 105.008 autorizações prévias pagas por operadora privada de saúde com aplicação de método não supervisionado de processamento de linguagem natural por modelagem de tópicos para identificação de pacientes suspeitos de infecção por COVID-19. Foram gerados seis modelos: três utilizando o algoritmo BERTopic e três modelos Word2Vec. O modelo BERTopic cria automaticamente grupos de doenças. Já no modelo Word2Vec, para definição dos tópicos relacionados a COVID-19, foi necessária análise manual dos 100 primeiros casos de cada tópico. O modelo BERTopic com mais de 1.000 autorizações por tópico sem tratamento de palavras selecionou pacientes mais graves – custo médio por autorizações prévias pagas de BRL 10.206 e gasto total de BRL 20,3 milhões (5,4%) em 1.987 autorizações prévias (1,9%). Teve 70% de acerto comparado à análise humana e 20% de casos com potencial interesse, todos passíveis de análise para inclusão em programa de promoção à saúde. Teve perda importante de casos quando comparado ao modelo tradicional de pesquisa com linguagem estruturada e identificou outros grupos de doenças – ortopédicas, mentais e câncer. O modelo BERTopic serviu como método exploratório a ser utilizado na rotulagem de casos e posterior aplicação em modelos supervisionados. A identificação automática de outras doenças levanta questionamentos éticos sobre o tratamento de informações em saúde por aprendizado de máquina.

Correspondência

R. P. Silva
Rua Aimberê 233, apto. 21, São Paulo, SP 05018-010, Brasil.
rildo.silva@alumni.usp.br

¹ Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil.

COVID-19; Processamento de Linguagem Natural; Atenção à Saúde; Critérios de Seleção de Pacientes; Instituições Privadas de Saúde



Introdução

A pandemia de COVID-19 ¹ reforçou a preocupação histórica de pesquisadores quanto à ameaça de novos vírus e mutação dos existentes. Isso implicou pressão sobre serviços de saúde já sobrecarregados ², por formas graves da doença (aproximadamente 25% dos pacientes vulneráveis ou com comorbidades) e alta taxa de letalidade (5,6% na primeira onda ³). Adicionalmente, alterações estruturais nos serviços de saúde, maior impacto em países de baixa e média renda ⁴, conflitos éticos na priorização do cuidado ⁵ e desafios financeiros acentuaram seu impacto. Essas dificuldades foram, ainda, agravadas pelo surgimento da COVID-19 longa ou síndrome pós-COVID-19 ^{6,7}, que acomete de 10% a 30% dos pacientes ⁸. É esperado que novas pandemias surjam no futuro ⁹ e a identificação precoce de pacientes será importante para a adoção correta e custo-efetiva de cuidados.

O tratamento de informações é um desafio, pelo seu volume crescente ¹⁰ ou pelas peculiaridades das diferentes áreas do conhecimento. Na saúde, os dados são incompletos, heterogêneos, multidimensionais, não estruturados e imprecisos ^{11,12}. Para tratamento desses problemas, foi proposta a descoberta de conhecimento através dos dados KDD (*knowledge discovery in database*) na mineração (*data mining*) de grandes volumes de dados (*big data*) ^{13,14}.

Técnicas de aprendizado de máquina (ML – *machine learning*) permitem que o algoritmo aprenda padrões não identificáveis por técnicas para classificação ou predição ¹⁵. Esse aprendizado pode ser supervisionado – quando existem rótulos que classifiquem o objeto de estudo – ou não supervisionado – a classificação não existe. Nesse caso, utilizam-se técnicas exploratórias para criação dos rótulos e posterior aplicação de técnicas supervisionadas ¹⁵. A rotulagem dos dados médicos é difícil e depende de trabalho especializado, sendo um fator limitante nos estudos da pandemia ¹⁶. Assim, técnicas exploratórias não supervisionadas são um passo importante na aplicação de ML em grandes volumes de dados para a descoberta de conhecimento.

A mineração de dados textuais (*text mining*) se resume à descoberta de padrões, como proposto por Fayyad et al. ¹⁰, enquanto o processamento de linguagem natural (NLP – *natural language processing*) é tido como ramo da inteligência artificial que lida com a linguagem humana ¹⁷ ou torna essa linguagem compreensível aos computadores ¹⁸, permitindo diferentes abordagens, entre elas o agrupamento de textos por tópicos (*topic modeling*). Os tópicos são grupos de objetos similares entre si, tratando-se de um caso particular de *clusterização*.

As operadoras de planos de saúde processam dados necessários à coesão regulatória ¹⁹ e assistencial. Entre eles, a autorização prévia é o processo de verificação da elegibilidade dos pacientes e da coerência entre a doença e o tratamento. Ela é solicitada antes do atendimento assistencial. Esse processo é regulado indiretamente pela Agência Nacional de Saúde Suplementar (ANS) através da garantia de prazos de atendimento ²⁰.

A análise das autorizações prévias oferece oportunidade de seleção precoce dos pacientes. Contudo, devido ao sigilo médico, não há informações sobre a Classificação Internacional de Doenças, 10ª revisão (CID-10), os procedimentos assistenciais solicitados não permitem a correlação correta com a doença a ser tratada e as informações complementares das autorizações prévias não são estruturadas. Há, portanto, oportunidade para soluções inovadoras na identificação de pacientes em operadoras de planos de saúde no Brasil. Trata-se de um importante setor econômico que cobre aproximadamente 25% da população brasileira com gastos equivalentes a 5,7% do produto interno bruto (PIB) ²¹.

Existem poucos estudos utilizando NLP em saúde no Brasil. Duval et al. ²² construíram um sistema de farmacovigilância utilizando o Twitter para a detecção de eventos adversos causados por medicamentos – usaram como modelo o medicamento doxiciclina para tratamento de malária. Moreira et al. ²³ propuseram um modelo híbrido pelo qual o NLP criou *clusters* de pacientes utilizando dados não estruturados. Esses *clusters* foram incorporados a dados estruturados, melhorando a acurácia do diagnóstico de pacientes com suspeita de demência. Diniz et al. ²⁴ criaram um sistema em celular para identificação de pacientes com ideação suicida que permitiu a quantificação individual do risco momento a momento (*digital phenotyping*), possibilitando a ação dos profissionais de saúde.

Não foram encontrados estudos utilizando dados da saúde suplementar, provavelmente pela dificuldade de acesso a dados nesse setor da saúde, limitados por barreiras de sigilo profissional e comercial. Este estudo preenche essa lacuna e contribui com a aplicação de métodos de ML em softwares livres através de um estudo de caso real.

O objetivo deste artigo é descrever um método de NLP não supervisionado para identificar pacientes suspeitos de infecção por COVID-19, por meio da análise de uma base de dados reais, de autorizações prévias emitidas por operadora de planos privados de saúde da modalidade de autogestão do Estado de São Paulo, Brasil.

Métodos

Desenho e população do estudo

Trata-se de um estudo observacional descritivo, com base em dados secundários de autorizações prévias de uma operadora de planos privados de saúde do Estado de São Paulo da modalidade de autogestão. As autorizações prévias são solicitadas por prestadores de serviços assistenciais ou beneficiários antes da realização de consultas, exames, internações e outros procedimentos eletivos. Os atendimentos de emergência têm liberação automática atendendo às regras da legislação. Para internações é emitida somente uma autorização abrangendo todo o período de internação do paciente. O pagamento da assistência para o prestador somente ocorre mediante apresentação da autorização prévia.

A base de dados estudada é anonimizada; contudo, cada autorização prévia é emitida para um beneficiário específico e há relação biunívoca entre autorização prévia e beneficiário. O método proposto seleciona autorizações que contêm informações sobre suspeita de infecção por COVID-19 e, portanto, considera-se que as autorizações selecionadas representam um paciente suspeito de infecção por COVID-19.

A operadora teve, no período, 29.336 beneficiários expostos, dos quais 14.663 (50%) eram do sexo feminino e 28.820 (98,2%) residiam no Estado de São Paulo. A idade média do grupo foi de 45 anos.

Base de dados e variáveis estudadas

Cada autorização contém um campo de texto livre, “indicacaoclinica”, no qual é indicado o motivo ou justificativa para a solicitação da autorização prévia. O preenchimento desse campo não é obrigatório. O prestador pode apenas anexar documentos justificando o pedido do procedimento. Nesse caso, é comum preencher o campo com texto “em anexo” ou não o preencher. A variável “indicacaoclinica” é a variável de interesse deste estudo.

Foram selecionadas autorizações prévias emitidas entre 1º de setembro de 2019 e 30 de junho de 2022 (n = 742.901). Foram excluídas aquelas sem preenchimento da justificativa (valores faltantes) no campo “indicacaoclinica” (n = 558.530, 75%). Foram incluídas neste estudo, portanto, 184.371 (25%) autorizações prévias das quais 105.008 contêm informação de pagamento. Cada autorização prévia contém pelo menos um evento de atenção à saúde identificado nas variáveis “estruturaevento” e “descricaoevento” correspondendo, respectivamente, ao código do evento solicitado e sua descrição. A autorização é classificada segundo: tipo (“tipotratamento”), regime (“regimeatendimento”) e objetivo de atendimento (“objetivotratamento”). O preenchimento do campo CID-10 não é obrigatório. Elas têm data de validade (“datavalidade”) e podem ser canceladas, reemitidas ou revalidadas de acordo com critérios administrativos da operadora. O Quadro 1 contém as variáveis presentes na base de dados e utilizadas neste estudo.

Processamento de linguagem natural

Foram aplicados dois modelos de NLP – BERTopic (<https://maartengr.github.io/BERTopic/index.html>) e Word2Vec – descritos brevemente a seguir.

Quadro 1

Variáveis da base de dados de autorizações prévias de operadora privada de planos de saúde. São Paulo, Brasil.

VARIÁVEL	DESCRIÇÃO DA VARIÁVEL	TIPO DA VARIÁVEL	TRANSFORMAÇÃO DA VARIÁVEL
autorizacao (autorizações prévias)	Número de autorização prévia para cada procedimento autorizado	Numérica	Não
tipoaautorizacao	Tipo de autorização de acordo com padrão TISS – consulta, remoção, prorrogação, resumo de internação e SADT	Texto	Não
dataautorizacao	Data de emissão da autorização	Data	Não
datavalidade	Data da validade da autorização	Data	Não
cid	Classificação Internacional de Doenças, 10ª revisão (CID-10) – relacionada à autorização informada pelo prestador de serviço solicitante. Não é campo obrigatório	Texto	Não
descricao_cid	Descrição do CID-10 relacionado à autorização	Texto	Não
regimeatendimento	Tipo de instalação utilizada para o atendimento de acordo com classificação da operadora – ambulatorial, domiciliar, hospital dia, internação hospitalar e pronto socorro	Texto	Não
tipotratamento	Tipo do tratamento executado – cirúrgico, clínico, obstétrico, pediátrico, psiquiátrico, odontológico	Texto	Não
objetivotratamento	Objetivo do tratamento – diagnóstico, paliativo, preventivo, reparador, terapêutico	Texto	Não
datasolicitacao	Data da solicitação da autorização prévia;	Data	Não
estruturacaoevento	Código do evento da autorização prévia. O prestador indica o evento de atenção à saúde que deseja realizar, que é analisado pela operadora e autorizado	Texto	Não
descricaoevento	Descrição do evento autorizado	Texto	Não
indicacaoclinica	Campo informado pelo prestador de serviço em que é feita a justificativa para o pedido do procedimento da autorização prévia. Esse campo não é obrigatório e é suportado por outras informações enviadas na forma de anexos. É um campo de texto livre sem nenhum tipo de validação automática	Texto	Sim *

SADT: serviço de apoio diagnóstico terapêutico; TISS: troca de informações na saúde suplementar.

* A transformação dessa variável é descrita no texto do artigo. É a variável de interesse para a análise de processamento de linguagem natural.

- **Modelo BERTopic**

O modelo BERTopic é um algoritmo não supervisionado para modelagem de tópicos baseado em vetores. A modelagem por tópicos é um método de mineração cujo objetivo é descobrir padrões ocultos considerando o contexto e classificar os respectivos textos em grupos similares ^{25,26}, chamados tópicos.

Inicialmente, cada documento, neste caso autorização prévia, é convertido em sua representação vetorial (*word embedding*) usando o modelo *Bidirectional Encoder Representations from Transformers* (BERT – representações de codificadores bidirecionais a partir de transformadores). A dimensionalidade dessa representação é reduzida utilizando-se a técnica *Uniform Manifold Approximation and Projection for Dimension Reduction* (UMAP – aproximação e projeção uniforme de Manifolds para redução de dimensões) e aplicado o algoritmo *Density-Based Clustering Based on Hierarchical Density Estimates* (HDBSCAN – agrupamento por densidade baseado em estimativas de densidade hierárquica) para

criar os tópicos de documentos que são similares semanticamente²⁷. Para a descrição de cada tópico utiliza-se o método frequência do termo/frequência inversa dos documentos (TF-IDF)^{28,29,30,31}. Os documentos não classificados pelo modelo são agrupados em um tópico específico contendo valores discrepantes (*outliers*). Neste trabalho, a aplicação dos métodos deu-se por meio de uma biblioteca gratuita baseada em Python²⁸ chamada BERTopic.

Foram utilizados dois parâmetros para definir a quantidade mínima de autorizações existentes em cada tópico criado: 500 ou mais (BERTopic +500) e 1.000 ou mais (BERTopic +1.000) definidos no parâmetro *min_topic_size* do modelo. Como é um modelo automático, o número total de tópicos criados depende desse parâmetro. O parâmetro *language* foi definido como multilingual para modelagem do texto em português e o modelo de vetorização (*embedding_model*) como all-MiniLM-L6-v2, que é o padrão do modelo.

Para identificação dos tópicos pertencentes a COVID-19, foi utilizado o método *get_topic_info()* do próprio modelo, que gera a descrição automática do tópico.

- **Modelo Word2Vec**

Word2Vec é um modelo de NLP que utiliza redes neurais para aprender a representação de palavras (*word embedding*) em um espaço vetorial de alta dimensão, capaz de capturar o contexto semântico e sintático das palavras em um determinado *corpus* de texto. Para a análise comparativa, foi utilizado o modelo *continuous Bag-of-Words*^{32,33} do algoritmo Word2Vec. Os textos da variável “indicacaoclinica” foram separados em palavras (*tokens*), utilizando-se a biblioteca NLTK (<https://www.nltk.org/>), sobre os quais foi aplicado o algoritmo Word2Vec da biblioteca Gensim (<https://pypi.org/project/gensim/>), sendo utilizados tamanho de vetor igual a 300, recalculados considerando sua média e categorizados em 20 *clusters* utilizando-se o algoritmo K-Means. Esses *clusters* foram considerados os tópicos desse modelo. Esse método não atribui automaticamente nomes aos tópicos. Para identificação dos *clusters* com casos suspeitos de infecção por COVID-19, cada um dos 20 *clusters* foi analisado manualmente pelo pesquisador principal. Para isso selecionou-se, em cada *cluster*, as primeiras 100 autorizações classificadas por ordem decrescente de gasto. Cada texto presente na variável “indicacaoclinica” foi analisado e o respectivo *cluster* classificado, ou não, no grupo COVID-19.

Cada um dos dois modelos foi aplicado sobre as descrições de solicitação de autorização da variável “indicacaoclinica” originais ou tratadas. O tratamento da variável é recomendável para melhorar a performance do modelo Word2Vec.

O tratamento da variável “indicacaoclinica” ocorreu da seguinte forma: conversão de todas as palavras em minúsculas, remoção de *stopwords* em português, exclusão de palavras mais comuns em saúde e exclusão de caracteres especiais. Não foram substituídos acentos ou outras características do português. As palavras COVID-19 e SARS-CoV-2 foram transformadas em *covid*. Também foi realizada a padronização das palavras relacionadas a CID-10 presentes na variável “indicacaoclinica”.

Avaliação da qualidade da classificação gerada pelos modelos

Chegou-se, portanto, a seis tipos diferentes de modelos: BERTopic +500, BERTopic +1.000 e Word2Vec, cada um deles com e sem tratamento de texto da variável “indicacaoclinica” (tratado e não tratado).

Para avaliar a qualidade da classificação, o autor principal analisou o modelo BERTopic +1.000, porque esse apresentou o maior custo médio por autorização. Assim, as primeiras 100 autorizações classificadas como eventos suspeitos ou relacionados à COVID-19 por esse modelo foram ordenadas em ordem decrescente de custo. O texto da “indicacaoclinica” de cada uma dessas autorizações foi analisado manualmente classificando-o em classes de interesse de estudo. Essa classificação manual foi comparada com a classificação automática gerada nesse modelo.

Para comparação com métodos tradicionais de pesquisa de linguagem estruturada SQL (*structured query language*), todas as autorizações prévias contendo as palavras *covid*, *sars*, *coronavirus* e *coronavirüs* maiúsculas ou minúsculas foram selecionadas e confrontadas com os modelos gerados utilizando-se como índice de ligação o número da autorização e identificando se faziam parte dos grupos designados como suspeitos de infecção por COVID-19.

Custo das autorizações prévias

O custo das autorizações prévias corresponde aos gastos assistenciais de cada autorização prévia. A base de pagamento contém as despesas pagas aos prestadores de serviços, líquidas de glosa. Os custos foram obtidos utilizando-se o número da autorização prévia como chave de ligação.

O valor total pago corresponde à soma de todas as despesas no período de setembro de 2019 a julho de 2022, identificadas na base de pagamento para cada autorização prévia. O número de autorizações pagas corresponde à contagem das autorizações com valor de gasto por autorização maior que BRL 0.

O custo médio por autorização paga corresponde à razão entre o gasto com autorizações pelo número de autorizações pagas. Neste estudo, foram considerados casos mais graves aqueles com maior custo médio por autorização prévia. Os gastos são apresentados em reais e sem ajuste inflacionário.

Acesso aos dados foi concedido por meio de um termo de sigilo e cooperação científica com a operadora e aprovado por Comitê de Ética em Pesquisa do Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (protocolo nº 55685722.9.0000.5440).

Resultados

Foram emitidas 742.901 autorizações nos 34 meses analisados, das quais 184.371 (24,9%) foram preenchidas com pelo menos um número ou palavra, fazem parte deste estudo e foram analisadas. Destas, 105.008 foram autorizações pagas (14,1%). O gasto total no período foi de BRL 374.089.836. Esse gasto tem desvio à direita ($D(105.008) = 0,438$, $p = 0,000$ – *skewness* 41,3) (Figura 1)

Os eventos assistenciais mais frequentes nas autorizações analisadas foram: consulta em pronto-socorro (6,1% das autorizações analisadas contêm esse evento), sessão de psicoterapia individual (5,7%) e pesquisa por RT-PCR para COVID-19 (5%). 96,2% das autorizações prévias não tem descrição de CID-10 e somente 587 (0,3%) têm o CID-10 B34.2 – Infecção por coronavírus, não especificada.

A variável “indicacaoclinica” teve 64.917 (35,2%) autorizações com apenas uma palavra ou número e 77,6% das autorizações tiveram até 5 palavras. Após o tratamento da variável, as palavras mais comuns foram “covid”, aparecendo 6.561 vezes, “pronto” (3.821) e “socorro” (3.692). A frase mais longa teve 104 palavras.

Quanto ao tipo de tratamento, 90,7% foram tratamentos clínicos, 7,8% cirúrgicos e 0,3% obstétricos. Quanto ao regime de atendimento, 81% foram ambulatoriais, 16,9% hospitalares e 1% domiciliares. Os atendimentos clínicos internados corresponderam a 15.741 autorizações – 8,5% do total (Tabela 1).

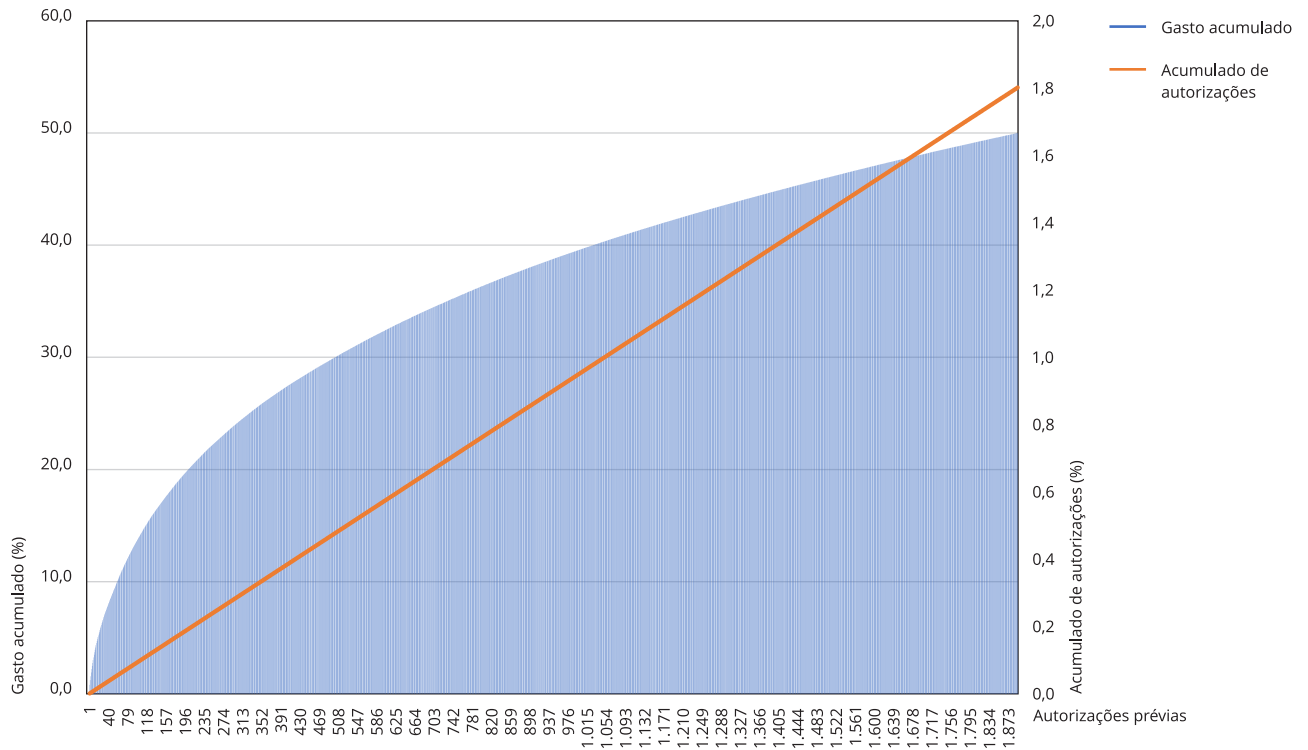
Em relação ao objetivo de atendimento, 75,1% foram para diagnóstico e 6,5% tratamento reparador – 18,3% das autorizações prévias não tiveram objetivo de atendimento preenchido. No regime ambulatorial, o objetivo “diagnóstico” foi mais frequente (80,6%). No regime internação hospitalar, há um grupo importante de “atendimentos reparadores” (34,5%) (Tabela 2).

Nos tópicos classificados como COVID-19, os modelos BERTopic sem tratamento apresentaram maiores custos médios por autorização paga – BRL 10.205 naquele com mais de 1.000 autorizações e BRL 10.138 naquele com mais de 500 autorizações por tópico. Correspondem, respectivamente, a 1,9% (1.987) e 2,3% (2.443) das autorizações pagas, e gastos de BRL 20,3 milhões (5,4% do gasto total) e BRL 24,8 milhões (6,6%). Os dois modelos apresentaram número expressivo de autorizações pagas consideradas discrepantes – 58,8% (61.723) no modelo BERTopic +1.000 e 48,3% (50.716) no modelo BERTopic +500 (Tabela 3).

Com o tratamento da variável “indicacaoclinica”, houve aumento do número de autorizações de casos suspeitos de infecção por COVID-19 no modelo BERTopic +500 autorizações (para 3,3% do total de autorizações pagas) e diminuição no modelo com mais de 1.000 palavras (1,7%) acompanhada de redução importante no total gasto – BRL 5,2 milhões e BRL 14 milhões, respectivamente – quando comparados aos mesmos modelos sem tratamento de palavras, resultando na diminuição dos custos médios por autorização nos dois modelos. Houve diminuição do número de autorizações prévias consideradas discrepantes – embora ainda elevadas (36,3% no modelo BERTopic +1.000 e 45,2% no modelo BERTopic +500) (Tabela 3).

Figura 1

Gasto percentual acumulado (até 50%) e percentual acumulado de autorizações prévias (%) de operadora de saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2022.

**Tabela 1**

Número de autorizações prévias analisadas por tipo de tratamento segundo regime de atendimento de autorizações de operadora de saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2022.

Regime de atendimento	Tipo de tratamento								Total	%
	Clínico		Cirúrgico		Obstétrico		Outros			
	n	%	n	%	n	%	n	%		
Hospitalar	15.741	50,4	13.085	41,9	571	1,8	1.830	5,9	31.227	16,9
Hospital dia *	310	16,1	1.208	62,8	5	0,3	400	20,8	1.923	1,0
Domiciliar	1.789	96,0	47	2,5	0	0,0	28	1,5	1.864	1,0
Ambulatorial	149.351	100,0	1	0,0	0	0,0	4	0,0	149.356	81,0
Pronto-socorro	1	100,0	0	0,0	0	0,0	0	0,0	1	0,0
Total	167.192	90,7	14.341	7,8	576	0,3	2.262	1,2	184.371	100,0

* Hospital dia: o regime de atendimento hospital dia corresponde a um regime intermediário de atendimento entre o atendimento ambulatorial e o regime de internação hospitalar. Considera-se que esse atendimento tenha duração de, no máximo, 12 horas.

Tabela 2

Número de autorizações prévias analisadas por objetivo de tratamento segundo regime de atendimento de autorizações de operadora de saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2022.

Regime de atendimento	Objetivo do atendimento												Total	%
	Sem informação		Diagnóstico		Reparador		Paliativo		Preventivo		Terapêutico			
	n	%	n	%	n	%	n	%	n	%	n	%		
Ambulatorial	28.687	19,2	120.384	80,6	43	0,0	22	0,0	3	0,0	217	0,1	149.356	81,0
Domiciliar	93	5,0	1.626	87,2	59	3,2	-	-	-	-	86	4,6	1.864	1,0
Hospital dia	463	24,1	393	20,4	1.066	55,4	-	-	-	-	1	0,1	1.923	1,0
Internação hospitalar	4.435	14,2	15.996	51,2	10.781	34,5	1	0,0	-	-	14	0,0	31.227	16,9
Pronto-socorro	-	-	-	0,0	1	0,0	-	-	-	-	-	0,0	1	0,0
Total	33.678	18,3	138.399	75,1	11.950	6,5	23	0,0	3	0,0	318	0,2	184.371	100,0

Tabela 3

Modelos e características das autorizações prévias pagas segundo suspeita de infecção por COVID-19 e valores discrepantes de autorizações emitidas por operadora de saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2020.

Modelos	Tópicos (n)	Valores discrepantes *		Tópicos de autorizações suspeitas de infecção por COVID-19					
		Autorizações prévias pagas (n)	Gasto (BRL)	n	Autorizações prévias pagas (n)	%	Gasto com autorizações prévias (BRL)	%	Custo médio por autorizações prévias pagas (BRL)
Sem tratamento das palavras									
BERTopic +500	55	50.716	155.110.004	3	2.443	2,3	24.768.350	6,6	10.138,50
BERTopic +1.000	23	61.723	166.228.195	2	1.987	1,9	20.277.859	5,4	10.205,26
Word2Vec	20	0	0	1	1.005	0,5	4.909.189	1,3	4.884,77
Com tratamento das palavras									
BERTopic +500	51	47.470	189.461.609	4	3.425	3,3	14.019.644	3,7	4.093,33
BERTopic +1.000	13	38.066	82.426.853	2	1.734	1,7	5.241.321	1,4	3.022,68
Word2Vec	20	0	0	3	5.989	5,7	30.072.836	8,0	5.021,35

BERTopic +500 = mínimo de 500 autorizações por tópico; BERTopic +1.000 = mínimo de 1.000 autorizações por tópico.

Nota: o modelo Word2Vec classifica todas as autorizações.

* Valores discrepantes correspondem a autorizações não classificadas em tópicos pelo modelo.

O tratamento da variável “indicacaoclinica” modificou substancialmente os indicadores do modelo Word2Vec. Para os casos classificados como COVID-19, sem tratamento, esse modelo apresentou menor número de autorizações pagas (n = 1.005, 0,5%), de gasto total (BRL 4.909.189, 1,3%) e de custo médio por autorização (BRL 4.885) que aqueles do modelo com tratamento de palavras: 5.989 – 5,7%, BRL 30,1 milhões – 8% e BRL 5.021 de custo médio, respectivamente (Tabela 3).

A comparação entre os seis modelos mostrou que o modelo BERTopic +1.000 sem tratamento tem número menor de autorizações classificadas como suspeitas de COVID-19 com gasto total elevado e o modelo Word2Vec com tratamento tem um número maior de autorizações classificadas como suspeitas de COVID-19 com gasto total maior (BRL 30 milhões) mas resultando em custo médio menor (Tabela 3).

A avaliação da qualidade de classificação do modelo BERTopic +1.000, mostra que, dos 100 primeiros casos analisados manualmente, 70 estão relacionados à suspeita ou infecção por COVID, claramente indicados no texto da variável “indicacaoclinica”. Esses pacientes tiveram gasto de BRL 11,5 milhões – 56,5% do gasto total identificado nesse modelo (Quadro 2).

Outros 20 pacientes apresentam sinais, sintomas ou doenças respiratórias que podem ou não estar relacionadas à COVID-19. O gasto, nesse grupo, foi de BRL 2,5 milhões. Outros oito casos são de recém-nascidos com desconforto respiratório, todos sem conexão com a doença, exceto um recém-nascido extremo filho de mãe com COVID-19. Os outros dois casos apresentam sinais e sintomas respiratórios sem relação com a doença (Quadro 2). O Quadro 3 mostra as 15 primeiras autorizações dessa avaliação de qualidade com a descrição original da autorizações prévias, a respectiva classificação manual e gasto por autorização. A análise dos 100 primeiros casos pode ser encontrada no Quadro 2.

O método tradicional, usando SQL e seleção de autorizações prévias contendo as palavras *covid*, *sars*, *coronavirus* e *coronavirus*, resultou em 3.703 autorizações pagas com gasto total de BRL 23.611.018 – custo médio de BRL 6.376.

Confrontando-se o método tradicional com os modelos NLP gerados, têm-se autorizações prévias selecionadas não classificadas pelos modelos, casos de interesse que foram perdidos. Essas autorizações se espalharam nos diferentes tópicos dos modelos, mas concentraram-se no tópico com valores discrepantes, onde não é possível fazer a classificação.

Nos modelos BERTopic, a maior perda de casos ocorreu no modelo não tratado com mais de 1.000 autorizações – 2.377 (64,2%) autorizações não foram classificadas pelo modelo, tiveram gasto total de BRL 8,7 milhões e custo médio de BRL 3.673. O modelo BERTopic com mais de 500 autorizações não tratado foi pouco melhor – 1.622 (43,8%) autorizações não classificadas, gasto de BRL 5,1 milhões e custo médio por autorização de BRL 3.214. Esses casos perdidos têm custo médio por autorização quase 3 vezes menor que aqueles classificados pelos modelos. O tratamento das palavras fez com que esses modelos deixassem de classificar os casos menos graves, os custos médios por autorização dos casos perdidos foram BRL 9.323 e BRL 7.217 nos modelos BERTopic +1.000 e +500, respectivamente.

Por outro lado, os modelos classificaram autorizações não selecionadas no método tradicional. As 362 autorizações excedentes no modelo BERTopic +500 sem tratamento que não contêm as palavras da pesquisa tradicional têm custo médio de BRL 17.196 – gasto de BRL 6,2 milhões. No modelo BERTopic +1.000 sem tratamento as autorizações prévias com a mesma característica (661 autorizações) têm custo médio de BRL 8.165 e gasto total de BRL 5,4 milhões. O modelo Word2Vec com melhor desempenho, nesse sentido – 2.703 autorizações, gasto de BRL 11.369.283 e custo médio por autorização de BRL 4.206 – é o tratado (Tabela 4).

Os modelos BERTopic geraram outros tópicos de interesse – relacionados a câncer (1.500 autorizações prévias e gasto de BRL 6.662.411), doenças ortopédicas (4.531 autorizações prévias e gasto de BRL 13.675.723) e doenças mentais (3.603 autorizações prévias e gasto de BRL 818.893). Esses tópicos variam em função do método empregado – os modelos BERTopic +1.000, tratados ou não, foram piores gerando poucos tópicos adicionais. Os tópicos formados por cada modelo encontram-se nos Quadros 4, 5, 6 e 7.

Quadro 2

Avaliação do modelo BERTopic +1.000 sem tratamento por classificação manual das 100 autorizações, ordenadas por custo de casos suspeitos de infecção por COVID-19 em operadora da saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2022.

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
1	Desconforto Respiratório. SOLICITO A REALIZAÇÃO DA INTERNAÇÃO DO RN EM SUA PRÓPRIA CARTEIRA POIS PERÍODO DE INTERNAÇÃO ULTRAPASSOU OS 30 DIAS NA CARTEIRA DA MÃE A PARTIR DO DIA 19/12/2021	709.892	Doença respiratória em recém-nascido/Sintomas
2	sd respiratoria aguda grave , covid pcr 07 dias , diabetes	676.338	COVID-19
3	covid confirmado evoluiu com hiperemia	650.212	COVID-19
4	INFECÇÃO POR COVID	515.771	COVID-19
5	sintomas gripais há 10 dias. desconforto respiratório. Com Taquidispneia	428.903	Doença respiratória/Sintomas
6	refere covid+ vem para avaliação. refere piora da dispneia e sO2 87 em casa	423.112	COVID-19
7	apresentando dispneia desconforto respiratório febre 38 e queda da saturação	415.292	Doença respiratória/Sintomas
8	COVID HA + DE 7 DIAS, TC MOSTRA ACOMETIMENTO ENTRE 15 E 50% DA ÁREA PULMONAR.	402.072	COVID-19
9	paciente com covid positivo apresentando dispneia a meios esforços e saturação de 88% em ar ambiente	390.321	COVID-19
10	Pneumonia por microorganismo	387.281	Doença respiratória/Sintomas
11	Paciente com COVID + com piora dos sintomas respiratórios nas últimas 24 horas	382.854	COVID-19
12	PNM VIRAL POR COVID-19?	378.524	COVID-19
13	BCP, COVID	352.845	COVID-19
14	RECEM NASCIDO PREMATURO EXTREMO, FILHO DE MÃE COVID POSITIVO, INTERNADO EM UTI NEONATAL NECESSITANDO DE SUPORTE VENTILATÓRIO, CLÍNICO E HEMODINÂMICO. SOLICITO INTERNAÇÃO DO MESMO POIS MÃE DE ALTA HOSPITALAR E RN NECESSITA PERMANECER INTERNADO PARA SUPORTE E TRATAMENTO.	337.206	Doença respiratória em recém-nascido/Sintomas
15	PACIENTE COVID 19 EVOLUINDO COM QUEDA DA SATURAÇÃO E DISPNEIA COM NECESSIDADE DE O2	331.786	COVID-19
16	Covid ? Falta de ar, tosse seca há dez dias com piora há dois dias TC com 25-50 de acometimento	309.762	COVID-19
17	Paciente com quadro de mialgia, tosse seca e febre há 07 dias. Há 05 dias realizou teste para covid-19 com resultado positivo. Febre desde então. Já em uso de azitromicina	302.743	COVID-19
18	Paciente em D14 de sintomas covid-19 positivo, da entrada em PS dessaturando. Foi colocado em sala de emergência devido a quadro clínico, com monitoração completa e cateter de O2 5l/min. Evoluindo com melhora da saturação. Tc torax: áreas de opacidades em vidro fosco e consolidação dispersas em ambos pulmões, de distribuição periférica.	281.346	COVID-19
19	Paciente hipertenso e hepatopata com infecção por covid 19, evoluiu com piora clínica da tosse, dispnéia e dessaturação.	276.431	COVID-19
20	COVID + HA 07 DIAS, DISPNEIA E PIORA SATURAÇÃO	270.206	COVID-19

(continua)

Quadro 2 (continuação)

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
21	B972-CORONAVIRUS, COMO CAUSA DE DOENÇAS CLASSIFICADAS EM OUTROS CAPÍTULOS J180-BRONCOPNEUMONIA NÃO ESPECIFICADA	242.635	COVID-19
22	COVID POSITIVO	229.489	COVID-19
23	Paciente covid positivo, com início dos sintomas há 10 dias refere dispnéia progressiva e tosse há 5 dias	226.263	COVID-19
24	Paciente broncodisplásica + portadora de AME 1, portadora de traqueostomia e gastrostomia, em quadro de broncoespasmo associado a hipoxemia. Internação em UTI devido necessidade de ventilação mecânica.	224.587	Doença respiratória/ Sintomas
25	INFECÇÃO POR COVID 19	215.832	COVID-19
26	COVID 19 POSITIVO PACIENTE EVOLUI COM PIORA TRANSFERIDO PARA UTI	203.934	COVID-19
27	COVID POSITIVO	198.838	COVID-19
28	Transferência da [omitido] com vaga cedida para internação 8º dia de sintomas: tosse, cefaléia, mialgia, dispnéia.... ontem relata ter tido o pior dia desde o início dos sintomas com dispnéia mais acentuada, motivo pelo qual voltou a procurar atendimento médico Na terça-feira dia 08/12 já havia comparecido à esta UPA e realizou o PCR para COVID (ainda sem resultado) e fez TC tórax (relata que exame foi normal e por isso foi liberado para tratamento em casa)	193.848	COVID-19
29	rn com desconforto respiratório	181.064	Doença respiratória em recém-nascido/Sintomas
30	covid positivo	180.297	COVID-19
31	pneumonia por covid com infecção secundária, febre e expectoração amarelada. desaturação. tc com imagem extensa leve	177.360	COVID-19
32	PNEUMONIA + COVID ?	167.661	COVID-19
33	Suspeita de COVID19	162.645	COVID-19
34	DISPNEIA + TOSSE + FEBRE + LINFOPENIA - COM SUSPEITA DE COVID-19	156.417	COVID-19
35	BCP, COVID-19 DISPNEIA, REBAIXAMENTO DE SAT O2, AR AMB - 91-92% AR AMB. TC CONSTA DERRAME IMPORTANTE MAIS LADO DIR, COM INFLT DIFUSA SEIO CARD, PERIF	143.477	COVID-19
36	Contato com COVID19 confirmado. Evolui com tosse, dispnéia e febre	142.496	COVID-19
37	covid + 09/03 apresentar febre, dispnéia, cefaleia	141.157	COVID-19
38	covid 19 positivo, paciente com queda do estado geral	139.158	COVID-19
39	Paciente com Covid-19 confirmado em D7 de sintomas evolui com piora clínica, dispnéia, queda de saturação e febre persistente	134.264	COVID-19
40	PACIENTE REFERE FALTA DE AR E FEBRE HA 11 DIAS (DIA 20/04 TEVE DIAGNOSTICO POSITIVO PARA COVID) HEMOPTISE HA 5 DIAS, PIORA DA DISPNEIA HOJE. SAT 88%. FEZ USO DE AZITROMICINA SEM MELHORA. NEGA ALERGIAS.	128.084	COVID-19
41	covid 19 + com saturação 92% em AA	122.251	COVID-19
42	COVID+ PCR de 27/09 início dos sintomas em 25/10 - D6 de sintomas retorna por tosse, fraqueza, inapetência, diarreia e vômito hipertenso, diabético, sobrepeso nega doença arterial coronariana	107.067	COVID-19

(continua)

Quadro 2 (continuação)

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
43	PACIENTE VEM PARA A CONSULTA COM QUADRO DE DISPNEIA E DESSATURAÇÃO. NÃO É VACINADA PARA COVID (NÃO QUIS TOMAR A VACINA POIS TEM MEDO). FEZ PCR HÁ 4 DIAS PARA COVID (POSITIVO). HOJE INICIOU INTENSA DISPNEIA E DESSATURAÇÃO. AO EXAME FÍSICO APRESENTA CREPITAÇÕES GROSSEIRAS EM HTD E SIBILOS DIFUSOS. APRESENTA DEFICIT MOTOR SECUNDÁRIO A AVC ISQUÊMICO PREGRESSO.	105.904	COVID-19
44	PACIENTE COM TOSSE SECA SECA E MAL ESTAR HÁ 1 SEMANA, SENDO DIAGNOTICADA COM COVID HÁ 4 DIAS VIA PCR. HOJE, QUEIXANDO-SE DE DISPNEIA, DOR TORACICA VENTILATÓRIO DEPENDENTE COM PIORA Á TOSSE, ALÉM DE DIARREIA.	102.470	COVID-19
45	Paciente com dispneia, queda de saturação de O2 para 92-94%, com esposa diagnosticada com Covid-19 e internada. TC torax: acometimento de 10% de pulmão em vidro fosco	99.690	COVID-19
46	DESSATURAÇÃO + DISPNEIA + COVID POSITIVO	95.268	COVID-19
47	pneumonia bacteriana	95.260	Doença respiratória/ Sintomas
48	PACIENTE COM TOSSE SECA, DISPNEIA E MIALGIA DIFUSA HÁ 5 DIAS, COM PIORA NAS ÚLTIMAS 24H. TC TÓRAX COM ACOMETIMENTO EM VIDRO FOSCO DE 25-50%, BILATERAL, ASSOCIADO A ELEVAÇÃO DE PCR(16,9).	88.353	Doença respiratória/ Sintomas
49	insuficiencia respiratoria	85.981	Doença respiratória/ Sintomas
50	PACIENTE DIAGNOSTICADO COM COVID 19 EM D5 DE SINTOMAS, RETORNA COM PIORA DE DISPNEIA, ESTADO SUBFEBRIL E DESATURAÇÃO	82.720	COVID-19
51	Paciente com dificuldade respiratória, cansaço aos minimos esforços, edema há um dia e piora progressiva	82.356	Doença respiratória/ Sintomas
52	SD RESPIRATÓRIA A/E COVID ?	82.310	COVID-19
53	Paciente acamada com sequela de AVC e demência em tratamento em domicilio de broncopneumonia apresentou piora clinica dessaturação com necessidade de mascara não reinalante. Internação em UTI.	80.711	Doença respiratória/ Sintomas
54	rn baixo peso com desconforto respiratorio	80.034	Doença respiratória em recém-nascido/Sintomas
55	PACIENTE COM QUEIXA DE INAPETENCIA FADIGA , FEBRE, EXAMES SUGESTIVOS DE COVID 19	78.061	COVID-19
56	Inf. de cateter?COVID?	76.414	COVID-19
57	PNEUMONIA COM SUSPEITA DE COVID-19	74.352	COVID-19
58	Insuficiência Respiratória. Covid	73.550	COVID-19
59	Paciente com dispneia, queda da saturação de o2 para 92-94% com esposa diagnosticada com covid 19 e internada. TC torax: acometimento de 10% de pulmão em vidro fosco	72.050	COVID-19

(continua)

Quadro 2 (continuação)

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
60	Paciente acamada, totalmente dependente para ABVD (home care), AVEH há um 3 anos, DM, CA de próstata há 8 anos. Vem trazido por remoção de casa com história de queda da PA arterial, febre de início hoje, tremores, oligúria e colúria. Logo após apresentar desconforto respiratório com queda de saturação. PA 70/40.	71.589	Doença respiratória/ Sintomas
61	PACIENTE INTERNA COM COVID 19 POSITIVO + DISPNEIA + DEPENDENTE DE O2 + DESCONFORTO RESPIRATÓRIO. EM ANEXO A DOCUMENTAÇÃO.	70.420	COVID-19
62	pneumonia bacteriana em paciente idoso	70.373	Doença respiratória/ Sintomas
63	52a, refere diagnostico de COVID + ha 11 dias (resultado de farmacia sic) nao traz o resultado. hoje, refere tosse seca frequente desde cedo. Nega febre, diarreia e falta de ar. Na entrada, sat 95% AA com FR 30 ipm e FC 104 bpm PA 160/100 AP HAS nega DM asma TBG (ex TBG ha 35a)	69.982	COVID-19
64	RECEM NASCIDO PREMATURO , PRIMEIRO GEMELAR , APRESENTOU SINDROME DO DESCONFORTO RESPIRATORIO. PERFURACAO GASTRICA , SUBMETIDO A LAPAROTOMIA EXPLORADORA. ENCONTRA-SE EM VENTILACAO MECANICA , INDICADO MANTER INTERNACAO EM UTI NEONATAL PARA CUIDADOS ITENSIVOS	69.078	Doença respiratória em recém-nascido/Sintomas
65	COVID19 HAS / DM // TOSSE E PERDA DE APETITE HA 4 DIAS , FEBRE NÃO AFERIDA // APRESENTOU NAUSEAS E HIPOTENSAO (100 X 60) COM SUDORESE DURANTE O EXAME	68.200	COVID-19
66	paciente portadora de laringotraqueomalacia e aneurisma de galeno apresentando piora do estridor, e taquidispneia. Ao exame fisico FR 79, FC 143, fiagdo rebaixado e pele moqueada. Nega febre ou outros sintomas. Devido o quadro de possivel insuficiencia cardiaca e respiratoria opto por internação em UTI para melhor estabilidade clinica	63.717	Outras doenças com sintomas respiratórios
67	DESCOFORTO RESPIRATORIO	62.369	Doença respiratória/ Sintomas
68	choque anafilático(reação á administração de hidrocortizida), dessaturação, insuficiência respiratória	61.023	Outras doenças com sintomas respiratórios
69	PACIENTE COM COVID-19 E ALTERAÇÕES NA TOMOGRAFIA, INTERNOU EM 21/08	60.471	COVID-19
70	COVID-19 HÁ 5 DIAS, EVOLUI COM DISPNEIA PROGRESSIVA E MIALGIA INTENSA, ANTECEDENTE DE LUPUS E TVP, UTILIZA MAREVAN, TC MOSTRA ACOMETIMENTO DE 22% DE PARENQUIMA PULMONAR, GASOMETRIA ARTERIAL COM S O2=77,6, PCO2=47 MMHG, PO2=42 MMHG, SOLICITO INTERNAÇÃO EM ENFERMARIA	59.406	COVID-19
71	INTERNAR APÓS 30 DIAS NO CARTÃO DO RN PRE TERMO MUITO BAIXO PESO SINDROME DO DESCONFORTO RESPIRATÓRIO INTERNAÇÃO REFERENTE A DATA 24/09/2021	59.192	Doença respiratória em recém-nascido/Sintomas
72	RN PARTO CESÁREA, IG 39+6, SOROLOGIAS NEGATIVAS, GBS NEG, BI, APGAR 6/8, NASCEU COM CHORO FRACO, CIANÓTICO, SAT 71-75% EM O2, FEITO O2 INALATÓRIO, EVOLUI COM DESCONFORTO RESPIRATÓRIO BSA 5. NECESSITA INTERNAÇÃO EM UTI NEO	59.002	Doença respiratória em recém-nascido/Sintomas

(continua)

Quadro 2 (continuação)

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
73	PACIENTE INTERNA POR COVID 19	57.447	COVID-19
74	paciente com covid 19+	56.930	COVID-19
75	Vaga cedida, pneumonia paciente veio do [omitido] por suspeita de pneumonia bacteriana	55.078	Doença respiratória/ Sintomas
76	Paicnete com esposo com covid desde dia 04/12/ começou com falta de ar, tosse com secreção, coriza e mialgia há 3 dias, nega febre. Estava saturando em casa entre 85-90%.	53.949	COVID-19
77	paciente com pneumonia por covid com piora dos sintomas	52.159	COVID-19
78	covid + (d9 sintomas) com piora da febre e da dispneia ha 3 dias sat o2 aa 85% fr 22 sat c/ O2 91% mv + ec difusos. internação em UTI	51.201	COVID-19
79	Paciente proveniente do [omitido] onde estava em reabilitação de sequela de AVC, deu entrada no [omitido] em 02/12, com queixa de desconforto respiratório, febre, RNC. Após ser avaliado pela equipe medica, realizado exames de imagem e laboratoriais, sendo evidenciado Sepse de Foco Pulmonar, sendo optado por internação para tratamento com antimicrobianos. Previsto aplicação de 4 frascos de 100UI de Botox.	48.812	Doença respiratória/ Sintomas
80	covid em paciente no grupo de risco	48.721	COVID-19
81	INSUFICIENCIA RESPIRATORIA PCR COVID + AP: DEPRESSÃO	48.307	COVID-19
82	PACIENTE COM QUADRO DE TOSSE, ADINAMIA, INAPETENCIA E DISPNEIA HÁ 1 SEMANA COM PIORA DESDE ONTEM NA ADMISSÃO SATO2 92%. LABORATORIO COM PCR ELEVADO, LEUCOCITOSE COM DESVIO 9% BASTÕES, TC DE TORAX CONSOLIDAÇÃO PULMONAR OCUPANDO QUASE A TOTALIDADE DO LOBO SUPERIOR ESQUERDO (PNEUMONIA LOBAR). DESCARTADO COVID-19. SOLICITO INTERNAÇÃO EM ENFERMARIA DE CLINICA MEDICA - PNM LOBAR	47.900	COVID-19
83	***** SOLICITO INTERNAÇÃO RETROATIVA DO DIA 31/12/2021 ***** PACIENTE ACAMADO , GTT - TQT HA 1 DIA COM QUEDA DO ESTADO GERAL, FEBRE, ALTERAÇÃO ASPECTO URINARIO, DESATURAÇÃO, BRADICARDIA SIC CONTACTANTE COVID	47.479	COVID-19
84	PACIENTE COVID 19 CONFIRMADO, APRESENTANDO TC TORAX COM ACOMETIMENTO > 50% SAT 02 90% AA E FR 24.	46.194	COVID-19
85	PACIENTE COVID POSITIVO, COM QUADRO DE DISPNEIA COM PIORA PROGRESSIVA, COM BAIXA SATURAÇÃO DE O2. DEU ENTRADA COM HIPOTENSÃO, COM MELHORA DA PA POS EXPANSÃO VOLEMICA	45.256	COVID-19
86	Paciente ha 05 dias apresenta dificuldade para se alimentar tosse e diarreia .refere desconforto respiratorio associado ao quadro .	45.217	Doença respiratória/ Sintomas
87	paciente previamente hígido sem antecedentes mórbidos, evoluindo com quadro respiratórios ha 8 dias	45.137	Doença respiratória/ Sintomas
88	PACIENTE COM FALTA DE AR, CANSAÇO, HIPOREXIA, MIALGIA E PROSTRACÃO INTENSA, TESTE DE COVID + TC TORAX APRESENTA ACOMETIDO PULMONAR 25 % SAT 95%	43.496	COVID-19
89	covid confirmado / protocolo de sepsse	43.196	COVID-19
90	Paciente em 11º dia de sintomas, com PCR COVID positivo, apresentando piora gradual com fadiga intensa, dispneia, tosse e SAT O2: 68% em AA	42.553	COVID-19

(continua)

Quadro 2 (continuação)

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
91	paciente no 11 dia d covid , evoluindo com mal estar, fadiga , dispenai e dessaturação tem tc de torax com 50% acometimnto	42.353	COVID-19
92	Paciente de 78 anos vem ao PS com quadro de Pneumonia + Delirium, relata que ontem estava bem, porem, durante a noite teve mta tosse, não dormiu direito, esta mais confuso e apatico, relata sensação de peso na cabeça.	41.929	Doença respiratória/ Sintomas
93	PNM VIRAL / COVID 19 ? SOLICITO AUTORIZAÇÃO PARA INTERNAÇÃO RETROATIVA - GUIA: 2022909741 COM DIÁRIA ZERADO (PACIENTE INTERNOU EM 24/12/2020)	41.357	COVID-19
94	INTERNAÇÃO RETROATIVA (13/04) - PACIENTE COM COVID POSITIVO HÁ 14 IDAS COM PROSTRAÇÃO. *** SOLICITADO NOVAMENTE POIS DIARIA LIBERADA NÃO TEMOS CONTRATO**	41.061	COVID-19
95	Paciente proveniente do [omitido] onde estava em reabilitação de sequela de AVC, deu entrada no [omitido] em 02/12, com queixa de desconforto respiratório, febre, RNC. Após ser avaliado pela equipe medica, realizado exames de imagem e laboratoriais, sendo evidenciado Sepse de Foco Pulmonar, sendo optado por internação para tratamento com antimicrobianos. Evolução/ intercorrências: Clínica e hemodinamicamente estável, menos secreto.	40.420	Doença respiratória/ Sintomas
96	DESCONFORTO RESPIRATÓRIO	40.263	Doença respiratória/ Sintomas
97	DOR E EMPASTAMENTO DE PANTURRILHA ESQUERDA APOS COVID	39.357	COVID-19
98	RNPT, AIG, DESCONFORTO RESPIRATORIO COM NECESSIDADE DE INTERNAÇÃO PARA SUPORTE VENTILATORIO E TRANSIÇÃO PARA DIETA ENTERAL	38.498	Doença respiratória em recém-nascido/Sintomas
99	PACIENTE INTERNADA EM 08/04 ALTA 19/04 EM ISOLAMENTO DEVIDO COVID 19 - SOLICITO INTERNAÇÃO COM DATA RETROATIVA DEVIDO ALTERAÇÃO DE CODIGO LIBERADO	38.284	COVID-19
100	PAciente HAS + DM + DLP + COVID confirmado em 19/08/20, inicio dos sintomas em 16/08/20. Procura PS com queixa de falta de ar + diarreia + mialgia e febre com inicio hoje. Na triagem SatO ² :95%.	38.254	Doença respiratória/ Sintomas

* A descrição do pedido de autorização prévia não foi alterada e está indicada como se encontra no pedido;

** Classificação do autor baseado na análise do campo "indicacaoclinica". A classificação foi independente da classificação gerada pelo modelo. Os casos classificados como COVID-19 indicam suspeita da infecção de paciente, cuja autorização prévia foi emitida nos termos contidos na "indicacaoclinica".

Quadro 3

Avaliação do modelo BERTopic +1.000 sem tratamento por classificação manual das 15 autorizações, ordenadas por custo de casos suspeitos de infecção por COVID-19 em operadora da saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2022.

AUTORIZAÇÕES PRÉVIAS	DESCRIÇÃO DO PEDIDO DE AUTORIZAÇÃO PRÉVIA *	GASTO POR AUTORIZAÇÃO (BRL)	CLASSE **
1	Desconforto Respiratório. SOLICITO A REALIZAÇÃO DA INTERNAÇÃO DO RN EM SUA PROPRIA CARTEIRA POIS PERIODO DE INTERNAÇÃO ULTRAPASSOU OS 30 DIAS NA CARTEIRA DA MÃE A PARTIR DO DIA 19/12/2021	709.892	Doença respiratória em recém-nascido/Sintomas
2	sd respiratoria aguda grave , covid pcr 07 dias, diabetes	676.338	COVID-19
3	covid confirmado evoluuino com hiperemia	650.212	COVID-19
4	INFECÇÃO POR COVID	515.771	COVID-19
5	sintomas gripais há 10 dias. desconforto respiratório. Com Taquidispneia	428.903	Doença respiratória/Sintomas
6	refere covid+ vem para avaliação. refere piora da dispneia e sO2 87 em casa	423.112	COVID-19
7	apresentando dispneia desconforto respiratorio febre 38 e queda da saturação	415.292	Doença respiratória/Sintomas
8	COVID HA + DE 7 DIAS, TC MOSTRA ACOMETIMENTO ENTRE 15 E 50% DA AREA PULMONAR.	402.072	COVID-19
9	paciente com covid positivo apresentando dispneia a medios esforços e saturação de 88% em ar ambiente	390.321	COVID-19
10	Pneumonia por microorganismo	387.281	Doença respiratória/Sintomas
11	Paciente com COVID + com piora dos sintomas respiratórios nas ultimas 24 horas	382.854	COVID-19
12	PNM VIRAL POR COVID-19 ?	378.524	COVID-19
13	BCP, COVID	352.845	COVID-19
14	RECEM NASCIDO PREMATURO EXTREMO, FILHO DE MÃE COVID POSITIVO, INTERNADO EM UTI NEONATAL NECESSITANDO DE SUPORTE VENTILATÓRIO, CLINICO E HEMODINAMICO. SOLICITO INTERNACAO DO MESMO POIS MAE DE ALTA HOSPITALAR E RN NECESSITA PERMANECER INTERNADO PARA SUPORTE E TRATAMENTO.	337.206	Doença respiratória em recém-nascido/Sintomas
15	PACIENTE COVID 19 EVOLUINDO COM QUEDA DA SATURAÇÃO E DISPNEIA COM NECESSIDADE DE O2	331.786	COVID-19

Nota: o quadro completo contendo as 100 primeiras autorizações prévias analisadas está no Quadro 2.

* A descrição do pedido de autorização prévia não foi alterada e está indicada como se encontra no pedido;

** Classificação do autor baseada na análise do campo "indicacaoclinica". A classificação foi independente da classificação gerada pelo modelo.

Os casos classificados como COVID-19 indicam suspeita da infecção de paciente cuja autorização prévia foi emitida nos termos contidos na "indicacaoclinica".

Tabela 4

Compara modelos com método tradicional de seleção de palavras na classificação de autorizações emitidas por operadora de saúde suplementar. São Paulo, Brasil, setembro/2019 a junho/2020.

Modelos	Autorizações prévias classificadas no modelo		Perdas quando comparadas ao método tradicional				Modelo encontrou, mas método tradicional perdeu	
	Autorizações prévias pagas (n)	Gasto (BRL)	Autorizações pagas (n)	%	Gasto (BRL)	%	Autorizações prévias pagas (n)	Gasto (BRL)
Sem tratamento das palavras								
BERTopic +500	2.081	18.543.341	1.622	43,8	5.067.677	21,5	362	6.225.009
BERTopic +1.000	1.326	14.880.970	2.377	64,2	8.730.048	37,0	661	5.396.889
Word2Vec	985	4.842.603	2.718	73,4	18.768.415	79,5	20	66.586
Com tratamento das palavras								
BERTopic +500	2.249	13.117.398	1.454	39,3	10.493.620	44,4	1.176	902.247
BERTopic +1.000	1.730	5.217.648	1.973	53,3	18.393.370	77,9	4	23.674
Word2Vec	3.286	18.703.552	417	11,3	4.907.466	20,8	2.703	11.369.284

BERTopic +500 = mínimo de 500 autorizações por tópicos; BERTopic +1.000 = mínimo de 1.000 autorizações por tópicos.

Método tradicional: pesquisa de linguagem estruturada (SQL – *structured query language*) considerando a presença de palavras *covid*, *sars*, *coronavírus* e *coronavírus*, maiúsculas e minúsculas.

Nota: o método tradicional encontrou 3.703 autorizações com gasto total de BRL 23.611.018.

Quadro 4

Número de autorizações prévias por tópicos gerados pelo modelo BERTopic +500 sem tratamento de palavras e respectiva descrição dos autores.

TÓPICOS MODELO BERTopic +500 SEM TRATAMENTO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
-1	85.505	-1_cid_de_em_tratamento	
0	13.551	0_joelho_direito_esquerdo_lesão	Lesão de joelhos direito ou esquerdo
1	8.848	1_rotina_check_up_azia	Exames de rotina/ <i>check-up</i>
2	8.567	2_prorrogação_05_01_03	Rotina administrativa
3	7.783	3_exame_srotina_sorologico_eroina	Exames de rotina/ <i>check-up</i>
4	4.663	4_exame_oct_raiox_bhcg	
5	4.032	5_paciente_com_quadro_tosse	Paciente com tosse
6	3.668	6_covid_19_com_sintomas	Com sintomas de COVID-19
7	2.883	7_mg_ev_continuidade_ciclo	
8	2.713	8_cid_10_e66_m54	
9	2.324	9_f41_f410_f415_	Grupo CID-10 (transtornos mentais e comportamentais)
10	2.213	10_esclarecer_esofagite_misto_esclarcer	
11	1.779	11_anexo_em_ema_pm	

(continua)

Quadro 4 (continuação)

TÓPICOS MODELO BERTopic +500 SEM TRATAMENTO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
12	1.479	12_f33_f34_fr33_deprssão	Grupo CID-10 (transtornos mentais e comportamentais)
13	1.290	13_abdominal_vomitos_dor_ nauseas	Dor abdominal com náuseas e vômitos
14	1.265	14_socorro_pronto_adulto_socoroo	
15	1.253	15_arterial_vascular_cerebral_ taquicardia	
16	1.240	16_m54_m51_m65_m75	Grupo CID-10 doenças osteomusculares
17	1.221	17_terapia_terpia_teraioia_teratoma	
18	1.081	18_exames_rotina_de_intratensão	
19	1.073	19_f32_f53_f328_f323	Grupo CID-10 (transtornos mentais e comportamentais)
20	1.048	20_local_no_paciente_md224	
21	1.022	21_tosse_febre_mialgia_coriza	Febre com coriza, tosse e mialgia
22	1.012	22_endoscopia_colonosopia_ polipectomia_colon	Endoscopia/Colonoscopia com polipectomia
23	1.012	23_dores_dore_arritimia_pateolo	
24	997	24_dor_ae_ver_amdominal	Dor abdominal a esclarecer
25	963	25_individual_psicoterapia_semana_2x	Psicoterapia individual
26	961	26_neoplasia_maligna_mama_tumor	Neoplasia maligna de mama
27	960	27_anexo_pedido_conforme_médico	
28	891	28_ico_atc2014_uniarterial_atc	
29	876	29_rotina_exame_de_reastreio	Exames de rotina/ <i>check-up</i>
30	871	30_avaliação_equoterapia_medico_ clinico	
31	838	31_i10_oncologico_oncológico_paciente	Paciente oncológico (CID-10 hipertensão)
32	781	32_cervical_cervicalgia_coluna_lombar	Ortopedia coluna
33	750	33_vacina_anual_vacinaçã_quadrupla	Vacinação anual
34	748	34_anexos_vida_metastase_segmento	
35	732	35_hematuria_duplo_ureterolitiase_ ureter	Hematúria com ureterolitiase
36	732	36_auditor_validação_conforme_médico	Rotina administrativa
37	693	37_gripais_sintomas_gripal_dias	Sintomas de gripe
38	690	38_psicoterapia_psicologia_individual_ ibndividual	Psicoterapia individual
39	689	39_covid_estima_auto_baixa	COVID-19
40	668	40_lombalgia_lombragia_discogenica_ lombaciatalgia	Ortopedia coluna
41	662	41_consulta_socorro_pronto_em	Consulta em pronto socorro
42	661	42_has_h40_psicoterapia_catarata	
43	659	43_ps_paciente_em_psa	
44	656	44_anexo_f10_f19_f103	

(continua)

Quadro 4 (continuação)

TÓPICOS MODELO BERTopic +500 SEM TRATAMENTO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
45	635	45_hernia_disco_hérnia_discal	Ortopedia coluna
46	567	46_ nao_ha_não_há	
47	560	47_individual_psicoterapia_sessao_sessão	Psicoterapia individual
48	544	48_maligna_neoplasia_rastreamento_de	Rastreamento para neoplasia maligna
49	518	49_segue_anexo_anexol_40313498	
50	517	50_suspeita_covid_de_h1n1	Suspeita de COVID-19 ou H1N1
51	512	51_trauma_traumatismo_queda_face	Trauma de face por queda
52	509	52_cid10_f41_ansiosos_outros	
53	506	53_sindrome_manguito_rotador_impacto	Síndrome do manguito rotador
54	500	54_pertinencia_tecnica_monitor_d22	Rotina administrativa

CID-10: Classificação Internacional de Doenças, 10ª revisão.

Nota: o tópico -1 é considerado *outlier* segundo o modelo.

* Inclui todas as autorizações inclusive valores zerados;

** Tópicos gerados automaticamente pelo modelo;

*** Análise qualitativa do nome gerado pelo tópico pelos autores.

Quadro 5

Número de autorizações prévias por tópicos gerados pelo modelo BERTopic +500 com tratamento de palavras e respectiva descrição dos autores.

TÓPICOS MODELO BERTOPIC +500 TRATADO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
-1	84.133	-1_transtorno_gripe_transtornos_direito	
0	38.255	0_nan_dd_pd_snc	
1	5.705	1_diárias_autorização_diária_uti	Rotina administrativa
2	4.957	2 Joelho_direito_esquerdo_fratura	Fratura de joelho esquerdo ou direito
3	3.208	3_f41_f52_f42_f10	Grupo CID-10 (transtornos mentais e comportamentais)
4	2.692	4_just_vie_ee_ah	
5	2.323	5_cide039_cidm224_cidi25_cidi839	
6	1.990	6_covid_teste_covide_sepse	Teste para COVID-19 e sepse
7	1.971	7_socorro_pronto_adulto_prontos	Consulta em pronto socorro
8	1.746	8_f33_f34_fr33_fffff33	
9	1.738	9_arteial_angina_vascular_sincope	
10	1.585	10_cervical_coluna_lombar_cervicalgia	Ortopedia coluna

(continua)

Quadro 5 (continuação)

TÓPICOS MODELO BERTOPIC +500 TRATADO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
11	1.438	11_abdominal_nauseas_vomitos_nega	
12	1.409	12_neoplasia_tumor_maligna_mama	Neoplasia maligna de mama
13	1.399	13_f32_corona_virus_cidb342	COVID-19
14	1.294	14_dores_intensas_olhos_dore	Dor no olho
15	1.288	15_renal_hematuria_calculo_duplo	Hematúria por cálculo renal
16	1.286	16_suspeita_covid_supeita_susp	Suspeita de COVID-19
17	1.258	17_terapia_reabilitação_membros_sup	
18	1.163	18_endoscopia_colonosopia_polipectomia_colon	Endoscopia/colonoscopia com polipectomia
19	1.160	19_covid_positivo_contato_tosse	Teste positivo para COVID-19, paciente com tosse
20	1.149	20_mg_ciclo_ansetron_sc	
21	1.082	21_ps_rotinas_exams_mandar	
22	1.051	22_cidf41_cidc41_tcg_cidf41p	
23	1.040	23_cidb07_cidm545_lesões_cid	
24	989	24_individual_psicoterapia_médica_clinica	
25	951	25_tosse_coriza_garganta_seca	Tosse, coriza e garganta seca
26	910	26_vacina_h1n1_vai_vacinaçã	
27	889	27_ico_atcidc2014_ic_has	
28	833	28_consulta_socorro_pronto_ortopedia	Consulta em pronto socorro ortopédico
29	802	29_avaliação_cidj111_avaliação_fono	
30	795	30_sindrome_transtorno_manguito_rotador	Síndrome do manguito rotador
31	773	31_psicoterapia_f81_f80_individual	Psicoterapia individual
32	698	32_i10_so10_ms10_	
33	685	33_cid_cidzoo_cidtontura_cidprovavel	
34	684	34_lombalgia_canelite_oa_aguda	
35	659	35_macular_visual_retina_acuidade	Acuidade visual, doença macular ou retiniana
36	652	36_individual_sessão_psicoterapia_sessao	Psicoterapia individual
37	606	37_hernia_disco_discal_umbilical	
38	593	38_cidf84_cidf41_	
39	592	39_cidf33_g12_diarias_	
40	583	40_urgencia_psicoterápico_urgência_pediatria	
41	582	41_trauma_cranio_traumatismo_queda	Traumatismo craniano por queda
42	563	42_alergologista_cidz10_alerologista_alergosita	Consulta com alergologista
43	556	43_maligna_neoplasia_rastreamento_rastreamentod	Rastreamento para neoplasia maligna

(continua)

Quadro 5 (continuação)

TÓPICOS MODELO BERTOPIC +500 TRATADO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
44	548	44_oncológico_oncologico_metastse_segmento	Seguimento oncológico e metástase
45	533	45_monitor__	
46	524	46_consulta_consultorio_hm_consultório	Consulta ambulatorial
47	522	47_f84_cardiologica_cardiologico_cardiológica	Consulta com cardiologista
48	514	48_abdominal_pilates_abdomen_paracentese	
49	510	49_has_desevolvimento_global_transtorno	
50	505	50_z00_z50_z0_zo	

CID-10: Classificação Internacional de Doenças, 10ª revisão.

Nota: o tópico -1 é considerado *outlier* segundo o modelo.

* Inclui todas as autorizações inclusive valores zerados;

** Tópicos gerados automaticamente pelo modelo;

*** Análise qualitativa do nome gerado pelo tópico pelos autores.

Quadro 6

Número de autorizações prévias por tópicos gerados pelo modelo BERTopic +1.000 sem tratamento de palavras e respectiva descrição dos autores.

TÓPICOS MODELO BERTopic +1.000 SEM TRATAMENTO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
-1	104.536	-1_de_cid_em_anexo	
0	11.726	0_com_paciente_dor_de	
1	9.206	1 Joelho_direito_esquerdo_ombro	Lesão de joelhos direito ou esquerdo ou ombro
2	8.858	2_rotina_rotura_check_up	Exames de rotina/ <i>check-up</i>
3	7.794	3_05_prorrogação_01_03	Rotina administrativa
4	7.781	4_exame_srotina_eroína_laringe	
5	4.669	5_exame_lab_examens_oct	
6	3.287	6_covid_19_com_paciente	Com sintomas de COVID-19
7	2.952	7_mg_ev_continuidade_para	
8	2.846	8_cid_10_i10_hd	
9	2.496	9_neoplasia_endoscopia_colonosopia_colon	Endoscopia/Colonoscopia e neoplasia de colón
10	2.326	10_f41_f410_f415_cis10	Grupo CID-10 (transtornos mentais e comportamentais)

(continua)

Quadro 6 (continuação)

TÓPICOS MODELO BERTopic +1.000 SEM TRATAMENTO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
11	2.196	11_esclarecer_esofagite_esclarcer_esclarecer	Esofagite a esclarecer
12	1.802	12_anexo_em_somatização_45	
13	1.479	13_f33_f34_fr33_deprssão	Grupo CID-10 (transtornos mentais e comportamentais)
14	1.330	14_vida_rotina_exame_anexo	
15	1.303	15_socorro_pronto_h360_h353	
16	1.224	16_terapia_terpia_teraioia_teratoma	
17	1.197	17_f32_f328_f323_ee	Grupo CID-10 (transtornos mentais e comportamentais)
18	1.187	18_m54_m51_m75_m65	Grupo CID-10 doenças osteomusculares
19	1.066	19_exames_rotina_de_mulher	Exames de rotina/ <i>check-up</i>
20	1.054	20_dor_f43_f51_ae	
21	1.046	21_local_no_paciente_md224	
22	1.010	22_dores_dore_arritimia_dorescovid	

CID-10: Classificação Internacional de Doenças, 10ª revisão.

Nota: o tópico -1 é considerado *outlier* segundo o modelo.

* Inclui todas as autorizações inclusive valores zerados;

** Tópicos gerados automaticamente pelo modelo;

*** Análise qualitativa do nome gerado pelo tópico pelos autores.

Discussão

O modelo BERTopic sem tratamento de palavras selecionou pacientes mais graves, enquanto o modelo Word2Vec com tratamento de palavras selecionou pacientes menos graves. Já em 1998, Hernández & Stolfo³⁴ discutiam a dificuldade em se trabalhar com dados de mundo real. Esse desafio é maior com uso de dados não estruturados. Os 100 casos manualmente analisados mostram diferenças quanto às formas de nomear o vírus, ampliadas pelas peculiaridades da língua portuguesa – acentos, por exemplo. Outro desafio é a amplitude das informações – grande parte das autorizações foram preenchidas com frases de até cinco palavras. Ainda assim, o modelo BERTopic conseguiu selecionar casos com a descrição “sintomas gripais há 10 dias. Desconforto respiratório com taquidispneia” como suspeitos de infecção pelo vírus. Observa-se que não há menção explícita à COVID-19 e enquanto respiratório tem acento, taquidispneia não tem, um exemplo do problema de dados não estruturados.

Essa dificuldade deve explicar por que existem poucos estudos utilizando NLP aplicados à detecção precoce da doença. Em revisão de uso de ferramentas de inteligência artificial aplicadas na resposta à pandemia, Syrowatka et al.³⁵ indicaram somente um estudo baseado em NLP para diagnóstico precoce ou triagem de pacientes. O maior volume de estudos (65 de 78) utilizaram técnicas de processamento de imagens de tórax. As autoras indicam que a maioria dos estudos analisados ainda estão em fase de pesquisa e poucos são utilizados para tomada de decisão³⁵. Uma revisão específica sobre uso de NLP na pandemia mostrou o uso de modelagem de tópico aplicados na busca de literatura relacionada à COVID-19 e à não aderência ao distanciamento social³⁶.

Quadro 7

Número de autorizações prévias por tópicos gerados pelo modelo BERTopic +1.000 com tratamento de palavras e respectiva descrição dos autores.

TÓPICOS MODELO BERTopicC +1.000 TRATADO			
TÓPICO	NÚMERO DE AUTORIZAÇÕES PRÉVIAS NOS TÓPICOS GERADOS PELO MODELO *	TÓPICOS GERADOS PELO MODELO **	DESCRIÇÃO DOS TÓPICOS RELEVANTES ***
-1	61.881	-1_psicoterapia_consulta_individual_vacina	
0	65.341	0_mg_direito_esquerdo_febre	
1	38.253	1_nan__	
2	3.222	2_f41_f52_f91_f42	Grupo CID-10 (transtornos mentais e comportamentais)
3	2.700	3_ok_just_partir_vie	
4	1.991	4_covid_teste_covide_sepse	Teste para COVID-19 e sepse
5	1.971	5_socorro_pronto_ortopedia_adulto	Consulta em pronto-socorro ortopédico
6	1.745	6_f33_f34_fr33_fffff33	
7	1.381	7_f32_f33_f38_	Grupo CID-10 (transtornos mentais e comportamentais)
8	1.289	8_dores_intensas_dore_peito	
9	1.286	9_suspeita_covid_supeita_familiar	Suspeita de COVID-19
10	1.221	10_terapia_terpia_teraoia_terapeutico	
11	1.049	11_cidf41_cidc41_tcg_cidf41p	
12	1.041	12_ps_yag_capsulotomia_laser	

CID-10: Classificação Internacional de Doenças, 10ª revisão.

Nota: o tópico -1 é considerado *outlier* segundo o modelo.

* Inclui todas as autorizações inclusive valores zerados;

** Tópicos gerados automaticamente pelo modelo;

*** Análise qualitativa do nome gerado pelo tópico pelos autores.

Em estudo comparando diferentes métodos de modelagem por tópicos em mídias sociais, Egger et al.³⁷ mostraram que o modelo BERTopic separou melhor os tópicos e suas ferramentas de análise possibilitam melhor entendimento das inter-relações entre os tópicos. Tais ferramentas são visuais e os autores afirmam que os tópicos precisam de interpretação humana.

Quanto à participação humana, é preciso uma visão holística e multidisciplinar, partindo da interpretação humana dos tópicos (dimensão modelagem) para o bem-estar do paciente (dimensão saúde) considerando aspectos financeiros (dimensão econômica).

Como exemplo do desafio dessa visão holística, observa-se que os modelos estudados têm comportamentos opostos: um seleciona casos graves e outro casos menos graves. A implantação de um programa de promoção à saúde no contexto da síndrome pós-COVID-19 é muito maior que a simples interpretação dos tópicos gerados por um modelo automático. Trata-se de um empreendimento multidisciplinar abrangendo também o desenho do programa, identificação e alocação correta dos pacientes, seu acompanhamento, avaliação dos desfechos e resultados financeiros.

Os pacientes com síndrome pós-COVID-19 necessitam de ampla gama de cuidados especiais, variando do restabelecimento das condições precedentes de saúde até a reabilitação³⁸. Nesse cenário, é importante observar que modelos gerados automaticamente e a interpretação de seus tópicos, ainda que interessantes, estão inseridos num processo altamente dependente de pessoas. Apesar de, no domínio da saúde, os recursos humanos serem especializados e caros, é imprescindível a participação

humana, não somente interpretando os tópicos gerados, mas também desenhando todo o programa alinhado com essa interpretação. Vale a pena utilizar um modelo de NLP na identificação precoce de doenças desde que uma equipe multidisciplinar leve à frente a tarefa de prover saúde de qualidade, acessível e sustentável aos pacientes.

Olhando especificamente a dimensão informacional, um modelo não supervisionado, especialmente quando não há tratamento de palavras, apresenta algumas vantagens. Não sofre influência do pesquisador. Outra vantagem é servir como suporte para os modelos supervisionados sendo empregados como técnicas exploratórias³⁹. A necessária interpretação humana casa-se perfeitamente num fluxo de descoberta de pacientes com os seguintes passos: (1) análise exploratória não supervisionada – objeto deste estudo; (2) interpretação humana e rotulagem, baseada no desenho do programa; (3) classificação dos casos; (4) aplicação dos rótulos em um modelo supervisionado com descoberta de novos pacientes. Um modelo supervisionado tem melhor desempenho e medidas diretas de avaliação da qualidade classificatória, mas a falta de rótulos em informações não estruturadas dificulta muito sua aplicabilidade.

Neste estudo foram empregados dois métodos indiretos de avaliação da qualidade. No primeiro, a análise e classificação humana dos pedidos de autorização do modelo BERTopic +1.000, selecionado em função de sua possível maior gravidade e simulando o passo classificação de casos por especialista. Esse exercício prático mostra a dependência da interpretação humana. Enquanto a maioria dos casos (90%) seria de interesse para uma avaliação criteriosa através de contato com o paciente, por exemplo, outros claramente foram classificados erroneamente (p.ex.: desconforto respiratório), mesmo que interessantes – um dos casos é um recém-nascido prematuro de mãe infectada por COVID-19, cuja análise pode ensejar um programa específico para grávidas nesse período pandêmico.

O segundo método de avaliação indireta da qualidade usou pesquisa de linguagem estruturada (SQL), indicando que os modelos BERTopic perdem um grupo significativo de pacientes suspeitos. Esses casos foram de menor gravidade. A perda não foi resolvida com alteração do número de documentos por tópicos – houve aumento de *outliers* – nem com o tratamento de palavras – os grupos ficaram menos identificáveis. Esses casos não classificados reforçam a necessidade de contexto semântico para a aplicação do método, que está atrelado à qualidade da informação do pedido de autorização. Somente 25% das autorizações prévias têm alguma informação e, entre elas, a maioria tem poucas palavras, dificultando a análise contextual pelo método. É antiga a discussão sobre qualidade dos dados e sua solução no processo de descoberta de conhecimento em bases de dados – KDD¹⁰. O uso de bases de dados reais, como a utilizada aqui, tem grande potencial, sendo inclusive passível de utilização nas evidências baseadas em dados reais desde que corrigidas as limitações impostas pela qualidade^{40,41}.

O modelo Word2Vec desempenhou melhor com o tratamento de palavras quando comparado a métodos tradicionais, em parte porque o tratamento envolveu a padronização das palavras COVID-19 escritas de diferentes formas. Apesar de vantajoso, isso expõe a dificuldade de manutenção de um modelo desse tipo e é preciso considerar se a pesquisa tradicional via SQL não seria melhor que esse modelo tratado. Contudo, deve-se considerar que formas tradicionais de extração de dados de textos estão sujeitas a erros humanos, a escolha *a priori* de palavras presentes nesse texto exigem conhecimento especializado⁴² e podem não aproveitar completamente as informações de mundo real. Opções tradicionais de análise de banco de dados para identificação dos pacientes com determinadas doenças em operadoras são limitadas – os CID-10 não são informados e os procedimentos pagos não permitem identificar a doença tratada (p.ex.: tomografia computadorizada de pulmão é paga da mesma forma para cânceres, infecções e *check-up*). Resta o acesso à ampla gama de informações não estruturadas, em que novos métodos, ainda que precisem de ajustes, podem ser mais eficazes.

Nesse ambiente real com baixa qualidade de informação, de alto volume autorizações prévias com valores faltantes ou preenchidos com apenas uma palavra, observa-se que o estudo demonstrou a viabilidade de um modelo não supervisionado para análise de autorizações prévias de operadora de plano de saúde sem nenhum tratamento prévio com uso de softwares livres, de fácil utilização e implementação. Esse tipo de modelo é especialmente útil na língua portuguesa onde “coronavírus” e “coronávirus” são palavras diferentes para o computador, mas com significados idênticos. Também endereça frases como – “FC: 65BPM FR: 26IPM PA:100/57MMGH SAT: 95% em AA. mantém quadro de desconforto respiratório” por “entender” que desconforto respiratório pode estar relacionado à COVID-19.

Inesperadamente o modelo gerou outros grupos de interesse. Notadamente um grupo de pacientes com câncer em que o tópico formado praticamente descreve o diagnóstico atribuído aos pacientes – “neoplasia, maligna, mama” e grupos de pacientes com problemas ortopédicos e transtornos mentais. Certamente são pacientes que podem se beneficiar de programas de promoção de saúde.

Por outro lado, um modelo não supervisionado selecionou autorização prévia pertencentes a pacientes com câncer. Isso levanta sérias preocupações sobre o tratamento ético e responsável das informações. Esse trabalho evidencia os problemas que esses modelos podem ocasionar no campo ético ⁴³, especialmente pelo foco na aplicação técnica de NLP desconsiderando-se a dimensão humana. Há necessidade da ampla participação humana em diferentes etapas da criação de um programa de promoção à saúde para pacientes com síndrome pós-COVID-19. Isso não torna o método menos importante, apenas reforça a necessidade de controle humano.

Até onde é dado saber, trata-se do primeiro estudo empregando essa técnica utilizando dados de saúde em saúde suplementar no Brasil.

Limitações do estudo

É um modelo pouco generalizável por conta de fatores como: (i) ser uma base proprietária; (ii) dificuldade de acesso a informações devido ao sigilo ético e legal; e (iii) pelo uso do modelo treinado em *corpus* não médicos de língua inglesa. Também foi observado que existe uma quantidade importante de autorizações com descrições pobres semanticamente, prejudicando a classificação. A avaliação da qualidade do modelo dependeu de análise manual do pesquisador principal, o que pode introduzir um viés que está sendo mitigado pela exposição da informação, e respectiva classificação.

Estudos adicionais

O modelo deve ser aprimorado por método supervisionado com a inclusão de rótulos criados por especialistas. Também pode ser enriquecido com outros métodos de aprendizado de máquina, como a análise das imagens anexadas às autorizações. É necessária a discussão ética sobre a aplicação de modelos automatizados, especialmente quando classificam pessoas em grupos de doenças. É preciso avaliar qual é o impacto dos regimes e objetivos de tratamento (p.ex.: ambulatorial e diagnóstico) no comportamento dos modelos. São necessários estudos adicionais sobre o inter-relacionamento de diferentes dimensões de conhecimento e respectivos profissionais na provisão de uma atenção integrativa, colaborativa e sustentável.

Conclusão

O modelo BERTopic sem tratamento de palavras selecionou pacientes suspeitos de infecção por COVID-19 mais graves que o modelo Word2Vec com tratamento de palavras. Por outro lado, o último modelo foi capaz de selecionar um grupo maior de casos suspeitos. Observa-se que a decisão sobre o melhor modelo depende da análise humana complementar e do programa de promoção à saúde desenhado.

Comparados a métodos tradicionais, observou-se que os modelos BERTopic não classificaram casos suspeitos, em sua maioria de menor gravidade, mas que podem ser relevantes em um modelo integrado de atenção à saúde. Reforça-se, assim, o caráter exploratório, seu uso intermediário para aplicação de um modelo supervisionado e a necessidade de comparação de resultados com métodos tradicionais de pesquisa.

Por outro lado, o modelo gerou tópicos de interesse para estudos futuros, especial atenção aos casos suspeitos de pacientes com câncer.

Os achados demonstram a importância da participação humana – análise dos tópicos gerados para correta classificação gerando informações para um modelo supervisionado, escolha do melhor modelo de acordo com a perspectiva de gerenciamento de saúde (pacientes mais graves *versus* menos graves), desenho de um programa de promoção à saúde (alinhado a essa escolha) e atenção aos aspectos éticos do uso de ferramentas de aprendizado de máquina em saúde.

Colaboradores

R. P. Silva contribuiu com a concepção e desenho do estudo, metodologia, aquisição e análise dos dados, redação e revisão; e aprovou a versão final. J. T. Pollettini contribuiu com a metodologia, análise dos dados e revisão crítica; e aprovou a versão final. A. Pazin Filho contribuiu com o desenho do estudo, metodologia, análise dos dados, redação e revisão crítica; e aprovou a versão final.

Informações adicionais

ORCID: Rildo Pinto da Silva (0000-0001-5718-2747); Juliana Tarossi Pollettini (0000-0002-4894-249X); Antonio Pazin Filho (0000-0001-5242-329X).

Conflito de interesses

R. P. Silva declara que é consultor, atuando na prestação de serviços para a saúde suplementar.

Referências

1. Adil MT, Rahman R, Whitelaw D, Jain V, Al-Ta'an O, Rashid F, et al. SARS-CoV-2 and the pandemic of COVID-19. *Postgrad Med J* 2021; 97:110-6.
2. Noronha KVMS, Guedes GR, Turra CM, Andrade MV, Botega L, Nogueira D, et al. The COVID-19 pandemic in Brazil: analysis of supply and demand of hospital and ICU beds and mechanical ventilators under different scenarios. *Cad Saúde Pública* 2020; 36:e00115320.
3. Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, et al. Epidemiology of COVID-19: a systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. *J Med Virol* 2021; 93:1449-58.
4. Victora CG, Hartwig FP, Vidaletti LP, Martorell R, Osmond C, Richter LM, et al. Effects of early-life poverty on health and human capital in children and adolescents: analyses of national surveys and birth cohort studies in LMICs. *Lancet* 2022; 399:1741-52.
5. Mannelli C. Whose life to save? Scarce resources allocation in the COVID-19 outbreak. *J Med Ethics* 2020; 46:364-66.
6. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. *BMJ* 2021; 374:n1648.
7. Hope AA, Evering TH. Postacute sequelae of severe acute respiratory syndrome coronavirus 2 infection. *Infect Dis Clin North Am* 2022; 36:379-95.
8. Pavli A, Theodoridou M, Maltezou HC. Post-COVID syndrome: incidence, clinical spectrum, and challenges for primary healthcare professionals. *Arch Med Res* 2021; 52:575-81.
9. Khan A, Khan M, Ullah S, Wei D-Q. Hantavirus: the next pandemic we are waiting for? *Interdiscip Sci* 2021; 13:147-52.
10. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Magazine* 1996; 17:37-54.
11. Dinov ID. Volume and value of big healthcare data. *J Med Stat Inform* 2016; 4:3.
12. Esfandiari N, Babavalian MR, Moghadam A-ME, Tabar VK. Knowledge discovery in medicine: current issue and future trend. *Expert Systems with Applications* 2014; 41:4434-63.
13. Cios KJ, Kurgan LA. Trends in data mining and knowledge discovery. In: Pal NR, Jain L, editors. *Advanced techniques in knowledge discovery and data mining*. London: Springer London; 2005. p. 1-26.
14. Idri A, Benhar H, Fernández-Alemán JL, Kadi I. A systematic map of medical data preprocessing in knowledge discovery. *Comput Methods Programs Biomed* 2018; 162:69-85.
15. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry MW, Mohamed A, Yap BW, editors. *Supervised and unsupervised learning for data science*. Cham: Springer International Publishing; 2020. p. 3-21.

16. Dogan O, Tiwari S, Jabbar MA, Guggari S. A systematic review on AI/ML approaches against COVID-19 outbreak. *Complex Intell Systems* 2021; 7:2655-78.
17. Lauriola I, Lavelli A, Aiolli F. An introduction to deep learning in natural language processing: models, techniques, and tools. *Neurocomputing* 2022; 470:443-56.
18. Junaid T, Sumathi D, Sasikumar AN, Suthir S, Manikandan J, Khilar R, et al. A comparative analysis of transformer based models for figurative language classification. *Comput Elect Eng* 2022; 101:108051.
19. Agência Nacional de Saúde Suplementar. TISS – padrão para troca de informação de saúde suplementar. <https://www.gov.br/ans/pt-br/assuntos/prestadores/padroao-para-troca-de-informacao-de-saude-suplementar-2013-tiss> (accessed on 20/Dec/2021).
20. Agência Nacional de Saúde Suplementar. Resolução Normativa nº 259, de 17 de junho de 2011. Dispõe sobre a garantia de atendimento dos beneficiários de plano privado de assistência à saúde e altera a Instrução Normativa – IN nº 23, de 1º de dezembro de 2009, da Diretoria de Normas e Habilitação dos Produtos DIPRO. *Diário Oficial da União* 2011; 20 jun.
21. Instituto Brasileiro de Geografia e Estatística. Conta-satélite de saúde: Brasil – 2010-2019. https://biblioteca.ibge.gov.br/visualizacao/livros/liv101928_informativo.pdf (accessed on 07/Jul/2022).
22. Duval FV, Silva FAB. Mining in Twitter for adverse events from malaria drugs: the case of doxycycline. *Cad Saúde Pública* 2019; 35:e00033417.
23. Moreira LB, Namen AA. A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Comput Methods Programs Biomed* 2018; 165:139-49.
24. Diniz EJS, Fontenele JE, Oliveira AC, Bastos VH, Teixeira S, Rabêlo RL, et al. *Boamente*: a natural language processing-based digital phenotyping tool for smart monitoring of suicidal ideation. *Healthcare (Basel)* 2022; 10:698.
25. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. *Springerplus* 2016; 5:1608.
26. Alghamdi R, Alfalqi K. A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications* 2015; 6:147-56.
27. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* 2018; 9 feb. <https://arxiv.org/abs/1802.03426>.
28. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. <https://maartengr.github.io/BERTopic/algorithm/algorithm.html> (accessed on 14/Dec/2022).
29. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. *arXiv* 2019, 27 aug. <https://arxiv.org/abs/1908.10084>.
30. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv* 2022; 24 may. <https://arxiv.org/pdf/1810.04805.pdf>.
31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need: 31st Conference on Neural Information Processing Systems (NIPS 2017), https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on 17/Oct/2023).
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv* 2013; 16 jan. <https://arxiv.org/abs/1301.3781>.
33. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf (accessed on 17/Oct/2023).
34. Hernández MA, Stolfo SJ. Real-world data is dirty: data cleansing and the merge/purge problem. *Data Min Knowl Discov* 1998; 2:9-37.
35. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJT, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digit Med* 2021; 4:96.
36. Chen Q, Leaman R, Allot A, Luo L, Wei C-H, Yan S, et al. Artificial intelligence in action: addressing the COVID-19 pandemic with natural language processing. *Annu Rev Biomed Data Sci* 2021; 4:313-39.
37. Egger R, Yu J. A topic modeling comparison Between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Front Sociol* 2022; 7:886498.
38. Leavell HR. The basic unity of private practice and public health. *Am J Public Health Nations Health* 1953; 43:1501-6.
39. Nadif M, Role F. Unsupervised and self-supervised deep learning approaches for biomedical text mining. *Brief Bioinform* 2021; 22:1592-603.
40. Liu F, Demosthenes P. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol* 2022; 22:287.
41. Raouf S, Kurzrock R. For insights into the real world, consider real-world data. *Sci Transl Med* 2022; 14:eabn6911.
42. Chen T, Dredze M, Weiner JP, Hernandez L, Kimura J, Kharrazi H. Extraction of geriatric syndromes from electronic health record clinical notes: assessment of statistical natural language processing methods. *JMIR Med Inform* 2019; 7:e13039.
43. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet* 2020; 395:1579-86.

Abstract

Patients with post-COVID-19 syndrome benefit from health promotion programs. Their rapid identification is important for the cost-effective use of these programs. Traditional identification techniques perform poorly especially in pandemics. A descriptive observational study was carried out using 105,008 prior authorizations paid by a private health care provider with the application of an unsupervised natural language processing method by topic modeling to identify patients suspected of being infected by COVID-19. A total of 6 models were generated: 3 using the BERTopic algorithm and 3 Word2Vec models. The BERTopic model automatically creates disease groups. In the Word2Vec model, manual analysis of the first 100 cases of each topic was necessary to define the topics related to COVID-19. The BERTopic model with more than 1,000 authorizations per topic without word treatment selected more severe patients – average cost per prior authorization paid of BRL 10,206 and total expenditure of BRL 20.3 million (5.4%) in 1,987 prior authorizations (1.9%). It had 70% accuracy compared to human analysis and 20% of cases with potential interest, all subject to analysis for inclusion in a health promotion program. It had an important loss of cases when compared to the traditional research model with structured language and identified other groups of diseases – orthopedic, mental and cancer. The BERTopic model served as an exploratory method to be used in case labeling and subsequent application in supervised models. The automatic identification of other diseases raises ethical questions about the treatment of health information by machine learning.

COVID-19; Natural Language Processing; Health Care; Selection Criteria; Proprietary Health Facilities

Resumen

Los pacientes con síndrome pos-COVID-19 pueden beneficiarse de los programas de promoción de la salud. Su rápida identificación es importante para el uso efectivo de estos programas. Las técnicas de identificación tradicionales no tienen un buen desempeño, especialmente en pandemias. Se realizó un estudio observacional descriptivo, con el uso de 105.008 autorizaciones previas pagadas por un operador de salud privado mediante la aplicación de un método no supervisado de procesamiento del lenguaje natural mediante modelado temático para identificar a los pacientes sospechosos de estar infectados por COVID-19. Se generaron 6 modelos: 3 con el uso del algoritmo BERTopic y 3 modelos Word2Vec. El modelo BERTopic crea automáticamente grupos de enfermedades. En el modelo Word2Vec para definir temas relacionados con la COVID-19, fue necesario el análisis manual de los primeros 100 casos de cada tema. El modelo BERTopic con más de 1.000 autorizaciones por tema sin tratamiento de palabras seleccionó a pacientes más graves: costo promedio por autorizaciones previas pagada de BRL 10.206 y gasto total de BRL 20,3 millones (5,4%) en 1.987 autorizaciones previas (1,9%). Además, contó con el 70% de aciertos en comparación con el análisis humano y el 20% de los casos con potencial interés, todos los cuales pueden analizarse para su inclusión en un programa de promoción de la salud. Hubo una pérdida significativa de casos en comparación con el modelo tradicional de investigación con lenguaje estructurado y se identificó otros grupos de enfermedades: ortopédicas, mentales y cáncer. El modelo BERTopic sirvió como un método exploratorio para ser utilizado en el etiquetado de casos y su posterior aplicación en modelos supervisados. La identificación automática de otras enfermedades plantea preguntas éticas sobre el tratamiento de la información de salud mediante el aprendizaje de máquina.

COVID-19; Procesamiento de Lenguaje Natural; Atención a la Salud; Criterios de Selección de Pacientes; Instituciones Privadas de Salud

Recebido em 19/Jan/2023

Versão final reapresentada em 26/Jun/2023

Aprovado em 04/Jul/2023