

Quais os derivados lácteos mais consumidos no Brasil durante a pandemia da COVID-19?¹

*Thallys da Silva Nogueira*²
*Anna Letícia Franco Monteiro*³
*Darlan Henrique Da Costa Silva*⁴
*Kennya Beatriz Siqueira*⁵
Priscila Vanessa Zabala Capriles Goliatt^{6,7}

Resumo: A pandemia da COVID-19 trouxe diversos impactos para a vida cotidiana. Um destes, está relacionado com a alimentação. A fim de analisar como foi o consumo de produtos lácteos durante a pandemia, a ferramenta Observatório do Consumidor foi empregada para analisar dados coletados da rede social Twitter com o objetivo de responder à pergunta “Quais foram os derivados lácteos mais consumidos no Brasil durante a pandemia da COVID-19?”. Foi observado que sorvete, leite condensado, queijos, doce de leite e a manteiga foram os cinco produtos lácteos que obtiveram maior número de menções a verbos que remetem ao consumo, caracterizando-os como os mais consumidos no período em questão. Este resultado mostra que a coleta, o processamento e a análise de dados de rede social feita pelo Observatório do Consumidor inova a forma de responder a perguntas de pesquisa de mercado de forma rápida, representativa e barata sem a necessidade da aplicação de questionários.

Palavras-chave: consumo de lácteos, mineração de dados, observatório do consumidor, pesquisa de mercado

What are the most consumed dairy products in Brazil during the COVID-19 pandemic?

Abstract: The COVID-19 pandemic has brought several impacts to everyday life. One of these is related to food. In order to analyze how the consumption of dairy products was during the pandemic, the Observatório do Consumidor tool was used to analyze data collected from the social network Twitter in order to answer the question “What were the most consumed dairy products in Brazil during the pandemic of COVID-19?”. It was found that ice cream, condensed milk, cheeses, dulce de leche and butter were the 5 dairy products that had the highest number of mentions of verbs that refer to consumption, characterizing them as the most consumed in the period in question. This result shows that the collection, processing and analysis of social network data carried out by the Observatório do Consumidor innovates the way to answer market research questions quickly, representatively and cheaply without the need for questionnaires.

Keywords: dairy consumption, data mining, consumer observatory, market research

¹ O presente trabalho foi realizado com o apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil: (a) Parte da tese de doutorado do primeiro autor, financiada pelo CNPq, (b) Bolsista do CNPq pelo Programa de Mestrado e Doutorado em Inovação MAI/DAI.

² Doutorando em Modelagem Computacional, Programa de Pós Graduação em Modelagem Computacional-UFJF, Juiz de Fora, MG. e-mail: thallys.nogueira@estudante.ufjf.br

³ Graduanda em Engenharia Computacional – UFJF. e-mail: anna.franco@estudante.ufjf.br

⁴ Graduando em Sistema de Informação – UFJF. e-mail: darlan.silva@ice.ufjf.br

⁵ Pesquisadora, Embrapa Gado de Leite - Juiz de Fora, MG. e-mail: kennya.siqueira@embrapa.br

⁶ Professora-orientadora, Programa de Pós Graduação em Modelagem Computacional-UFJF. e-mail: capriles@ice.ufjf.br

⁷ Orientadora

Introdução

Analisar o consumo dos derivados lácteos por meio de dados de redes sociais é o que o Observatório do Consumidor (OC) tem feito atualmente. Semanalmente, cerca de 165.511 publicações do Twitter denominados *tweets* são coletadas por meio de um conjunto de 293 palavras-chave¹ que se dividem em 10 categorias de lácteos². As redes sociais podem ser consideradas como um dos principais meios de comunicação atualmente e seu uso se intensificou muito durante o período de quarentena causado pela descoberta do novo coronavírus (COVID-19). Devido ao alto poder de contágio da doença e os elevados números de pessoas infectadas e levadas a óbito, em março de 2020, a Organização Mundial da Saúde declarou pandemia global, impondo diversas medidas como o isolamento social a fim de desacelerar o avanço e crescimento da doença enquanto desenvolvia-se a vacina para a mesma.

A mudança repentina na rotina das pessoas trouxe consigo diversos impactos. Com um alto número de pessoas em isolamento, o consumo tanto de conteúdo de redes sociais quanto de certos alimentos aumentou servindo como entretenimento e válvula de escape para alívio do estresse causado pelo novo normal imposto pela pandemia. Pesquisas (SIQUEIRA *et al.*, 2020a; SIQUEIRA *et al.*, 2020b) mostraram que alterações nas tendências de consumo ocorreram, com um incremento do consumo de caráter indulgente com o objetivo de minimizar as situações de estresse e ansiedade.

O propósito do OC é o de inovar a maneira de se conduzir pesquisas de mercado tradicionais por meio da análise de dados de redes sociais. Neste sentido, o foco deste trabalho está diretamente relacionado com a extração de informações contidas em dados coletados do Twitter sobre os lácteos no Brasil para responder à pergunta: “Quais os derivados lácteos mais consumidos no Brasil durante a pandemia da COVID-19?” no período compreendido entre 07/05/2020⁴ e 20/11/2021⁵.

Material e Métodos

A arquitetura do OC é segmentada em 3 principais módulos que são (i) coleta e armazenamento de dados, (ii) processamento de dados e por fim (iii) mineração de dados e extração de informações. O fluxo de tarefas é iniciado com a seleção das palavras-chave (nome dos produtos lácteos de interesse) e submissão das mesmas ao algoritmo de coleta e armazenamento dos dados. As publicações coletadas foram armazenadas em um banco de dados relacional. Com o objetivo de extrair informações relevantes destes dados, o processamento dos dados é realizado padronizando cada um dos *tweets* com a remoção de *stopwords*, *emojis* e *links* de sites que de forma recorrente aparecem nos *tweets* e não carregam informação interessante consigo. Com os dados padronizados inicia-se então a mineração dos dados para a extração das informações de interesse.

¹ Lista de palavras-chave construída com a supervisão e orientação de especialistas da Embrapa Gado de Leite e representantes da Empresa de Assistência Técnica e Extensão Territorial (EMATER).

² Bebidas Lácteas, Creme de Leite, Doce de Leite, Iogurte, Leite, Leite Condensado, Leite Fermentado, Manteiga, Queijos e Sorvete.

³ Alimentos que são consumidos em situações de estresse por proporcionarem alívio e satisfação.

⁴ Início da coleta de dados do Observatório do Consumidor.

⁵ Marca de 60% da população brasileira com o protocolo inicial de vacinação contra a COVID-19 (primeira e segunda dose) concluída (Mathieu *et al.*, 2021).

Para responder à pergunta problema mencionada na introdução, utilizou-se a linguagem de programação *Python* (VAN ROSSUM *et al.*, 2009) em conjunto com diversas bibliotecas sendo a *spaCy* (HONNIBAL *et al.*, 2017), *NLTK* (*Natural Language Toolkit*) (BIRD *et al.*, 2019) as mais utilizadas. Ambas foram utilizadas para extração de informações referentes ao consumo dos lácteos no conteúdo dos *tweets*.

Utilizando a biblioteca *spaCy*, a ocorrência da relação sintática entre verbo e substantivo foi mapeada em todos os *tweets* coletados com o objetivo de identificar os principais verbos e substantivos que indicam respectivamente o ato de consumir e os nomes dos lácteos que são analisados pelo OC.

Resultados e Discussão

Ao todo 19.033.823 *tweets* sobre lácteos foram coletados pelo OC. Destes, 70,3% dos *tweets* apresentaram pelo menos uma ação no contexto dos lácteos e possuíam ao menos um verbo podendo este ser referente ao consumo ou não. Como o objetivo deste trabalho é o de extrair informações sobre o consumo, analisou-se em específico a ocorrência de cinco principais verbos⁶ que remetem ao consumo com suas respectivas flexões verbais. Para afirmar que de fato houve o consumo de lácteos foram considerados verbos flexionados nos tempos verbais do passado e do presente.

Após realizar a filtragem dos dados, 1.705.383 *tweets* foram selecionados por serem publicados durante a pandemia e apresentarem algum dos verbos de interesse. Os resultados mostraram que os cinco lácteos mais consumidos foram: sorvete (604.784), leite condensado (189.435), queijos (55.483), doce de leite (83.523), manteiga (61.107). A Figura 1 mostra em detalhes a proporção de menções a cada um dos verbos de consumo por lácteo.

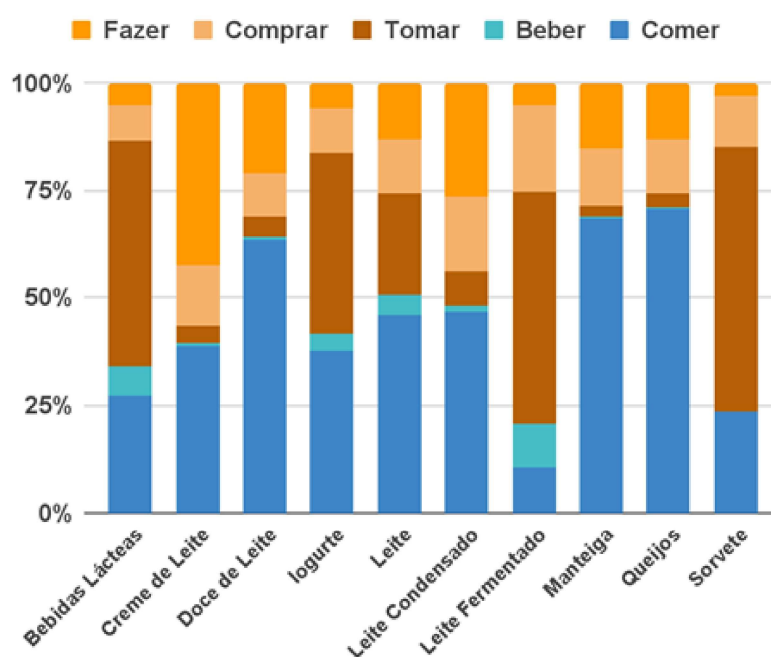


Figura 2. - Proporção de menções aos verbos de consumo por derivado lácteo durante a pandemia.

Fonte: resultados da pesquisa.

⁶ Comer, beber, tomar, comprar e fazer.

Cada produto possui um modo de consumo distinto. Da Figura 1 é possível verificar que os verbos **fazer** e **comprar**, que estão intimamente ligados ao ato de ter produtos necessários para seguir passos requeridos em receitas culinárias, os produtos creme de leite e leite condensado foram os que obtiveram maior expressão representando respectivamente 44,4% e 23,1%. Para o verbo **tomar** os produtos sorvete, leite fermentado, bebidas lácteas e iogurte tiveram os maiores números de menções a essa ação. Vale destacar que o sorvete foi consideravelmente o produto mais “tomado” no tempo em questão, com 86,9% dessa ação. O verbo beber não obteve uma expressão tão significativa como os demais, entretanto, os produtos lácteos líquidos que são os casos de leite fermentado, bebidas lácteas, leite e iogurte foram os mais expressivos em número de menções a esse verbo. Por fim, o verbo comer esteve presente em todos os produtos analisados, sendo os de maior destaque: queijos, manteiga, doce de leite e leite condensado. Os *tweets* mostraram que os queijos foram mais consumidos isoladamente ou em lanches delivery e a manteiga esteve presente em receitas e no café da manhã. Já o doce de leite e o leite condensado, produtos de caráter indulgente, foram amplamente consumidos em diversos momentos e opções de consumo, mostrando como esses lácteos fizeram parte da rotina alimentar dos brasileiros.

Conclusões

Esse trabalho mostrou que a coleta, o processamento e a análise de dados de rede social feita pelo Observatório do Consumidor inova a forma de responder a perguntas de uma pesquisa de mercado de forma rápida, representativa e barata sem a necessidade da aplicação de questionários. Com essa ferramenta foi possível identificar os derivados lácteos mais consumidos durante a pandemia de Covid-19 no Brasil.

Para trabalhos futuros espera-se ampliar as questões a serem respondidas com perguntas como: “Qual o sentimento após o consumo?”, “Quem consumiu?”, “Quando consumiu?” e “Quais estados brasileiros que obtiveram maior consumo?” dentre outras, além de explorar dados de outras redes sociais.

Agradecimentos

Agradecemos à Embrapa Gado de Leite, ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora e ao Grupo de Modelagem Computacional Aplicada por todo apoio e incentivo no desenvolvimento deste e diversos outros trabalhos produzidos.

Referências

SIQUEIRA, K. B. *et al.* Análise exploratória da imagem dos lácteos em tempos de coronavírus. **Indústria de laticínios**, n. 143, p. 64-66, 2020. ISSN 1678-7250.

SIQUEIRA, K. B. *et al.* O impacto da pandemia no consumo de lácteos no Brasil. **Indústria de laticínios**, n. 147, p. 36-38, 2020. ISSN 1678-7250.

MATHIEU, E., RITCHIE, H., ORTIZ-OSPINA, E., ROSER, M., HASELL, J., APPEL, C., GIATTINO, C., RODÉS-GUIRAO, L. A global database of COVID-19 vaccinations. **Nature Human Behaviour**, v. 5, n. 947–953, 2021.. doi: 10.1038/s41562-021-01122-8

VAN ROSSUM G. *et al.* **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace; 2009.

HONNIBAL, M. *et al.* **spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing**. 2017.

BIRD S. *et al.* **Natural language processing with Python: analyzing text with the natural language toolkit**. " O'Reilly Media, Inc." 2009.