

**XXV Workshop de Iniciação Científica da Embrapa Gado de Leite**  
Juiz de Fora, MG – 5 de março de 2020

**Análise exploratória de leite e derivados mais comentados no Twitter e Google Trends**

Thallys Nogueira<sup>1</sup>, Nedson D. Soares<sup>2</sup>, Emerson Campos<sup>3</sup>, Kennya Siqueira<sup>4</sup>, Priscila Capriles Goliatt<sup>5</sup>

<sup>1</sup>Mestrando em Modelagem Computacional - UFJF/Minas Gerais. e-mail: thallysnogueira@ice.ufjf.br

<sup>2</sup>Mestrando em Ciência da Computação - UFJF/Minas Gerais. e-mail: nedson@ice.ufjf.br

<sup>3</sup>Graduando em Engenharia Mecatrônica -IF Sudeste- Juiz de Fora/Minas Gerais, email: emersonwendelim@gmail.com

<sup>4</sup>Pesquisadora, Embrapa Gado de Leite - Juiz de Fora/Minas Gerais, email: kennya.siqueira@embrapa.br

<sup>5</sup>Pesquisadora-orientadora, Programa de Pós Graduação em Modelagem Computacional-UFJF/Minas Gerais. email: capriles@ice.ufjf.br

**Resumo:** Nos dias de hoje, com o grande uso das redes sociais, um enorme volume de dados é gerado a todo momento, provenientes das postagens de milhares de internautas conectados. Por esse motivo, analisar o que os usuários dizem pode fornecer informações importantes para novas tendências no mercado. Utilizando a técnica de mineração de dados, o objetivo deste trabalho foi realizar uma análise exploratória do leite e seus derivados no Twitter e Google Trends, quantificando quais são os produtos lácteos mais comentados nessas redes. Os resultados obtidos mostram que os quatro produtos mais comentados no intervalo de 12 a 19 de fevereiro de 2020 foram o queijo, sorvete, leite condensado e doce de leite. Com exceção do queijo, todos produtos mais citados são considerados alimentos com maior apelo de indulgência no Brasil. Foi possível observar também que há uma maior concentração de pessoas falando sobre estes produtos nas regiões Sul e Sudeste do Brasil, que são as maiores consumidoras de lácteos do Brasil. Portanto, este trabalho sugere que a técnica de mineração de dados pode facilmente identificar fatores inerentes a tendências e mercado consumidor brasileiro de lácteos.

**Palavras-chave:** lácteos, mineração de dados, rtweet, tendências de mercado, consumo, Google Trends

**Exploratory analysis of milk and dairy products most commented on Twitter and Google Trends**

**Abstract:** Nowadays, with the social networks, a huge volume of data is generated every day, from the posts of thousands of Internet users. So, analyzing this data can provide important information about new trends in the market. Using the data mining technique, the objective of this work was to carry out an exploratory analysis of milk and dairy products on Twitter and Google Trends, quantifying which dairy products are most commented on these networks. The results showed that the four most commented products in the period of February, 12 to 19,2020 were cheese, ice cream, condensed milk and dulce de leche. Except by cheese, all the most commented dairy products are considered foods with the greatest appeal for indulgence in Brazil. It was also possible to observe that there is a higher concentration of people talking about these products in the South and Southeast regions of Brazil. Then, this work suggests that the data mining technique could easily identify inherent factors to trends and Brazilian dairy consumers.

**Keywords:** products, data mining, rtweet, market trends, consumption, Google Trends

**XXV Workshop de Iniciação Científica da Embrapa Gado de Leite**  
Juiz de Fora, MG – 5 de março de 2020

### **Introdução**

O leite é um dos produtos mais consumidos no mundo e está presente em alimentos como manteiga, biscoitos, queijos, chocolates, dentre outros. De acordo com Carvalho (2020), a cadeia produtiva do leite movimenta cerca de US\$ 10 bilhões, empregando 3 milhões de pessoas. Segundo Siqueira (2019), dentre os lácteos mais consumidos no Brasil em 2016, o leite UHT e os queijos representavam, conjuntamente, cerca de 49,7% do valor de vendas das categorias de produtos lácteos no Brasil.

No entanto, entender o perfil do consumidor de lácteos e as tendências desse mercado é uma tarefa complicada. Mas com o advento das redes sociais, a captura dessas informações tornou-se mais rápida, menos dispendiosa e confiável. Com um crescente número de usuários conectados na rede, empresas e produtores podem se aproveitar desse grande volume de dados gerados com o intuito de obter novos *insights* sobre as atuais tendências de consumo.

O Twitter é uma rede social que oferece aos seus usuários um espaço para conversação, compartilhamento de conteúdos escritos, fotografias e vídeo. De acordo com In Digital (2020) essa é a sexta rede social mais usada no Brasil, com cerca de 48% da população utilizando essa plataforma.

O Google Trends é uma ferramenta capaz de acompanhar a evolução do número de buscas por uma determinada palavra-chave ao longo do tempo. Com ela é possível analisar e entender o que usuários buscam no site do Google.

Diante disso, o objetivo geral deste trabalho foi realizar uma análise exploratória do leite e seus derivados mais comentados no Twitter e Google Trends como forma de entender melhor as tendências do mercado lácteo nacional.

### **Material e Métodos**

Utilizando a técnica de mineração de dados (Cortês, 2002), foi desenvolvido um *script* em R para a realização de uma busca no Twitter utilizando 16 palavras-chave como “queijo” e “sorvete”, obtidas a partir de Árvore do Leite (2020) e de Siqueira (2019). Utilizando o Google Trends, foi feita uma comparação dos quatro derivados do leite mais comentados no Twitter no período de 12 a 19 de fevereiro de 2020.

### **Resultados e Discussão**

A pesquisa identificou 91.202 postagens no Twitter referentes a leite e derivados. Os derivados com maior número de comentário são apresentados na Figura 1.

**XXV Workshop de Iniciação Científica da Embrapa Gado de Leite**  
Juiz de Fora, MG – 5 de março de 2020

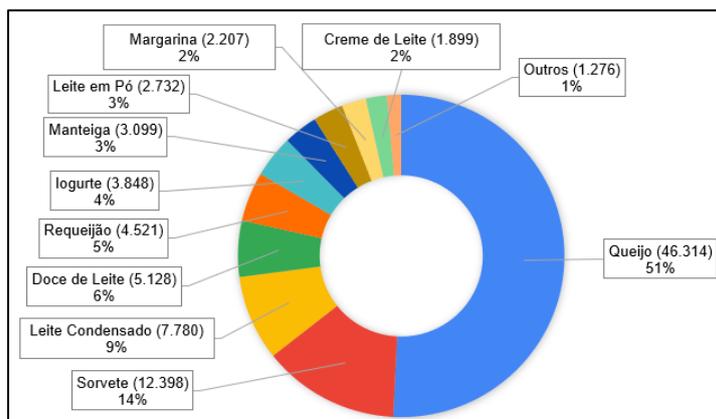


Figura 1. Gráfico apresentando os derivados lácteos mais comentados no Twitter.

Em laranja claro estão representados os produtos com número de comentários pouco representativo: farinha láctea (818), leite desnatado (256), leite integral (179), leite pasteurizado (16) e leite UHT (7).

Pela Figura 1, pode-se observar que dentre os principais *trends*, a palavra “queijo” apresentou maior destaque com 50,8% em aparições nos comentários dos internautas, seguido de sorvete, leite condensado e doce de leite. Esses quatro produtos mais comentados foram analisados no Google Trends e apresentados na Figura 2. Nessa figura, valores mais próximos de 100 indicam maior interesse dos usuários por um determinado tema.

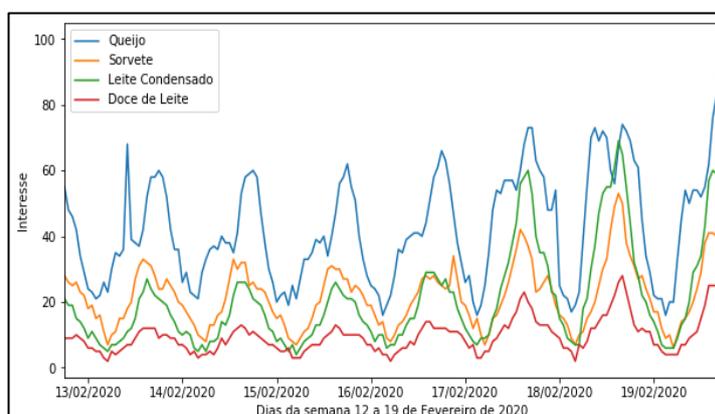


Figura 2. Gráfico apresentando o percentual de interesse pelos produtos lácteos por dias no intervalo de 12 a 19 de fevereiro de 2020, de acordo com o Google Trends.

É possível observar pela Figura 2 que há um padrão e que a palavra “queijo” manteve-se superior às demais palavras pesquisadas, durante todo o período estudado. Já a expressão “leite condensado” apresentou um maior interesse em certas horas dos dias quando comparado à palavra “sorvete” a partir do dia 16 de fevereiro de 2020. É importante destacar que os 4 termos mais comentados não correspondem aos produtos mais consumidos (exceto queijos), mas sim aos produtos que tem maior apelo de indulgência no Brasil (exceto queijos).

Dentre os usuários que comentaram os produtos e que estavam com a localização dos dispositivos ativadas no momento da postagem, pode-se observar que as regiões que mais

**XXV Workshop de Iniciação Científica da Embrapa Gado de Leite**  
Juiz de Fora, MG – 5 de março de 2020

comentaram sobre os derivados do leite foram as regiões Sul e Sudeste, que coincidem com as regiões de maior renda e consumo *per capita* do Brasil.



Figura 3: Mapa do Brasil apresentando a localização (pontos em azul) das postagens no Twitter sobre leites e derivados.

### Conclusões

Este trabalho sugere que as técnicas de mineração de dados podem facilmente identificar fatores e características inerentes a tendências e mercado consumidor de lácteos no Brasil. Para trabalhos futuros, uma sugestão é a utilizar dados de outras redes sociais para fazer uma análise comparativa do que tem sido comentado utilizando técnicas de aprendizado de máquina e análise de sentimentos para classificação destes comentários, bem como ampliar o período de análise.

### Agradecimentos

Ao Programa Residência Zootécnica Digital da Embrapa Gado de Leite pela concessão da bolsa.

### Referências

ÁRVORE DO LEITE. **Árvore Genealógica do Leite**. Available at: <<https://www.arvoredoleite.org/>>. Accessed on: feb. 26 2020.

CARVALHO L.A, NOVAES L.P, MARTINS C.E *et al.* **Sistema de produção de Leite (Cerrado)**. Available at: <<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Leite/LeiteCerrado/introducao.html>>. Accessed on: feb. 26 2020.

CORTÊS S.C, PORCARO R.M, LIFSCHITZ S. **Mineração de dados - Funcionalidades, Técnicas e Abordagens**. 2002.

RELATÓRIO IN DIGITAL 2020. Available at: <<https://datareportal.com/reports/digital-2020-brazil>>. Accessed on: feb. 26 2020.

SIQUEIRA K.B. Consumo de leite e derivados no Brasil. In: **Anuário do leite 2019**. Embrapa. 2019