



## MINERAÇÃO DE DADOS NA PRODUÇÃO DO QUEIJO MINAS ARTESANAL DATA MINING IN THE PRODUCTION OF MINAS ARTESANAL CHEESE

BARBOSA, T.H.G.<sup>1</sup>.; BARBOSA, T.N.T.<sup>2</sup>

<sup>1</sup>Discente. Bacharelado em Sistemas de Informação. Instituto Federal do Norte de Minas Gerais - *Campus* Porteirinha; <sup>2</sup>Docente. Bacharelado em Sistemas de Informação. Instituto Federal do Norte de Minas Gerais - *Campus* Porteirinha.

### Introdução

As técnicas de mineração de dados são métodos alternativos de fornecer extração de conhecimento e descoberta de padrões, podendo auxiliar na solução de problemas, bem como evidenciar oportunidades ocultas em relacionamentos de dados. Desta forma, pode-se diagnosticar o comportamento do negócio, e a depender do resultado, uma intervenção pode ser necessária (Cardoso e Machado, 2008).

Com base nesse contexto, o emprego de técnicas de mineração de dados poderia ser útil em áreas do ramo alimentício, como por exemplo, na produção do Queijo Minas Artesanal, que segundo a EPAMIG (2019), queijos artesanais enfrentam desafios na prevenção e correção de defeitos, devido à variação não controlável do leite. Matsumoto et al. (2016) afirmam que o queijo, rico em proteínas e gorduras, é suscetível à contaminação por coliformes, ameaçando a segurança do consumidor. Sendo assim, este trabalho se propôs aplicar técnicas de mineração de dados com imagens, a fim de auxiliar o processo de produção do Queijo Minas Artesanal na identificação dos queijos que possuem características que indiquem a presença de coliformes no processo de fabricação.

### Material e Métodos

#### *Coleta de Imagens*

Foram coletadas imagens de queijos saudáveis através de fábricas localizadas na região de Porteirinha-MG, bem como imagens de queijos com a presença de coliformes através de banco de dados público disponível na internet.

#### *Software e Pré-Processamento de Imagens*

Foi utilizado a ferramenta WEKA que é um software de código aberto para o pré-processamento (aplicação de filtros) e classificação dos dados. As imagens coletadas foram divididas em 2 classes (Coliforme e Bom), para que posteriormente fosse utilizado o filtro *COLOR LAYOUT*, que permite calcular médias de valores entre cores RGB e pixels presentes nas imagens, e transformá-las em atributos. A seguir, ocorreu a aplicação do filtro *INTERQUARTILE RANGE*, para detectar valores discrepantes e valores extremos dos atributos das imagens. Por fim, o filtro *RemoveWithValues* foi utilizado para a remoção destes atributos discrepantes.

#### *Algoritmo e Classificação*

Para o treinamento da base de dados, foi utilizado o algoritmo *Random Forest* que é um algoritmo classificador que utiliza o método de árvores de decisão construídas de forma agregada (LORENZETT, Cassio D. C.; TELÖCKEN, Alex V., 2016). E como técnica de classificação, foi



utilizada a Validação Cruzada que envolve a subdivisão aleatória do conjunto de treinamento em  $k$  partes de igual tamanho, em que  $k-1$  partes são utilizadas para treinar os modelos e a parte restante é destinada a avaliar o desempenho (SANTOS, Hellen G., et al., 2019), definida em 5 partes.

## Resultados e Discussão

Para o treinamento correto do algoritmo, houve a necessidade de uma coleta de imagens de queijo que apresentavam uma boa nitidez, para que fosse possível a detecção das características dos queijos considerados de boa qualidade e os que apresentavam a presença de coliformes (Figura 1). Dessa forma, as imagens foram transformadas em atributos, e com a aplicação do filtro *InterQuartileRange*, foi possível detectar os valores discrepantes dos atributos (Figura 2), e removê-los posteriormente, evitando assim, anomalias nos resultados de classificação. O algoritmo *Random Forest* em conjunto com técnica de Validação Cruzada produziu uma taxa de acerto de classificação de aproximadamente 83% (Figura 3). A taxa de sucesso na classificação de queijos com coliformes foi 100%, enquanto a taxa de acerto dos queijos normais foi de aproximadamente 66% (Figura 3), os erros encontrados podem ser resultado da diversidade de características encontradas nas imagens.

Os resultados podem variar conforme o banco de dados, já que as imagens, após a filtragem de atributos, baseiam-se na identificação de anomalias pela concentração das cores RGB e quantidade de pixels. As dificuldades enfrentadas no trabalho estiveram principalmente relacionadas à coleta de dados para mineração, dado seu impacto direto nos resultados obtidos.

## Considerações finais

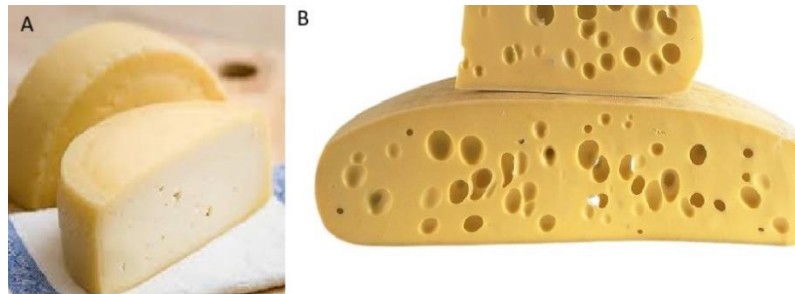
Os resultados obtidos indicam que as técnicas de mineração de dados foram eficientes na detecção de imagens de queijos que apresentavam características de presença de coliformes. A utilização de técnicas de mineração de dados pode beneficiar o processo de produção do Queijo Minas Artesanal evitando a comercialização prematura de produtos contaminados.

## Agradecimentos

Gostaria de expressar minha gratidão a todos os indivíduos e entidades que tornaram possível a realização deste estudo. Agradeço ao meu orientador pelos valiosos insights e orientações. Agradeço à instituição pelo apoio financeiro e infraestrutural que viabilizou este estudo.

## Referências

- CARDOSO, O.N.P. & MACHADO, R.T.M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista de Administração Pública [online]. 2008. Disponível em: <<https://doi.org/10.1590/S0034-76122008000300004>>. Acesso em 15 jan. 2022.
- EPAMIG. Queijo Minas Artesanal principais problemas de fabricação – Manual Técnico de Orientação ao Produtor. Belo Horizonte, 2019.
- LORENZETT, Cassio D. C.; TELÖCKEN, Alex V. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. Universidade De Cruz Alta. Cruz Alta-RS, 2016. Disponível em: <[https://web.archive.org/web/20180424102752id\\_/http://revistaeletronica.unicruz.edu.br/index.php/computacao/article/viewFile/4023/737](https://web.archive.org/web/20180424102752id_/http://revistaeletronica.unicruz.edu.br/index.php/computacao/article/viewFile/4023/737)>. Acesso em 29 ago. 2023.
- MATSUMOTO, et al. Contaminação por coliformes fecais em queijos prontos para o consumo. Revista Saúde em Foco. São Paulo, 2016. Disponível em: <[https://portal.unisepe.com.br/unifia/wpcontent/uploads/sites/10001/2018/06/002\\_contaminacao\\_coliformes.pdf](https://portal.unisepe.com.br/unifia/wpcontent/uploads/sites/10001/2018/06/002_contaminacao_coliformes.pdf)>. Acesso em 30 mar. 2023.
- SANTOS, Hellen G., et al. Machine learning para análises preditivas em saúde: exemplo de aplicação para prever o óbito em idosos de São Paulo, Brasil. Cadernos de Saúde Pública. São Paulo-SP, 2019. Disponível em: <<https://doi.org/10.1590/0102-311X00050818>>. Acesso em 29 ago. 2023.



**Figura 1.** A: Queijo em boas condições; B: Queijo com a presença de coliformes. Arquivo pessoal (2021).

```
@attribute 'MPEG-7 Color Layout25' numeric
@attribute 'MPEG-7 Color Layout26' numeric
@attribute 'MPEG-7 Color Layout27' numeric
@attribute 'MPEG-7 Color Layout28' numeric
@attribute 'MPEG-7 color Layout29' numeric
@attribute 'MPEG-7 Color Layout30' numeric
@attribute 'MPEG-7 Color Layout31' numeric
@attribute 'MPEG-7 Color Layout32' numeric
@attribute class {COLIFORME,BOM}
@attribute Outlier {no,yes}
@attribute ExtremeValue {no,yes}

@data
46,12,16,15,13,18,13,11,15,17,15,16,14,15,16,14,13,15,15,15,16,9,15,17,15,20,24,47,16,14,16,13,9,COLIFORME,yes,no
52,9,16,18,16,21,14,19,17,13,16,16,16,15,16,15,13,15,15,15,15,8,16,22,19,15,25,48,16,13,13,15,10,COLIFORME,no,no
38,19,20,8,13,18,18,12,20,17,10,12,14,18,6,16,13,15,17,19,17,26,15,17,17,16,16,36,15,14,15,16,15,COLIFORME,no,no
26,20,17,16,14,15,17,7,14,15,15,20,21,22,9,13,16,14,19,13,15,20,20,16,16,11,18,41,10,14,16,20,14,COLIFORME,yes,no
38,19,20,8,13,18,18,12,20,17,10,12,14,18,6,16,13,15,17,19,17,26,15,17,17,16,16,36,15,14,15,16,15,COLIFORME,no,no
26,20,17,16,14,15,17,7,14,15,15,20,21,22,9,13,16,14,19,13,15,20,20,16,16,11,18,41,10,14,16,20,14,COLIFORME,yes,no
51,12,27,14,17,15,16,15,13,17,16,21,18,15,15,14,15,15,15,15,16,11,12,24,23,16,22,47,17,7,14,16,14,BOM,no,no
52,7,17,14,19,10,15,17,19,18,15,16,17,16,16,16,15,16,17,17,14,12,10,21,14,18,44,17,20,13,16,14,BOM,no,no
58,14,15,22,16,20,16,15,17,16,18,15,13,15,18,15,16,14,16,15,16,23,14,14,21,15,19,36,17,17,12,16,13,BOM,no,no
41,12,12,7,20,14,15,18,16,11,16,16,12,14,13,18,17,17,17,14,14,14,20,19,17,18,41,17,14,16,15,15,BOM,no,no
28,25,16,10,19,6,19,14,15,18,13,17,15,16,10,19,17,15,15,13,15,28,18,15,17,16,22,34,14,16,14,16,12,BOM,no,no
19,13,14,11,19,3,18,15,19,16,12,13,16,16,14,16,18,14,17,17,17,29,12,15,15,19,34,18,16,15,16,14,BOM,no,no
10,16,7,16,11,15,16,18,17,15,17,20,15,15,14,14,13,15,15,15,21,15,21,15,18,47,12,16,12,16,14,BOM,no,no
12,3,24,10,18,14,16,14,16,15,12,14,17,16,16,16,16,17,15,16,17,18,24,13,20,16,17,45,9,18,12,16,14,BOM,no,no
25,8,24,9,17,11,10,15,16,16,15,17,16,16,15,15,19,17,12,15,17,28,14,14,22,17,18,34,16,17,11,15,14,BOM,no,no
```

**Figura 2.** Resultado do filtro *InterQuartileRange*. Arquivo pessoal (2021).

```
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      10           83.3333 %
Incorrectly Classified Instances    2           16.6667 %
Kappa statistic                    0.6364
Mean absolute error                 0.38
Root mean squared error             0.4199
Relative absolute error             77.1692 %
Root relative squared error         84.5087 %
Total Number of Instances          12

=== Confusion Matrix ===

 a b  <-- classified as
 7 0 | a = COLIFORME
 2 3 | b = BOM
```

**Figura 3.** Resultados da Classificação. Arquivo pessoal (2021).