

Universidade Federal de Minas Gerais  
Departamento de Ciência da Computação

DCC057 Mineração de Dados  
**Trabalho Prático 1 — Proposta**

Alexander Thomas Mol Holmquist  
13 de janeiro de 2021

# 1 Introdução

Se não puramente por interesse, então como conhecimento prático; ao ver um cogumelo, como saber se ele é comestível? O perigo de cogumelos vem exatamente de que as variedades venenosas geralmente são similares às não-venenosas. É possível encontrar enciclopédias que os classifiquem corretamente, mas neste projeto buscamos uma maneira mais simples de distingui-los, através da mineração de padrões frequentes.

## 2 Objetivos

O objetivo do projeto é estritamente gerar um conjunto de atributos que possam ser aplicados no dia-a-dia para rápida e rudimentar análise do perigo sob ingestão, ao porventura se deparar com espécie de cogumelo não conhecida.

## 3 Motivação

Com a mineração de dados, será possível construir um conhecimento mais concreto na área. Se obtermos um conjunto de poucas características (menor ou igual a cinco) que possam determinar com precisão de teste de 50% se um dado cogumelo da base de dados é venenoso ou não, considere-se isso um sucesso. Caso não exista tal conjunto sintetizante, então o objetivo terá falhado. Contudo, no sentido de que a exploração realizada possibilitou um melhor entendimento do contexto, mesmo em tal situação será coerente afirmar que houve ganho.

## 4 Restrições

Imagina-se, considerando o pequeno tamanho da base de dados, e o fato de que só são incluídas duas famílias de fungos, que a representatividade da população não seja suficiente para aplicar os padrões observados universalmente.

## 5 Plano Inicial

O plano para o projeto, conforme os horizontes estabelecidos inicialmente, é:

1. Decisão das ferramentas a serem utilizadas
2. Tratamento básico dos dados, para os tornar mais legíveis
3. Carregamento dos dados e análise exploratória
4. Decisão do algoritmo a ser utilizado e sua implementação
5. Aplicação do algoritmo, verificação dos resultados
6. Iteração sobre esse processo, registrando a experiência e detalhes importantes. Talvez seja necessário buscar outras bases de dados
7. Produção do relatório e conclusão do projeto

## 6 Dados

A tabela de dados a ser utilizada dispõe 22 atributos referentes ao aspecto físico, e a classe (comestível ou venenoso / não recomendado), de cada cogumelo amostrado. Todos estes são cogumelos lamelados, pertencentes às famílias *Agaricus* e *Lepiota*.

A única tabela a ser utilizada está disponível publicamente[1]. Foi modificada no passado duas vezes (como está registrado na própria pasta disponível para download), mas sem perda significativa de qualidade.

É informado ainda que havia uma terceira classe indicativa de fungos não sabidamente comestíveis, e portanto não recomendados. Esta foi integrada na classe "venenoso". Portanto, por "venenoso", deve-se entender "venenoso, com alta probabilidade".

## 7 Bibliografia

[1]<https://archive.ics.uci.edu/ml/datasets/mushroom>