

# TRABALHO FINAL DATA MINING



MANOELA KOHLER

[Prof.manoela@ica.ele.puc-rio.br](mailto:Prof.manoela@ica.ele.puc-rio.br)

# TRABALHO SUGERIDO

## **Problema de classificação:**

- 27 atributos numéricos e categóricos que descrevem o estado de saúde de cavalos.
- Três classes de saída que indicam o que aconteceu com o cavalo: morreu, viveu ou se foi submetido à eutanásia. A ideia é prever se um cavalo pode sobreviver ou não baseado nas condições médicas passadas.



# TRABALHO SUGERIDO

## Sugestões:

- Análise exploratória;
- *Missing values*;
- Atributos desnecessários;
- Balanceamento;
- Transformação dos tipos dos atributos (dependendo do modelo de machine learning usado);
- Teste de diferentes modelos;



# TRABALHO SUGERIDO

- O trabalho pode ser feito individualmente ou em grupo de até 5 pessoas.
- O prazo de entrega é da 13 de janeiro de 2018.
- A base é uma sugestão. Sintam-se livres para trabalhar em qualquer outra base de interesse.
- Vocês podem utilizar R, Python, Rapidminer ou qualquer outro software ou linguagem que prefiram.



# TRABALHO SUGERIDO

- A base sugerida encontra-se disponível no classroom. Base de treino “horse.csv” e base de teste: “horseTest.csv” ;
- Ao terminar o trabalho, enviem-me um relatório (pdf, doc ou ppt) com o relato dos experimentos e resultados.
- Dúvidas: [prof.manoela@ica.ele.puc-rio.br](mailto:prof.manoela@ica.ele.puc-rio.br)

