Trabalho Final de Data Mining

Davi Oliveira Francisco

Matrícula: 231101072

# Objetivo

Este documento visa a documentação de todo processo de exploração, visualização, tratamento dos dados. Além disso, também é o registro dos testes, otimização e seleção de modelos de machine learning clássicos de classificação para tratar do problema proposto como trabalho final da matéria de Data Mining.

# Introdução

O trabalho proposto trata de um problema de classificação, no qual também devem ser desenvolvidos passos de pré-processamento como: análise de dados (EDA), tratamento de valores faltantes, atributos desnecessários, balanceamento de classes, transformações dos atributos e teste de diferentes modelos.

A base de dados proposta contém 27 atributos numéricos e categóricos sobre cavalos e, com base nessas informações, o objetivo do modelo é prever com boa precisão o destino do cavalo que são 3 possíveis: viveu, morreu ou eutanasiado.

Nesse contexto, foi desenvolvido um notebook em python para realizar todos os passos necessários do trabalho e, como adicional, foi criado também um app para realizar as previsões. O app tem como objetivo levar a capacidade de predição dos modelos desenvolvidos para o usuário final. Nele é possível prever usando diversos modelos que foram serializados a partir de uma classe criada para unificar o objeto do modelo e do encoder em um aquivo pickle único.

# Análise exploratória dos dados

O primeiro passo do trabalho envolve a análise exploratória da base de dados fornecida. Depois de importar os dados, for