

DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA - CTC - UFSC

INE5687 - Projeto em Ciência de Dados

Proposta do projeto em Ciência de Dados

**Avaliação comparativa de modelos de Machine Learning na  
classificação de comportamento canino**

Matheus Cadorin Luca

Mariana Mazzo Heitor

Yuri Rodrigues de Souza

Setembro de 2025

## Motivação e Contribuição

Devido a questões relacionadas a razões morais, econômicas e psicológicas, os tutores são motivados a entender aspectos comportamentais, psicológicos e de saúde de seus animais, com o objetivo de monitorar seu bem-estar. Por isso, uma análise mais detalhada e precisa dos padrões de movimento e posturas corporais dos cães, através de medições de atividades, beneficia tanto os tutores quanto os próprios animais ao permitir a identificação de comportamentos repetíveis e níveis de atividade e repouso, assim como a supervisão de suas alterações. Isso auxilia, por exemplo, a melhorar a compreensão sobre comportamentos e reações do cachorro, como a ansiedade de separação.<sup>1</sup>

Além disso, a capacidade de extrair comportamentos mais específicos tem o potencial de servir como um índice do bem-estar e estado de saúde do animal, ajudando a detectar hábitos relacionados ao stress ou à dor. Observa-se alguns exemplos de condições de saúde, como os estágios iniciais da artrite, responsável pela redução de movimento do animal, que são frequentemente despercebidos por tutores. Portanto, essa análise se torna valiosa para a avaliação de saúde, suscetibilidade a doenças e eficácia de medicamentos - especialmente em cães com doenças crônicas - e, no futuro, para identificar problemas comportamentais e resposta a tratamentos, assim como o gasto energético do animal.

Vale ressaltar que, atualmente, a pesquisa comportamental baseia-se muitas vezes na anotação manual de gravações de vídeo, um método que, além de demorado e trabalhoso, reduz a precisão devido à sobreposição de outros objetos no raio de visão da câmera. Dessa forma, o avanço da aplicação de algoritmos de Machine Learning, em conjunto com os sensores de movimento, oferece uma alternativa mais eficiente para a detecção automática de comportamentos, fornecendo informações adicionais e captando comportamentos sutis do cão, o que beneficia tanto os tutores de animais como a pesquisa comportamental. Em vista disso, devido ao custo acessível de sensores contendo acelerômetro e giroscópio de três eixos, a solução proposta torna-se uma alternativa barata e eficiente para os tutores monitorarem seus cachorros durante os períodos em que os animais estão sozinhos ou fora dos limites da visão humana.

O objetivo deste projeto, portanto, se trata de, a partir de dados retirados de sensores com acelerômetro e giroscópio de três eixos, avaliar a acurácia e a precisão de diversos modelos de Machine Learning na classificação de comportamentos caninos - classificados, de acordo com *“Dog behaviour classification with movement sensors placed on the harness and the collar”*, como *galopar, deitar, sentar, cheirar, ficar em pé, trotar e*

---

<sup>1</sup> “O transtorno de ansiedade de separação em cachorro envolve ansiedade intensa e persistente no animal que está longe da pessoa a quem ele tem maior apego, no caso, o tutor.”

Fonte: [Ansiedade de separação cachorro: como ajudar o peludo | Petz](#)

*caminhar* - a fim de comparar os resultados e, dessa forma, determinar o melhor algoritmo para o tratamento dessa problemática.

## Objetivos

### Objetivo Geral

Realizar uma avaliação comparativa da performance de diferentes algoritmos de Machine Learning na tarefa de classificação de comportamentos caninos, utilizando dados de sensores de movimentos acoplados aos animais. Busca-se identificar o modelo com melhor a precisão para essa finalidade específica.

### Objetivos específicos

- Identificar os padrões e as correlações mais relevantes entre os dados dos sensores por meio da análise exploratória.
- Extrair características fundamentais a partir dos dados brutos que possam aprimorar a capacidade dos modelos de distinguir entre as diferentes classes;
- Avaliar a performance de múltiplos algoritmos de classificação para estabelecer qual modelo oferece maior precisão para os dados analisados.

## Tecnologias utilizadas

A implementação deste projeto será realizada na linguagem Python, utilizando o ambiente Jupyter Notebooks, seja em um *IDE* local ou na plataforma em nuvem Google Colab. Para a etapa de engenharia e análise de dados, serão utilizadas as bibliotecas Pandas/Polars, para a manipulação estruturada dos dados, e NumPy, para operações numéricas. A visualização de dados e a geração de gráficos serão conduzidas a partir das bibliotecas Matplotlib, Seaborn e Plotly. O framework TensorFlow será a tecnologia central do projeto, responsável pela construção, compilação e treinamento dos modelos de Machine Learning. Adicionalmente, a biblioteca Scikit-learn será utilizada como suporte para as tarefas de pré-processamento, divisão dos conjuntos de dados e para a extração das métricas de avaliação de performance dos modelos. Em relação aos modelos de Machine Learning, será realizado um estudo, durante a execução do projeto, para selecionar quais serão utilizados na avaliação comparativa, não estando definido ainda.

# Origem dos dados

O projeto utilizará o conjunto de dados [Inertial data for dog behaviour classification](#), disponível publicamente na plataforma do Kaggle. Os dados derivam da pesquisa, “*Dog behaviour classification with movement sensors placed on the harness and the collar*” publicada no periódico *Applied Animal Behaviour Science*.

A coleta de dados foi conduzida em um ambiente semi-controlado na faculdade de medicina veterinária da Universidade de Helsinque (Finlândia). Participaram do estudo 45 cães, de porte médio a grande, representando 27 raças distintas.

Os cães foram equipados com dois sensores de movimento ActiGraph GT9X Link, que possuem um acelerômetro e um giroscópio de três eixos cada. Um sensor foi posicionado na altura da cernelha nas costas do animal, enquanto o outro foi fixado em sua coleira. As sessões foram gravadas em vídeo para permitir a anotação manual dos comportamentos com o auxílio de um software especializado.

O conjunto de dados é composto por duas fontes de dados e possui cerca de 10 milhões de registros. As fontes de dados são:

## Dados informacionais sobre os cães

Coluna	Tipo de dado	Descrição
DogID	Inteiro	Número de identificação do cão.
Breed	Categórico	Raça do cão.
Weight	Inteiro	Peso do cão(Kg).
Age Months	Inteiro	Idade em meses do cão.
Gender	Inteiro	Sexo do cão. 1 = Fêmea, 2 = Macho.
Neutering Status	Inteiro	Estado de castração do cão. 1 = castrado, 0 não castrado.

DogID	Breed	Weight	Age months	Gender	NeuteringStatus
16	CrossBreed	13	20	1	0

### Dados dos sensores vestidos nos cães

Coluna	Tipo de dado	Descrição
DogID	Inteiro	Número de identificação de um cão
TestNum	Inteiro	Número da execução do teste (1 ou 2)
t_sec	Flutuante	Tempo em segundos desde o início da execução do teste.
ABack	Flutuante	Medida de acelerômetro do sensor das costas para os eixos x, y e z.
ANeck	Flutuante	Medida de acelerômetro do sensor da coleira para os eixos x, y e z.
GBack	Flutuante	Medida de giroscópio do sensor das costas para os eixos x, y e z.
GNeck	Flutuante	Medida de giroscópio do sensor da coleira para os eixos x, y e z.
task	Categórico	O comando solicitado em dado momento.
Behavior	Categórico	Comportamento anotado. Podendo ser até três anotações por registro.
PointEvent	Categórico	Pequenos eventos anotados separadamente.

DogID	TestNum	t_sec
16	1	90.81

ABack_x	ABack_y	ABack_z	ANeck_x	ANeck_y	ANeck_z
-0.272949	0.204102	1.185547	-0.394043	-0.1875	-0.80127

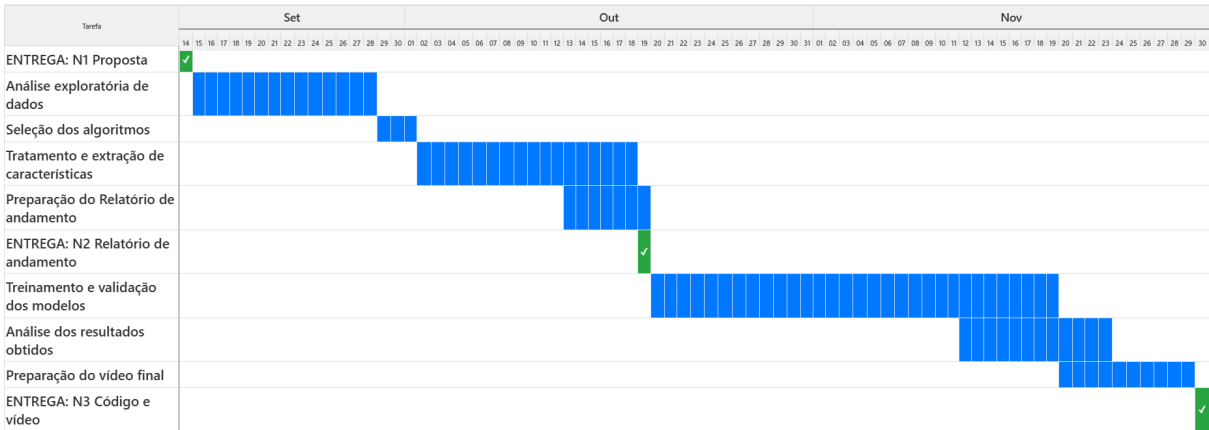
GBack_x	GBack_y	GBack_z	GNeck_x	GNeck_y	GNeck_z
-39.489749	62.500004	35.766604	-11.718751	52.185062	-75.378423

Task	Behavior_1	Behavior_2	Behavior_3	PointEvent
Task walk	Walking	<undefined>	<undefined>	<undefined>

# Eventuais riscos

- O conjunto de dados pode apresentar ruídos nos sensores ou um desbalanceamento significativo entre as classes de comportamento, ocasionando em um enviesamento nos modelos e fazendo com que eles performem mal em classes minoritárias.
- Os modelos treinados podem não generalizar bem para cães de outras raças, tamanhos ou em condições diferentes das registradas, pois o dataset representa um número limitado de cães.
- A escolha de features inadequadas pode limitar significativamente o desempenho de todos os modelos avaliados.
- O treinamento de múltiplos modelos de machine learning pode ser computacionalmente intensivo, exigindo muito tempo e recursos de hardware.

# Cronograma



ENTREGA: N1 Proposta: 14/09

Análise exploratória de dados 15/09 - 28/09

Seleção dos algoritmos 29/09 - 01/10

Tratamento e extração de características 02/10 - 18/10

Preparação do Relatório de andamento 13/10 - 19/10

---

**ENTREGA: N2 Relatório de andamento 19/10**

---

**Treinamento e validação dos modelos 20/10 - 19/11**

**Análise dos resultados obtidos 12/11 - 23/11**

**Preparação do vídeo final 20/11 - 29/11**

---

**ENTREGA: N3 Código e vídeo 30/11**

## **Referências**

- [1] KUMPULAINEN, P. et al. Dog behaviour classification with movement sensors placed on the harness and the collar. *Applied Animal Behaviour Science*, v. 241, p. 105393, ago. 2021.
- [2] AZAMJON MUMINOV; MUKHRIDDIN MUKHIDDINOV; CHO, J. Enhanced Classification of Dog Activities with Quaternion-Based Fusion Approach on High-Dimensional Raw Data from Wearable Sensors. *Sensors*, v. 22, n. 23, p. 9471–9471, 4 dez. 2022.
- [3] HUSSAIN, A. et al. Long Short-Term Memory (LSTM)-Based Dog Activity Detection Using Accelerometer and Gyroscope. *Applied Sciences*, v. 12, n. 19, p. 9427, 20 set. 2022.
- [4] KUMPULAINEN, P. et al. Dog activity classification with movement sensor placed on the collar. *Proceedings of the Fifth International Conference on Animal-Computer Interaction*, 4 dez. 2018.