

# UM ESTUDO DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAR IMAGENS DE REDES SOCIAIS E EXTRAIR INFORMAÇÕES QUE IDENTIFIQUEM ENGAJAMENTO



**Erick H. Orsi, Dr. Alexandre Noma**

Centro de Matemática, Computação e Cognição, Universidade Federal do ABC  
Av. dos Estados, 5001, Santo André, SP  
{erick.orsi@aluno.ufabc.edu.br, alexandre.noma@ufabc.edu.br}

**Resumo.** Este projeto teve como objetivo extrair informações do Facebook e utilizá-las em algoritmos de classificação supervisionada (ALPAYDIN, 2014) para prever a quantidade de curtidas que uma imagem deve receber na rede social com base em algumas informações presentes nela.

**Palavras-chave.** aprendizado de máquina, redes sociais, classificação supervisionada de imagens, engajamento.

## INTRODUÇÃO

Neste trabalho, desejou-se automatizar os processos de extração de informações de redes sociais e classificação dos dados em plataformas de aprendizado de máquina para medir o engajamento do público. Foi desenvolvido um programa para extrair as imagens do perfil público *aacdFacebook* do Facebook e o conteúdo destas imagens. Após extrair os dados (*tags*, quantidade de curtidas, etc.), construiu-se um conjunto rotulado de amostras para treinamento. Desta forma, pôde-se classificar uma nova imagem entre duas classes: "muitas" ou "poucas" curtidas.

## Metodologia

Foram extraídas imagens do Facebook por meio da ferramenta *Facebook Graph API*, assim como os componentes visuais (*tags*) de cada imagem usando a ferramenta *Computer Vision API* da Microsoft Azure. Em seguida, foi realizada uma análise exploratória dos dados (MORETTIN; BUSSAB, 2010) para que os dados analisados fossem finalmente aplicados em algoritmos de classificação.

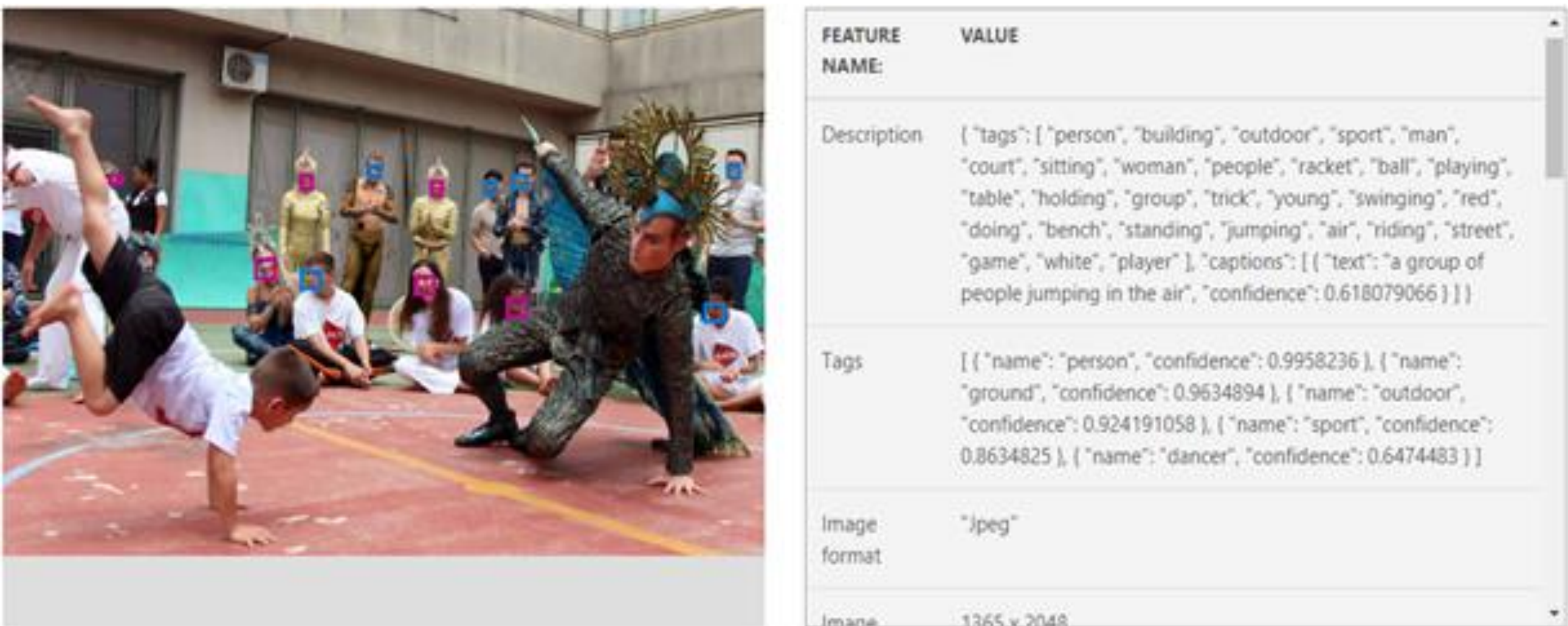


FIGURA 1: Exemplo de extração de *tags*.

FONTE: [3].

Um vetor de características foi composto pelas *tags*, representadas numericamente por 0 (ausência) ou 1 (presença) em cada imagem. A variável resposta foi a quantidade de curtidas binarizada, sendo 0 a resposta para "poucas" e 1 para "muitas".

## Conclusão

Em geral, as classificações apresentaram resultados promissores, sendo a maior porcentagem de acerto por volta de 70%. Os resultados reforçam que a presunção de independência entre as *tags* gera uma perda de informação de contexto da imagem. A presença de uma *tag* sozinha pode ser insuficiente para gerar uma grande quantidade de curtidas, mas uma combinação delas pode motivar e sensibilizar mais uma pessoa devido ao contexto ou à situação ilustrada pela imagem.

## ACERTO DOS MODELOS

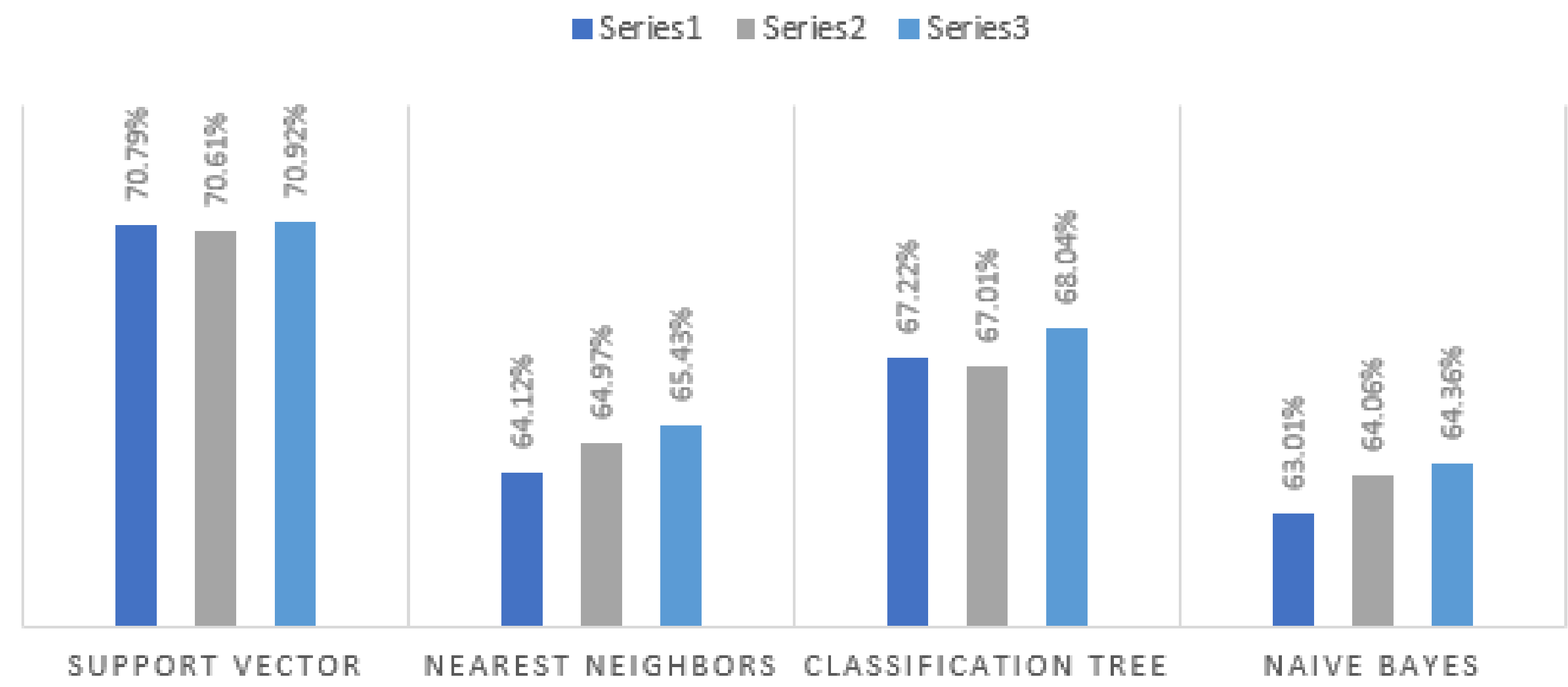


FIGURA 2: Porcentagem de acerto dos modelos com base em amostras de treinamento e teste, selecionadas aleatoriamente.

## REFERÊNCIAS

[1] ALPAYDIN, E. *Introduction to Machine Learning*. 3 ed. London: Cambridge. 2014.

[2] MORETTIN, P. A.; BUSSAB, W. *Estatística Básica*. 6 ed. São Paulo: Saraiva. 2010.

[3] <<https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>>

## AGRADECIMENTOS

Este trabalho foi financiado pelo CNPq.