

# Trabalho Pedagógico Complementar

## *Machine Learning* Aplicado 2

Elloá B. Guedes<sup>1</sup>

<sup>1</sup>Laboratório de Sistemas Inteligentes  
Escola Superior de Tecnologia  
Universidade do Estado do Amazonas  
Av. Darcy Vargas, 1200 – Manaus – Amazonas

ebgcosta@uea.edu.br

### 1. Instruções Gerais

O trabalho pedagógico complementar para a disciplina *Machine Learning* Aplicado 2 tem como objetivo o desenvolvimento de competências relacionadas à ementa da disciplina as quais não foram alcançadas pelos discentes em cópia no tempo hábil da ministração do módulo.

O foco deste trabalho consiste na proposição, implementação e avaliação de Máquinas de Vetores de Suporte (SVMs, do inglês *Support Vector Machines*) em uma tarefa de classificação multiclasse a partir de características extraídas automaticamente de imagens com Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*). Muito embora CNNs possam abranger o *pipeline* completo de tarefas de classificação em Visão Computacional, uma abordagem alternativa e também efetiva consiste em utilizar CNNs pré-treinadas em grandes bases de dado como extratoras automáticas de características visuais as quais provêm entrada para outros tipos de classificadores, que discriminam então as classes do problema considerado. Dentre esses classificadores, destacam-se as SVMs. Essa abordagem alternativa, contemplada nas técnicas de *transfer learning*, mostra-se especialmente efetiva quando há poucos exemplos na tarefa alvo considerada [Sarkar et al. 2018].

O contexto da atividade será a tarefa de classificação de sementes de cacau no teste de corte a partir de uma base de dados proposta na literatura, a qual contempla 1.400 imagens [Santos et al. 2019]. A partir das contribuições de um trabalho anterior na literatura [Malcher and Guedes 2022], devem avaliar a utilização das mesmas arquiteturas de CNNs nele listadas como base para a transferência de aprendizado com a utilização de SVMs classificadoras. Nesse último ponto, devem propor uma busca em grade com diferentes configurações de parâmetros e hiperparâmetros, justificando a escolha dos valores. Buscar-se-á superar o *baseline* de 89,79 % na mesma tarefa, repetindo-se os protocolos experimentais do trabalho relacionado.

As seguintes atividades devem ser contempladas no trabalho pedagógico:

1. **Análise exploratória da base de dados.** Deve compreender a distribuição de exemplos por classe, a inspeção de qualidade e formato das imagens e a visualização de imagens por classe;
2. **Preparação de Dados.** Deve contemplar o redimensionamento das imagens para fins de padronização, a normalização dos valores dos pixels e a elaboração de

partição *holdout* que será utilizada em todos os experimentos por todos os integrantes;

3. **Extração de Características Visuais.** Todas as arquiteturas mencionadas no trabalho correlato devem ser utilizadas sem as camadas finais de classificação, com pesos fixos, para produzir vetores de características visuais de todas as imagens na base de dados;
4. **Elaboração da Busca em Grade.** Conceber uma grade de busca com SVMs especialmente focando em múltiplos valores de parâmetros de regularização e de funções de kernel;
5. **Execução dos Experimentos.** Executar os experimentos de avaliação dos *pipelines* propostos, coletando métricas como tempo de treinamento e desempenho no conjunto de testes, tais como acurácia, precisão, revocação e  $F_1$ -Score;
6. **Análise dos Resultados.** Compreendendo uma tabela descritiva dos resultados, matrizes de confusão dos três melhores modelos e definição da solução de referência identificada perante o problema proposto.

Serão avaliados os seguintes artefatos:

1. **Repositório no GitHub.** De caráter público, incluindo todos os integrantes da equipe e contendo as evidências de código e de progresso, registrando o trabalho dos múltiplos integrantes;
2. **Código-Fonte Python.** Deve ser incluído no repositório sob a forma de um único ou múltiplos Jupyter Notebooks, contendo todas as atividades de programação consideradas para atender ao que se pede no projeto. As células devem demonstrar o resultado da execução do código.
3. **Relatório.** Seguindo o *template* da Sociedade Brasileira de Computação e possuindo não menos que 5 páginas, deve contemplar as seções de introdução, metodologia, resultados e discussão e considerações finais. Deve incluir figuras, gráficos e referências.

Recomenda-se fortemente a adoção de estratégias de integração contínua, preferencialmente utilizando plataformas como o GitHub Actions ou ferramentas similares, para gerenciar o *pipeline* de desenvolvimento. O foco no versionamento e colaboração via GitHub é crucial para evitar o isolamento do desenvolvimento em ambientes efêmeros, como o Google Colab. A integração contínua e o uso do Git atestam a participação contínua e transversal de todos os membros da equipe, assegurando que as contribuições sejam rastreáveis e verificáveis. Esta metodologia não apenas garante uma distribuição mais justa e transparente de atividades, permitindo uma supervisão clara dos *commits* e *merges*, mas também promove o aprendizado e a capacitação coletiva ao exigir que todos os envolvidos compreendam e sigam o fluxo de trabalho colaborativo padrão da Engenharia de Software e MLOps.

Práticas para aferir a integridade acadêmica nos artefatos entregues pelas equipes serão utilizadas na avaliação do trabalho de forma complementar à verificação dos critérios elencados.

## 2. Datas Importantes

1. 18/12 – Definição das equipes, indicação do Github de cada equipe com todos os integrantes;

2. 19/12 – Disponibilidade para reunião com os discentes para esclarecimento de dúvidas do trabalho. Sob demanda, a ser solicitado por e-mail;
3. 21/01 – Entrega do repositório do Github e do código fonte Python;
4. 26/01 – Entrega do relatório;
5. 29/01 – Atribuição das notas.

### **3. Considerações de Acompanhamento**

As datas elencadas permitem que os estudantes desenvolvam as atividades durante o recesso dos módulos, de tal modo a evitar intersecções com novos módulos com pendências pedagógicas. Porém, considerando o Calendário Institucional da UEA, a docente responsável pelo trabalho pedagógico complementar estará em período de férias, o que exigirá o contato das equipes por e-mail para eliciação de dúvidas e outros aspectos de acompanhamento. As mensagens serão respondidas em até 48 horas. O contato por aplicativo de mensagens não será permitido, pois não se trata de via de comunicação institucional formal.

### **Referências**

- Malcher, D. and Guedes, E. (2022). Classificação inteligente do teste de corte do cacau com redes neurais convolucionais profundas. In *Anais do XIII Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*, pages 31–40, Porto Alegre, RS, Brasil. SBC.
- Santos, F., Palmeira, E., and Jesus, G. (2019). An image dataset of cut-test-classified cocoa beans. *Data in Brief*, 24:103916.
- Sarkar, D., Bali, R., and Ghosh, T. (2018). *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing, Birmingham, 1st edition.