

Disciplina: Aprendizagem de Máquina
Período: 2024.1
Professor: César Lincoln Cavalcante Mattos

Projeto Final

Trabalhando em **grupos de 3 ou 4 pessoas** (não misturar alunos da graduação e da pós-graduação), escolha um problema e desenvolva um projeto de aprendizagem de máquina, documentando sua execução em um **artigo** com o seguinte formato:

1. **Introdução:** Resumo do projeto, contextualização do problema, apresentação dos objetivos e motivação para a investigação feita (classificação, regressão, agrupamento, etc).
2. **Fundamentação teórica:** Breve descrição técnica do problema escolhido, descrição dos dados (atributos, saídas, etc) e revisão de trabalhos similares.
3. **Metodologia:** Detalhamento da metodologia de aplicação do(s) modelo(s) de aprendizagem e de análise dos resultados, incluindo as métricas utilizadas na avaliação dos modelos.
4. **Experimentos:** Descrição dos experimentos computacionais realizados, apresentação de tabelas/figuras e discussão dos resultados obtidos.
5. **Conclusão:** Comentários finais sobre o trabalho e propostas de investigações futuras.

Alunos da pós-graduação: Ao menos 1 modelo/algoritmo não apresentado na disciplina deve ser usado no projeto.

O projeto possui os seguintes marcos relevantes (as datas serão fixadas posteriormente):

- **Data 1:** Discussão sobre as ideias para os projetos finais.
- **Datas 2:** Finalização dos projetos no laboratório.
- **Datas 3:** Entrega dos trabalhos pelo SIGAA ou por e-mail (cesarlincoln@dc.ufc.br).
- **Datas 4:** Apresentação dos artigos.

Observações:

- Os trabalhos serão apresentados para a turma (~25 minutos) em ordem a definir em sorteio.
- Tirem dúvidas sobre o projeto com antecedência ao longo das aulas ou por e-mail.

- A equipe deve indicar como cada participante contribuiu para o projeto.
- O projeto deverá usar dados reais (não sintéticos).
- Sejam criteriosos na metodologia de preparação dos dados e treinamento/avaliação dos modelos.
- Use mais de um modelo de aprendizagem de máquina para o problema escolhido, permitindo comparações de desempenho.
- As implementações podem ser feitas em qualquer linguagem de programação e deverão ser entregues junto com o trabalho. A equipe poderá usar implementações já existentes.
- Recomenda-se o uso de \LaTeX . Você pode seguir o modelo de um dos principais congressos da área: Neurips, ICLR, ICML, AAAI, AISTATS, UAI, etc.

Sugestões de fontes de dados reais:

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- OpenML: <https://www.openml.org/search?type=data>
- Kaggle: <https://www.kaggle.com/datasets>
- Google Dataset Search: <https://toolbox.google.com/datasetsearch>
- Carnegie Mellon University Machine Learning and AI Datasets: <https://guides.library.cmu.edu/machine-learning/datasets>
- Portal Brasileiro de Dados Abertos: <https://dados.gov.br/>
- Dados do Íris Lab: <https://irislab.ce.gov.br/>
- U.S. Government's Open Data: <https://www.data.gov/>
- EU Open Data Portal: <https://data.europa.eu/euodp/en/home>
- UK Data Service: <https://www.ukdataservice.ac.uk/>
- Awesome Public Datasets: <https://github.com/awesomedata/awesome-public-datasets>