

Aprendizado de Máquina Interpretável: Estado da Arte e Avaliações Empíricas

Código do projeto: ML01

Responsável: Prof. Marcelo de Souza Laretto

Linha de pesquisa: Inteligência de Sistemas

Número de vagas: 1

Descrição geral

A demanda por métodos de machine learning (ML) interpretáveis está presente em vários contextos, como diagnósticos médicos, análise de crédito, mineração de processos etc. Todavia, ainda existe uma dicotomia entre interpretabilidade e desempenho, uma vez que modelos de aprendizado de máquina mais complexos e sofisticados tendem a obter resultados mais robustos em detrimento da interpretabilidade. Diante desses desafios, há um interesse crescente no desenvolvimento de métodos voltados a obter interpretações/explicações para as predições fornecidas pelos algoritmos de ML.

Este projeto de pesquisa terá por objetivos principais: (i) realizar um levantamento do estado da arte nesta na área de Aprendizado de Máquina Interpretável (iML), através de uma revisão sistemática; (ii) realizar avaliações empíricas de desempenho de alguns métodos de iML, com ênfase em indicadores de desempenho e de interpretabilidade. Dependendo da evolução dos estudos nesta área, pretendemos investigar potenciais extensões a alguns métodos existentes de iML.

Perfil desejado

Sólidos conhecimentos em programação (obrigatório); conhecimentos avançados em estatística e em aprendizado de máquina (desejável).

Referências

Robeer, M.J. Contrastive Explanation for Machine Learning. Master Thesis Business Informatics, Utrecht University, 2018. Disponível em: https://studenttheses.uu.nl/bitstream/handle/20.500.12932/30669/MJRobeer_Thesis-Public.pdf