

Título: Ativando λ-justiça: Não-discriminação algorítmica em Árvores de Decisão

Data: 22/04/2022

Horário: 14h00

Local: Videoconferência

Resumo:

Com o avanço tecnológico, várias entidades optam por usar modelos computacionais para classificar indivíduos, com o objetivo de negar ou conceder um benefício. Por exemplo, em concessões de empréstimo bancário, onde o banco classifica se uma pessoa é apta a receber um empréstimo ou não. Apesar de aplicações de Inteligência Artificial (IA) serem úteis para tomadas de decisão, elas não são livres de discriminação. Quando um algoritmo é treinado com dados historicamente discriminatórios, ou a base de dados é desbalanceada no que diz respeito às características minoritárias, o modelo tende a propagar a discriminação presente nos dados de treinamento. Para classificar indivíduos semelhantes de forma semelhante, ou seja, rotular analogamente as pessoas com habilidades e características similares para a realização de uma tarefa, são necessárias restrições de justiça, que por sua vez, podem alterar as classificações, prejudicando a acurácia. Neste trabalho, são definidas uma métrica para calcular o quão justo é um modelo, e duas propriedades para mitigar o problema gerado pela propagação de discriminação de indivíduos ao mesmo tempo que lidam com o trade-off entre utilidade e justiça. As propriedades são ativadas na etapa de pós-processamento. Além disso, propomos a ativação das propriedades definidas para o modelo de Árvore de Decisão.

Banca examinadora:

- Prof. Dr. Javam de Castro Machado (MDCC/UFC) - Orientador
- Prof. Dr. César Lincoln Cavalcante Mattos (MDCC/UFC)
- Prof. Dr. Victor Aguiar Evangelista de Farias (UFC)
- Prof. Dr. Sérgio Lifschitz (PUC-Rio)