Machine Learning Methods for Precision Medicine Research Designed to Reduce Health Disparities: A Structured Tutorial

Basu, S., Faghmous, J. H., & Doupe, P. (2020). Machine learning methods for precision medicine research designed to reduce health disparities: A structured tutorial. *Ethnicity and Disease*, *30*, 217–228. https://doi.org/10.18865/ed.30.S1.217

(Basu, Faghmous, & Doupe, 2020)

Resumo:

A investigação de medicina de precisão concebida para reduzir as disparidades na saúde envolve frequentemente o estudo de conjuntos de dados a vários níveis para compreender como as doenças se manifestam desproporcionadamente num grupo em detrimento de outro, e como os escassos recursos de cuidados de saúde podem ser dirigidos precisamente para aqueles que correm maior risco de doença. Neste artigo, fornecemos um tutorial estruturado para investigadores médicos e de saúde pública sobre a aplicação de métodos de aprendizagem de máquinas para conduzir investigação em medicina de precisão concebida para reduzir as disparidades de saúde. Revemos termos e conceitos-chave para a compreensão de documentos de aprendizagem de máquinas, incluindo aprendizagem supervisionada e não supervisionada, regularização, validação cruzada, ensacamento, e reforço. As métricas são revistas para avaliar os aprendentes de máquinas e as principais famílias de abordagens de aprendizagem, incluindo a aprendizagem baseada em árvores, aprendizagem profunda, e aprendizagem em conjunto. Destacamos as vantagens e desvantagens de diferentes abordagens de aprendizagem, descrevemos estratégias para interpretar modelos de "caixa negra", e demonstramos a aplicação de métodos comuns num exemplo de conjunto de dados com estatísticas de fonte aberta código em R. Ethn Dis. 2020;30(Sup.

Introdução :

Investigação em medicina de precisão concebida para reduzir as disparidades na saúde envolve frequentemente o estudo de vários níveis conjuntos de dados para compreender como as doenças manifestar-se de forma desproporcionada num só grupo sobre outro, e quão escasso os recursos dos cuidados de saúde podem ser dirigidos precisamente para aqueles que correm maior risco de doença.1 Aplicação apropriada de os métodos de aprendizagem de máquinas podem ajudar assegurar que maximizamos a clínica utilidade das ferramentas que desenvolvemos, tomando em conta a de cada paciente individual biologia, estilo de vida, e ambiente.

Neste artigo, fornecemos um tutorial estruturado para médicos e público em geral. investigadores da saúde sobre a aplicação da máquina métodos de aprendizagem na investigação em medicina de precisão concebidos para reduzir a saúde disparidades. A aprendizagem mecânica refere-se a utilizando abordagens algorítmicas - comummente referidas como aprendentes, preditivas modelos, ou estimadores - para categorizar dados ou prever um resultado. O termo aprendizagem de máquinas engloba uma vasta conjunto de métodos com um objectivo comum: ligar as entradas a uma saída com precisão aperfeiçoando repetidamente as regras que regem como os dados de entrada se relacionam com a saída resultado.

Conclusões:

A terminar, revemos alguns dos princípios considerados mais críticos para os documentos de aprendizagem de máquinas de alta qualidade.25-27

Primeiro, um padrão importante para reprodutibilidade e extensão na literatura de aprendizagem de máquinas é a partilha de código estatístico e subjacente dados. A desidentificação dos dados e a partilha do código estatístico bruto é particularmente importante dado o problema que muitos trabalhos de investigadores têm foi considerado não reprodutível.28

Em segundo lugar, é importante para a máquina problemas de aprendizagem a serem pré-especificados, para que os investigadores não se sintam tentados a utilizar a abordagem apenas para produzir (potencialmente falso-positivo) associações.

Terceiro, o utilizador final de uma máquina O aprendente deve ser tido em conta. Os diferentes públicos precisam de ser capaz de interpretar um aprendiz, ou apenas utilizar o aprendente através da introdução de dados e que o aprendente forneça resultados automaticamente (por exemplo, no back-end de um registo médico electrónico).

Finalmente, é fundamental que os investigadores que utilizam métodos de aprendizagem automática ter empatia com os dados, ou a perspectiva que a qualidade e o tipo de dados devem correspondem bem ao tipo de pergunta que está a ser feita e à utilização futura do método. Se um determinado pergunta não pode ser bem respondida com um pequeno conjunto de dados, é improvável que a pergunta seja melhor respondida com um conjunto de dados maior do mesmo tipo e qualidade dos dados.4

Por exemplo, a utilização abundante de dados de reclamações de seguros ou de registos médicos electrónicos, que são grandes e amplamente disponíveis na literatura médica, é problemática para estudos clínicos, uma vez que os modelos de previsão não irão realmente prever a presença de doença, mas sim de códigos de facturação de diagnóstico que podem estar mal correlacionados com a doença real (e sofrer de enviesamentos de selecção e erros de classificação errados). À medida que os métodos de aprendizagem de máquinas evoluem, o respeito pelos princípios-chave das boas práticas de aprendizagem de máquinas servirá para ajudar a melhorar a utilidade, fiabilidade e impacto da aprendizagem de máquinas.