

# LỰA CHỌN ĐẶC TRƯNG CHO MÔ HÌNH TUYẾN TÍNH SỬ DỤNG CRITERION RELAXATION

Lê Hoàng Sinh - 1901028

# Tóm tắt

- Lớp: CS2205.APR2023
- Link Github: <https://github.com/sinh14/CS2205.APR2023>
- Link YouTube video: <https://youtu.be/k-WChA-UYhE>
- Ảnh + Họ và Tên:



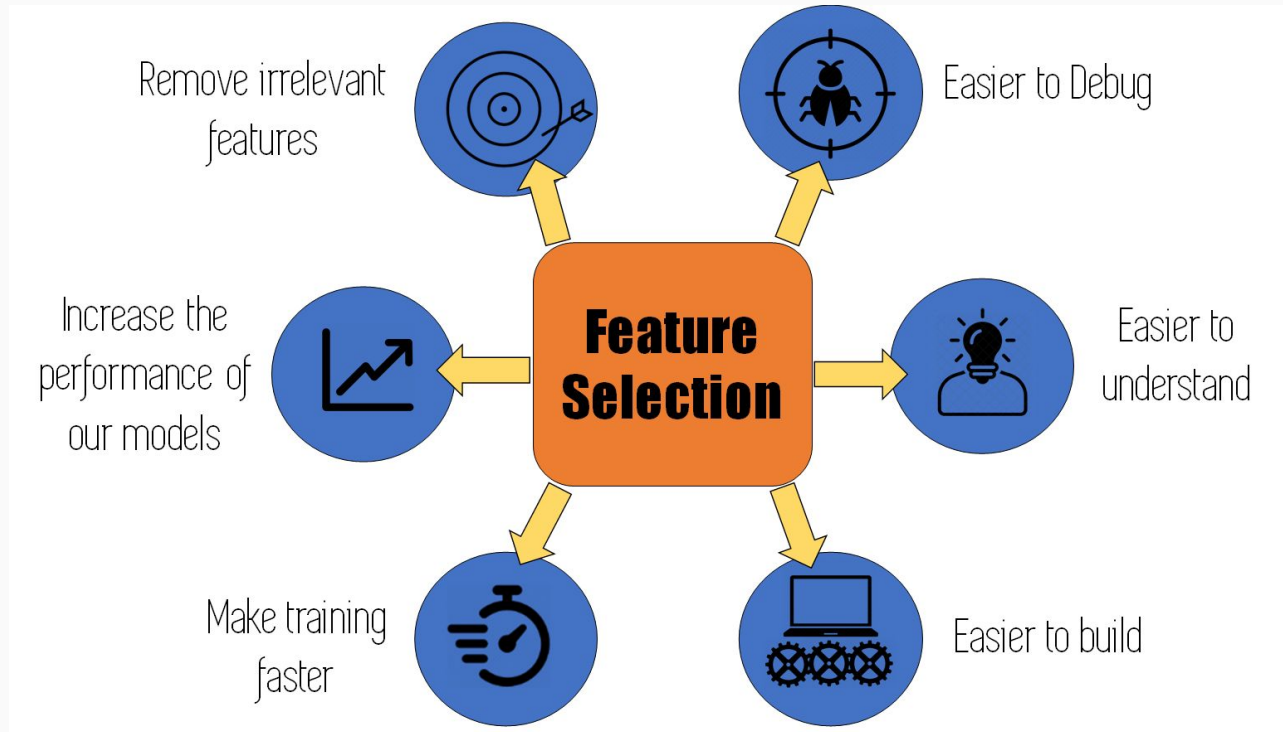
Lê Hoàng Sinh

# Giới thiệu

- Lựa chọn đặc trưng là quá trình chọn ra một tập hợp con của các đặc trưng từ tập hợp ban đầu của các đặc trưng có sẵn trong dữ liệu.
- Có nhiều phương pháp lựa chọn đặc trưng khác nhau, bao gồm tiêu chí thống kê, thuật toán học máy, kỹ thuật đặc trưng nhúng và kết hợp các phương pháp khác nhau để tạo ra tập hợp cuối cùng của các đặc trưng quan trọng.
- Sử dụng forward selection, backward selection và stepwise selection vẫn được sử dụng rộng rãi cho các bài toán hồi quy tuyến tính. Tuy nhiên vẫn còn nhược điểm đối với dữ liệu lớn

=> Cải thiện quá trình chọn lựa đặc trưng dựa trên forward, backward, stepwise selection bằng cách sử dụng kỹ thuật criterion relaxation

# Giới thiệu



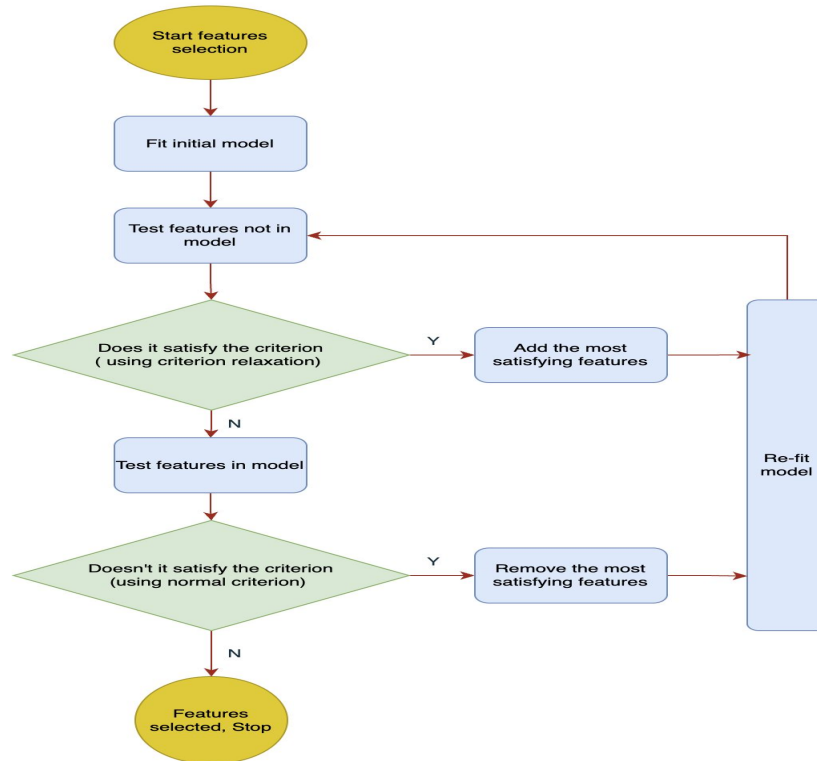
# Mục tiêu

- Cải thiện tốc độ của việc lựa chọn đặc trưng bằng cách đề xuất phương pháp lựa chọn đặc trưng mới
- Đảm bảo hiệu suất tốt của mô hình khi sử dụng tập dữ liệu đầu ra.

# Nội dung và Phương pháp

- Thu thập các bộ dữ liệu từ các nguồn khác nhau.
- Áp dụng các phương pháp Feature Selection để chọn ra tập con các đặc trưng quan trọng.
- Xây dựng các mô hình Machine Learning trên các tập các đặc trưng được chọn.
- Phân tích một số hạn chế của các phương pháp lựa chọn đặc trưng Forward, Backward, Forward dropping.
- Thực hiện lựa chọn đặc trưng sử dụng kỹ thuật Criterion relaxation.
- So sánh các phương pháp dựa trên thực nghiệm để đưa ra nhận xét kết luận.

# Nội dung và Phương pháp



# Kết quả dự kiến

- Hiểu sâu về các kỹ thuật Feature Selection và các phương pháp được sử dụng trong Machine Learning.
- Áp dụng Criterion relaxation đem lại kết quả tốt hơn về mặt thời gian trong việc lựa chọn đặc trưng và vẫn duy trì hiệu suất tốt.



# Tài liệu tham khảo

- [1] Stephen A Billings and Hua-Liang Wei. Sparse model identification using a forward orthogonal regression algorithm aided by mutual information. *IEEE Transactions on Neural Networks*, 18(1):306–310, 2007.
- [2] Christophe Cuvreur and Yoram Bresler. On the optimality of the backward greedy algorithm for the subset selection problem. *SIAM Journal on Matrix Analysis and Applications*, 21(3):797–808, 2000.
- [3] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse over-complete representations in the presence of noise. *IEEE Transactions on information theory*, 52(1):6–18, 2005.
- [4] A. Aravkin, J. Burke, A. Sholokhov, and P. Zheng, ‘Analysis of Relaxation Methods for Feature Selection in Mixed Effects Models’, *arXiv [stat.ME]*. 2022.
- [5] P. Saha, S. Patikar, and S. Neogy, ‘A Correlation - Sequential Forward Selection Based Feature Selection Method for Healthcare Data Analysis’, in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, 2020, pp. 69–72.