**Predicting Algae Blooms**

Nhóm 11

1. Người hỏi: Pham Dinh Nhien

*mình k hiểu chổ bạn chọn ntree vs mtry bạn cho chọn nhiều ntree nhưng mtry bạn chon là 2*

Trả lời:

Nhóm mình dùng 5 biến thể RandomForest với ntree thay đổi(150,250,550,600,750) và mtry được giữ nguyên.

Bởi vì, giá trị của mtry từ 1-3 là khoảng giá trị tốt, mà 2 là tốt hơn cả. Nếu chọn các giá trị mtry >= 4 sẽ cho NMSE >1 cho nhiều dự đoán ở các loại tảo, nghĩa là hiệu quả rất tệ cho dự đoán.

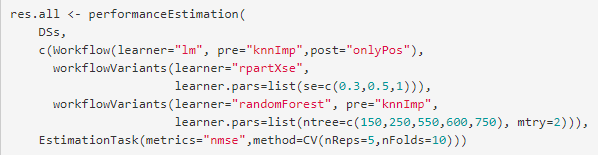
2. Người hỏi: Trần Tiến Đức

*nãy mình thấy các bạn áp dụng workflow vào mô hình dự đoán thì cho mình hỏi là workflow hoạt động theo cơ chế như thế nào để có thể tự động hóa việc tạo mô hình dự đoán?*

Trả lời:

Ta gọi một workflow là một giải pháp của một nhiệm vụ dự đoán. Điều này thường sẽ bao gồm lệnh gọi đến một số thuật toán học tập (ví dụ: cây hồi quy, hồi quy tuyến tính đa biến, rừng ngẫu nhiên) nhưng nó cũng có thể bao gồm một số bước xử lý trước dữ liệu hoặc bất kỳ bước nào khác mà người dùng cho là hữu ích để giải quyết công việc.

Ví dụ:



* Workflow() của đối tượng “lm”( hồi quy tuyến tính đa biến) gọi đến mô hình học tập hồi quy tuyến tính đa biến, với knnImp () để điền vào các giá trị chưa biết của dữ liệu bằng cách sử dụng k- phương pháp tiếp cận láng giềng gần nhất và xử lý hậu kỳ các dự đoán của mô hình tuyến tính với hàm onlyPos () về cơ bản cắt bớt bất kỳ giá trị dự đoán âm nào về 0.
* WorkflowVariants() là biển thể của Workflow() dùng để tạo điều kiện thuận lợi cho việc tạo các biến thể khác nhau của quy trình làm việc. Chẳng hạn, với đối tượng rừng ngẫu nhiên, ta đang sử dụng 5 phiên bản.