Dữ liệu:

* Em ưu tiên dùng dữ liệu 1 lần / ngày, do dữ liệu đồng bộ hơn, đầy đủ cho mỗi điểm dữ liệu, và dễ xử lý dữ liệu hơn.

Về mô hình:

* Em phát triển trước mô hình dự đoán Độ kiềm dựa trên các input đầu vào. (Chưa phải là mô hình dự báo ngày tiếp theo)

Xử lý dữ liệu, từ dữ liệu thô ban đầu, sẽ đi qua các bước xử lý sau:

* Làm sach dữ liệu, xử lý các dữ liệu khuyết, dữ liệu lỗi, xử lý các ký tự tiếng Việt(đổi tên)
* Lược bỏ các dữ liệu sẽ không dùng trong mô hình
* Gom nhóm dữ liệu theo từ ao vụ.
* Mã hóa dữ liệu dạng phân loại (non-numeric), dùng mã hóa dạng OnehotEncoder, ví dụ từ Season chứa các giá trị là ‘Dry’ và ‘Rainy’ sẽ được mã hóa thành 2 cột ‘Season\_Dry’ và cột ‘Season\_rainy’ chứa giá trị 0 và 1
* Sau khi gom nhóm theo từng vụ ao, đối với mỗi ao riêng biệt, kết hợp các thông số vật lý của ngày hiện tại và độ kiềm của ngày hôm trước làm input để dự đoán độ kiềm của ngày hiện tại.
* Tách dữ liệu ngẫu nhiên thành 2 phần, phần train và phần test.
* Scale dữ liệu về scale chuẩn (chuẩn hóa dữ liệu)

Xử lý dữ liệu:

Thông tin:

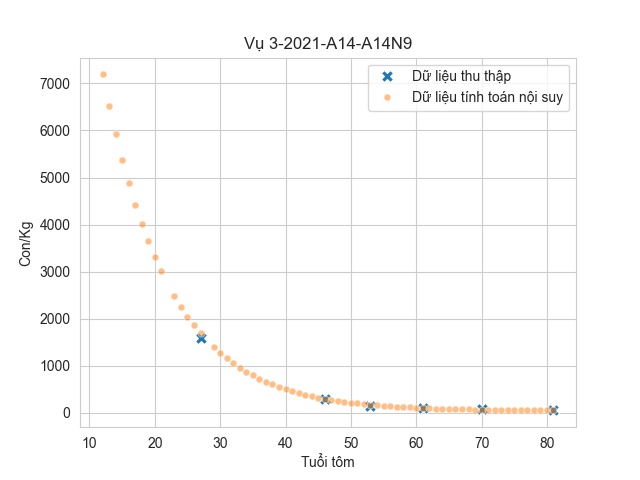
|  |  |
| --- | --- |
| Column | Non-Null Count |
| ------ | -------------- |
| Date | 1160 non-null |
| Season | 1160 non-null |
| Vụ nuôi | 1160 non-null |
| module\_name | 1160 non-null |
| ao | 1160 non-null |
| Ngày thả | 1160 non-null |
| Time | 1160 non-null |
| Nhiệt độ | 1160 non-null |
| pH | 1160 non-null |
| Độ mặn | 1160 non-null |
| TDS | 1160 non-null |
| Độ đục | 1160 non-null |
| DO | 1160 non-null |
| Độ màu | 1160 non-null |
| Độ trong | 1160 non-null |
| Độ kiềm | 1160 non-null |
| Độ cứng | 1160 non-null |
| Amoni | 1160 non-null |
| Nitrat | 1160 non-null |
| Nitrit | 1160 non-null |
| Phosphate (PO43-) | 1160 non-null |
| Silica | 1160 non-null |
| Canxi | 1160 non-null |
| Kali | 1160 non-null |
| Magie | 1160 non-null |
| Loại ao | 1160 non-null |
| Công nghệ nuôi | 1160 non-null |
| area | 1160 non-null |
| Giống tôm | 1160 non-null |
| Tuổi tôm | 1160 non-null |
| Con/Kg | 107 non-null |
| Mực nước | 1160 non-null |
| Thay nước (cm) | 1160 non-null |
| WQI | 1160 non-null |

Gom nhóm theo thông tin từ “Vụ nuôi”, “Module Name” và “Ao” thành 18 “unit”:

|  |
| --- |
| UNIT |
| Vụ 1-2022-A1-D3N6 |
| Vụ 1-2022-A1-D5N9 |
| Vụ 1-2022-B5-D1N6 |
| Vụ 1-2022-B5-D4N9 |
| Vụ 2-2021-A13-D5N9 |
| Vụ 2-2021-A13-D6N13 |
| Vụ 3-2021-A13-A13N1 |
| Vụ 3-2021-A13-A13N5 |
| Vụ 3-2021-A14-A14N13 |
| Vụ 3-2021-A14-A14N9 |
| Vụ 2-2022-A1-D5N9 |
| Vụ 2-2022-A1-D4N7 |
| Vụ 2-2022-A2-D4N9 |
| Vụ 2-2022-A2-D5N10 |
| Vụ 2-2022-B11-D4N4 |
| Vụ 2-2022-B11-D5N5 |
| Vụ 2-2022-B4-D3N5 |
| Vụ 2-2022-B4-D5N9 |

Nội suy thông tin “Con/Kg” Theo mỗi Unit, theo hàm số:

Biểu đồ đại diện nội suy của 1 unit:



Mô hình:

Output: Độ kiềm.

Input:

Thông số cố định: season, loại ao, công nghệ, giống, ngày thả, diện tích ao, tuổi tôm

Thông số dễ đo: Nhiệt độ, pH, Độ mặn, TDS, Độ đục, DO, mực nước, độ trong

Thông số phòng thí nghiệp: không dùng

* Input: Season, Loại ao, Công nghệ nuôi, Giống tôm, Ngày thả, Nhiệt độ, pH, Độ mặn, TDS, Độ đục, DO, Độ màu, Độ trong, area, Tuổi tôm, Mực nước, Độ kiềm ngày trước.
* Output: Độ kiềm

Mô hình RandomForest (Classic machine learning model):

* Với các thông số được tuning dựa trên bộ:
  + n\_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]
  + 'max\_features': ['auto', 'sqrt']
  + 'max\_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None]
  + 'min\_samples\_split': [2, 5, 10]
  + 'min\_samples\_leaf': [1, 2, 4]

Kết quả:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  |  |  | MRSE | MEA | MAPE | R2 |
| Random Forest Regressor | With the best hyperparameters after tuning | | 17.929 | 12.666 | 8.207% | 0.558 |
| SVRRegresor | kernel: RBF (Radial Basis Function) | | 19.971 | 14.235 | 9.209% | 0.489 |
|  | | | | | | |
| ANN |  | Number of neurons |  |  |  |  |
| 1 hidden layer | 10 | 21.318 | 14.880 | 9.795% | 0.418 |
| 30 | 20.841 | 14.437 | 9.524% | 0.444 |
| 2 hidden layers | 30 15 | 20.417 | 13.846 | 9.078% | 0.466 |
| 16 8 | 20.594 | 14.175 | 9.367% | 0.457 |
| 3 hidden layers | 32 16 8 | 20.689 | 14.047 | 9.243% | 0.452 |
| 15 10 5 | 21.737 | 14.939 | 9.763% | 0.395 |