* 1. Tóm Tắt:

Chất thải hữu cơ động vật rắn (OWPs)[1] từ trang trại chăn nuôi có giá trị phân bón (hàm lượng N, P, K) cao, nhưng cũng có thể gây ra các vấn đề về môi trường khi được sử dụng ở lượng quá mức. Do thành phần của chúng biến đổi lớn, việc phát triển các phương pháp nhanh chóng, đáng tin cậy và chi phí thấp để xác định phân tích hóa học của chúng là rất quan trọng. Phổ cận hồng ngoại (NIRS) cung cấp khả năng phân tích nhanh chóng các mẫu và không yêu cầu chuẩn bị mẫu nhiều, và các nghiên cứu trước đây đã chứng minh rằng NIRS có thể xác định các tham số cấu thành quan trọng nhất của phân bón động vật rắn. Sự phát triển gần đây của máy quang phổ giá thấp đã cho phép thiết bị có thể trang bị cảm biến để đo thành phần trong thời gian thực, và một số ứng dụng đã được thương mại hóa để phân phối OWPs lỏng. Phân tích phổ của những sản phẩm không đồng nhất về bề mặt, độ ẩm này là một thách thức đối với các ứng dụng như vậy, vì việc thu thập phổ phải thực hiện trên các mẫu thô mà không cần chuẩn bị. Để đánh giá độ chính xác mà NIRS xác định nội dung chất khô, chất hữu cơ, amoni, nitơ, phospho, kali, canxi và magiê, các nhà nghiên cứu đã tạo ra một cơ sở dữ liệu hiệu chuẩn lớn đại diện cho phân bón động vật rắn thô gặp ở Brittany. Tổng cộng có 490 mẫu phân OWPs động vật rắn từ các trang trại chăn nuôi đã được thu thập vào đầu mùa xuân từ 270 trang trại ở Brittany (Pháp), trong 2 chiến dịch được tiến hành vào năm 2018 và 2019. Mục tiêu của việc thu thập là thu thập mẫu chất thải hữu cơ động vật rắn mà đại diện cho đa dạng về thành phần của các sản phẩm hữu cơ này (chẳng hạn như loài động vật, chế độ chăn nuôi, cách lưu trữ và tiềm năng phân hủy thành phần hữu cơ.

* 1. Mô Tả Dataset:

Dataset này cung cấp một bộ dữ liệu về thành phần hóa học và phổ gần hồng ngoại của các loại phân bón hữu cơ động vật rắn, bao gồm phân bón từ bò, lợn và gia cầm. Dữ liệu được thu thập với mục tiêu chính là cung cấp các mẫu phân bón hữu cơ động vật rắn đại diện cho đa dạng về thành phần của các sản phẩm này. Điều này bao gồm việc thu thập mẫu từ nhiều nguồn khác nhau, bao gồm các loại động vật khác nhau, phương pháp chăn nuôi, cách lưu trữ khác nhau (ngăn xếp và gom với nhau). Dữ liệu hóa học được thu thập thông qua các phân tích hóa học chi tiết, bao gồm đo lượng chất khô, chất hữu cơ, nitơ tổng và nitơ amoni, photpho, kali, canxi và magie. Các phân tích này tuân theo các phương pháp phân tích được chứng nhận bởi Tổ chức tiêu chuẩn Pháp (AFNOR). Dữ liệu cũng bao gồm thông tin về phổ gần hồng ngoại (NIR) của các mẫu phân bón. Các mẫu đã được quét bằng một thiết bị NIR có tên là Q-interline AgriQuant B8 tại Đan Mạch. Dữ liệu phổ NIR này cung cấp thông tin chi tiết về các bước sóng và độ hấp thụ của các mẫu phân bón. Dữ liệu bao gồm tổng cộng 490 mẫu phân bón động vật rắn được thu thập từ 270 trang trại ở vùng Brittany, Pháp, trong 2 chiến dịch thu thập dữ liệu vào các mùa xuân của năm 2018 và 2019. Dữ liệu hóa học và phổ NIR này đã được lưu trữ và tổ chức dưới dạng tệp "chemical\_analysis.xlsx" và "spectra-1.csv." Tệp "chemical\_analysis.xlsx" chứa thông tin hóa học chi tiết về mỗi mẫu, trong khi tệp "spectra-1.csv" chứa dữ liệu phổ NIR tương ứng. Dữ liệu đã được phân tích và được công bố dưới dạng tài liệu nghiên cứu. Dữ liệu gốc (raw data) đã được đặt trong một kho lưu trữ công cộng có tên "Data INRAE" và có sẵn cho mọi người để sử dụng và tham khảo [2]. Nó có thể giúp nghiên cứu về thành phần hóa học của phân bón hữu cơ động vật rắn, đặc biệt là trong ngữ cảnh của việc sử dụng chúng làm phân bón.

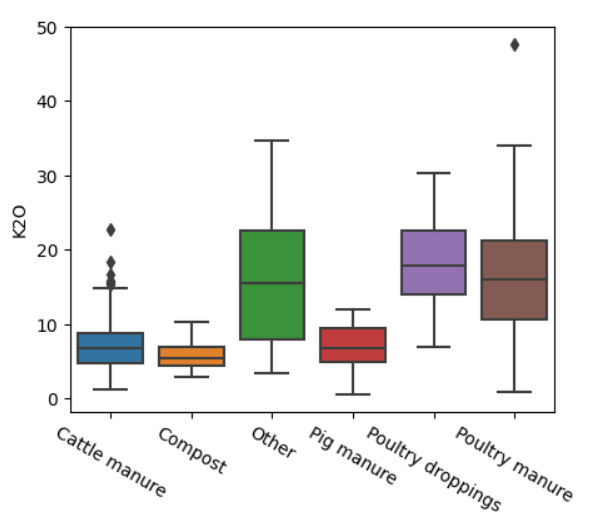
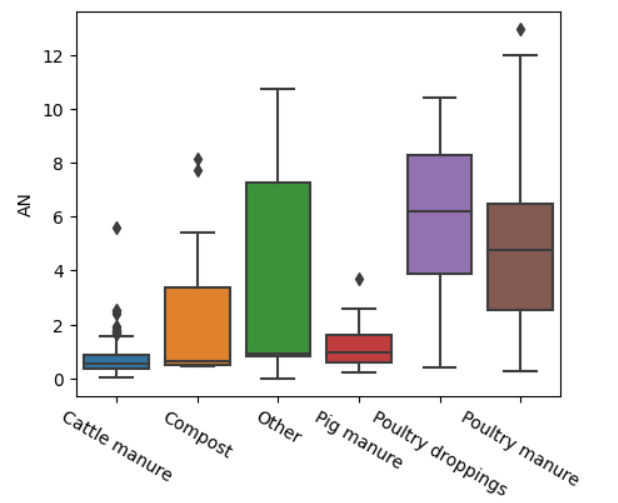
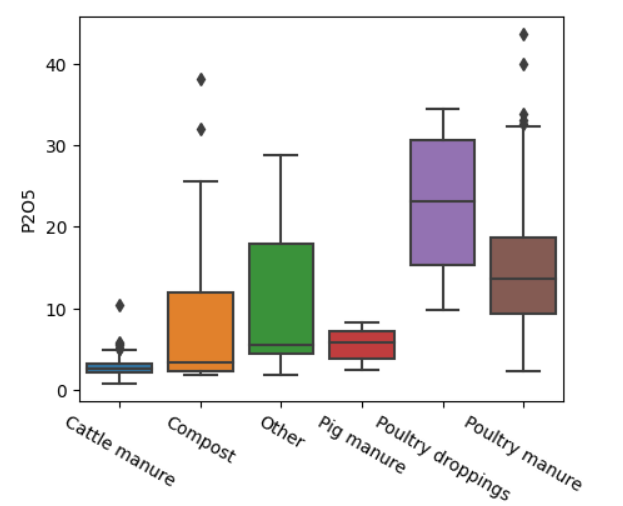
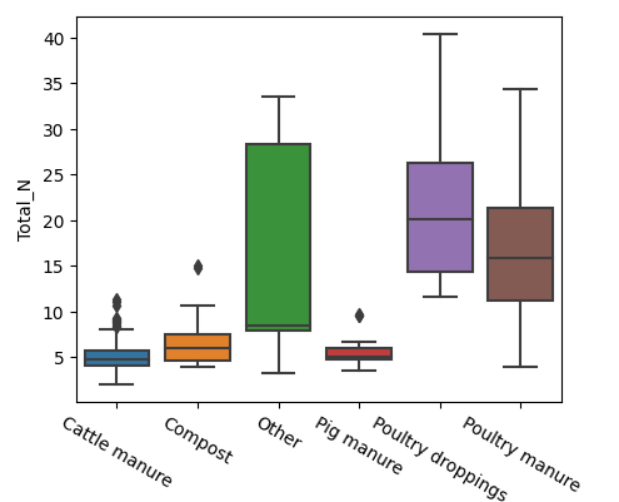
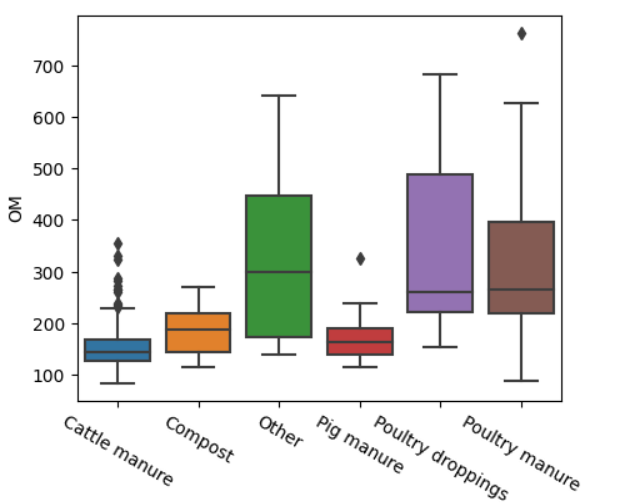
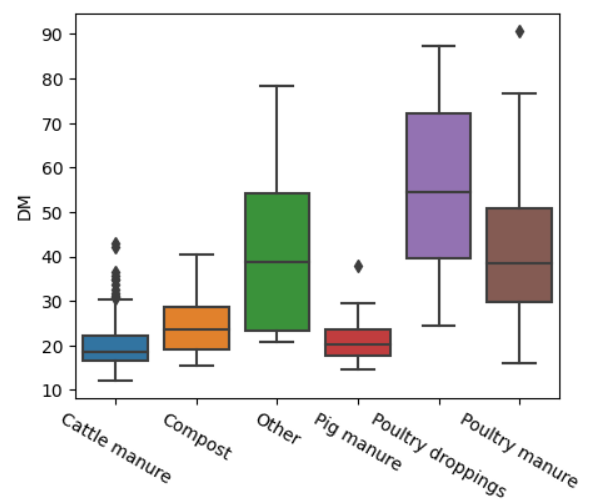
* 1. Lý do lựa chọn Dataset:
* Vì nó chứa thông tin về các mẫu phân bón động vật rắn có tính đa dạng và tỷ lệ mẫu bị mất cân bằng lớn. Điều này là quan trọng khi thực hiện các nhiệm vụ liên quan đến việc xử lý dữ liệu không cân bằng và phân tích dữ liệu phức tạp.
* NIRs từ các mẫu thô làm cho nó trở thành một nguồn tài nguyên quý bởi vì nó cho phép việc thu thập dữ liệu từ các mẫu thô mà không cần chuẩn bị mẫu phức tạp.
* Nghiên cứu giá trị phân bón của các sản phẩm hữu cơ và thực hiện các phân tích liên quan đến hóa học, spectroscopy và dữ liệu không cân bằng.
  1. Mô tả chi tiết:

Dữ liệu phổ NIR của 490 chất thải hữu cơ động vật được thu thập, bao gồm giá trị trung bình của các thông số thành phần hóa học theo loại chất thải hữu cơ (Bảng 1); Boxplots của 6 thông số thành phần hóa học (Hình 1); mối quan hệ giữa chất khô, chất hữu cơ và tổng hàm lượng nitơ (Hình 2).

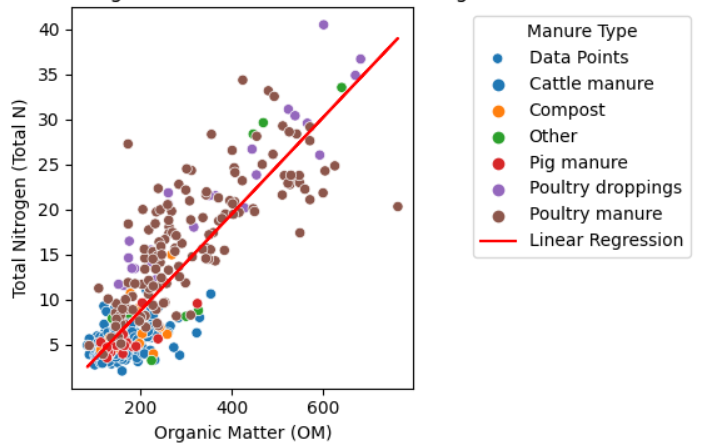
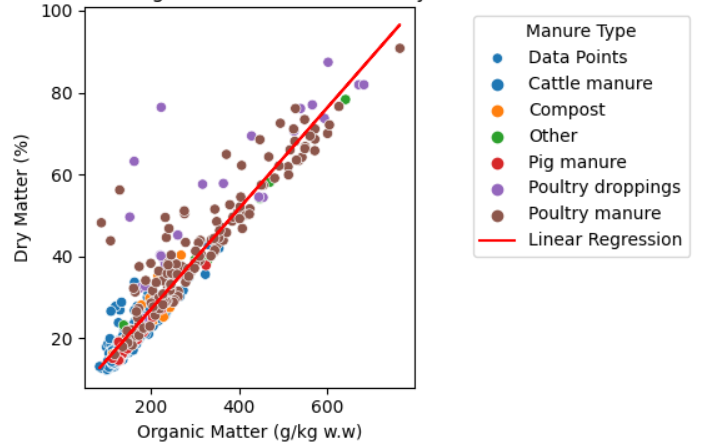
| Product | Sample size | Dry matter | OM | AN | Total N | P2O5 | K2O | CaO | MgO |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cattle manure | 276 | 19.9 | 153.2 | 0.7 | 5.0 | 2.8 | 7.2 | 4.2 | 1.8 |
| Compost | 16 | 24.6 | 183.5 | 2.2 | 7.1 | 10.0 | 5.7 | 10.5 | 5.2 |
| Pig manure | 18 | 213 | 172.3 | 1.2 | 5.7 | 5.6 | 6.6 | 6.3 | 2.1 |
| Poultry droppings | 27 | 54.6 | 350.2 | 6.1 | 21.1 | 22.6 | 18.4 | 62.0 | 7.7 |
| Poultry manure | 144 | 42.2 | 310.9 | 4.8 | 16.6 | 15.1 | 16.2 | 22.9 | 5.6 |
| Other | 9 | 40.4 | 320.2 | 3.7 | 14.9 | 10.6 | 15.5 | 13.4 | 4.1 |

Bảng 1: Giá trị trung bình của các thông số thành phần hóa học theo loại chất thải hữu cơ

(DM: dry matter content; OM: organic matter content AN: ammonium nitrogen content; Total N: total nitrogen content P2O5: phosphorus oxide content; K2O: potassium oxide content; CaO: calcium oxide content; MgO: magnesium oxide content).



Hình 1: Boxplots (Whiskers represent 1.5 times the interquartile range)



Hình 2: (a) Relation between Organic Matter (OM) and Total Nitrogen (Total N)

(b) Relation between Organic Matter (OM) and Total Nitrogen (Total N)