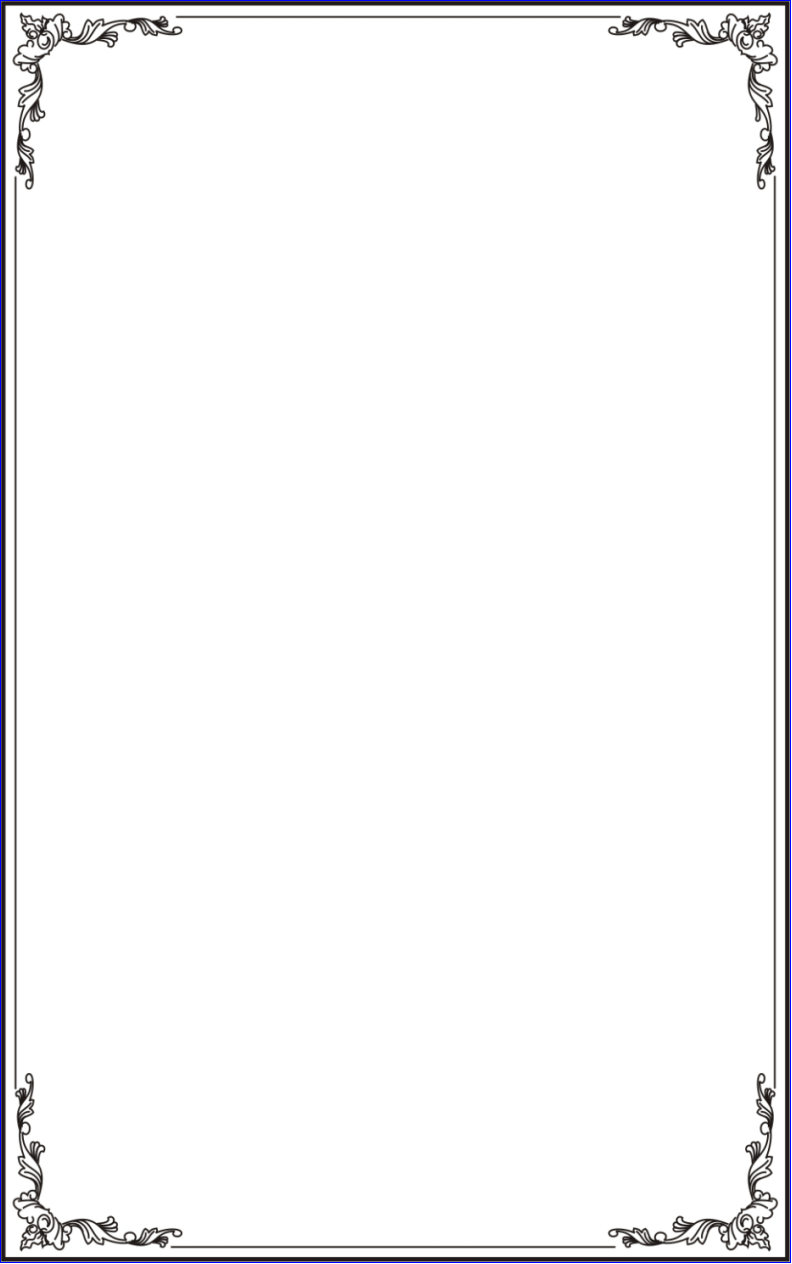
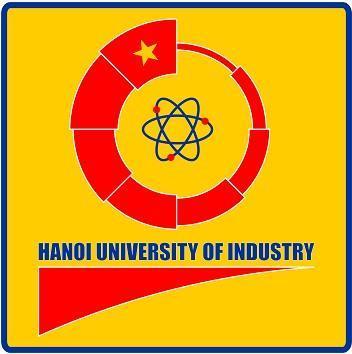
**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**



BÁO CÁO

HỌC PHẦN: HỆ HỖ TRỢ QUYẾT ĐỊNH

**ĐỀ TÀI: HỆ THỐNG DSS DỰ BÁO SẢN XUẤT TRONG NÔNG NGHIỆP**

Giảng viên hướng dẫn : TS. Giang Thành Trung

Nhóm : 04

Lớp : 20243IT6058001 - K16

Sinh viên thực hiện : Trần Đặng Phương Thảo - 2021601876

Phan Thùy Linh - 2021601685

Nguyễn Thị Hồng Nhung - 2021602687

**HÀ NỘI – 2025**

# **LỜI MỞ ĐẦU**

Trong bối cảnh nền kinh tế toàn cầu đang ngày càng phát triển và sự cạnh tranh giữa các ngành công nghiệp trở nên khốc liệt, ngành nông nghiệp cũng không đứng ngoài xu hướng chuyển đổi số. Công nghệ thông tin và các phương pháp phân tích dữ liệu đang mở ra những cơ hội mới cho việc tối ưu hóa quy trình sản xuất và quản lý trong lĩnh vực nông nghiệp. Một trong những công cụ đáng chú ý trong việc hỗ trợ các quyết định sản xuất là hệ thống hỗ trợ quyết định (DSS), giúp các nhà quản lý đưa ra quyết định chính xác và hiệu quả dựa trên phân tích dữ liệu.

Với mục tiêu nâng cao năng suất và hiệu quả trong sản xuất nông nghiệp, việc dự báo sản lượng cây trồng là vô cùng quan trọng. Đề tài của chúng em tập trung vào việc xây dựng một hệ thống DSS dự báo sản xuất trong nông nghiệp sử dụng mô hình hồi quy tuyến tính. Hồi quy tuyến tính, với khả năng phân tích mối quan hệ giữa các yếu tố đầu vào và sản lượng nông sản, đã và đang chứng minh được tính hiệu quả trong việc dự báo các chỉ tiêu sản xuất.

Thông qua việc thu thập và xử lý các dữ liệu về thời tiết, đất đai, giống cây trồng và các yếu tố tác động khác, mô hình hồi quy tuyến tính sẽ giúp dự báo sản lượng một cách chính xác, từ đó giúp người nông dân đưa ra các quyết định hợp lý trong kế hoạch sản xuất.

Chúng em xin gửi lời cảm ơn chân thành đến TS. Giang Thành Trung, người đã luôn hướng dẫn và cung cấp những ý kiến quý báu trong suốt quá trình thực hiện đề tài này. Mặc dù nhóm chúng em đã cố gắng để hoàn thiện báo cáo, nhưng không thể tránh khỏi những thiếu sót. Chúng em rất mong nhận được sự góp ý của thầy để đề tài này được hoàn thiện và ứng dụng thực tế hơn.

# **MỤC LỤC**

[LỜI MỞ ĐẦU 1](#_heading=h.30j0zll)

[MỤC LỤC 2](#_heading=h.1fob9te)

[CHƯƠNG 1: GIỚI THIỆU VỀ HỆ THỐNG DSS DỰ BÁO SẢN XUẤT TRONG NÔNG NGHIỆP 4](#_heading=h.3znysh7)

[1.1.Hệ hỗ trợ ra quyết định 4](#_heading=h.2et92p0)

[1.1.1 Hệ hỗ trợ ra quyết định là gì? 4](#_heading=h.tyjcwt)

[1.1.2. Kiến trúc chung hệ hỗ trợ ra quyết định 5](#_heading=h.4d34og8)

[1.1.3 Mô hình ra quyết định 6](#_heading=h.17dp8vu)

[1.1.4. Mục đích xây dựng hệ hỗ trợ ra quyết định 7](#_heading=h.26in1rg)

[1.1.5. Tính cần thiết và phạm vi ứng dụng của hệ hỗ trợ ra quyết định 7](#_heading=h.lnxbz9)

[1.1.6. Môi trường ra quyết định và các yếu tố ảnh hưởng 8](#_heading=h.44sinio)

[1.1.7. Mô hình Hồi Quy Tuyến Tính (Linear Regression) 10](#_heading=h.3j2qqm3)

[1.2. Mục tiêu của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 12](#_heading=h.3whwml4)

[1.2.1. Mục tiêu chính của dự án 12](#_heading=h.2bn6wsx)

[1.2.2. Vấn đề chính cần giải quyết 13](#_heading=h.qsh70q)

[1.3. Phạm vi và giới hạn Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 14](#_heading=h.3as4poj)

[1.3.1. Phạm vi hỗ trợ 14](#_heading=h.1pxezwc)

[1.3.2. Giới hạn và giả định của hệ hỗ trợ 15](#_heading=h.49x2ik5)

[1.4. Phát biểu bài toán 16](#_heading=h.2p2csry)

[1.4.1. Một số yếu tố ảnh hưởng đến năng suất khi lựa chọn các yếu tố đầu vào 16](#_heading=h.147n2zr)

[1.4.2. Bài toán 16](#_heading=h.3o7alnk)

[CHƯƠNG 2: THIẾT KẾ HỆ HỖ TRỢ DỰ BÁO TRONG NÔNG NGHIỆP 20](#_heading=h.41mghml)

[2.1. Phân tích thực trạng 20](#_heading=h.2grqrue)

[2.1.1. Thực trạng khi chưa có hệ hỗ trợ quyết định (DSS) 20](#_heading=h.vx1227)

[2.1.2. Ưu và nhược điểm khi chưa có hệ hỗ trợ quyết định 20](#_heading=h.3fwokq0)

[2.1.3. Đề xuất xây dựng hệ thống DSS dự báo sản xuất nông nghiệp 21](#_heading=h.1v1yuxt)

[2.2. Thiết kế Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 22](#_heading=h.4f1mdlm)

[2.2.1. Dashboard tổng hợp dữ liệu 22](#_heading=h.2u6wntf)

[2.2.2. Thiết kế mô hình hoạt động tổng thể của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 24](#_heading=h.19c6y18)

[2.2.3. Thiết kế mô hình chức năng của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 25](#_heading=h.3tbugp1)

[2.2.4. Phân tích chi tiết từng chức năng của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 26](#_heading=h.28h4qwu)

[CHƯƠNG 3: XÂY DỰNG HỆ HỖ TRỢ DỰ BÁO SẢN XUẤT TRONG NÔNG NGHIỆP VÀ ÁP DỤNG THỰC TẾ 34](#_heading=h.3l18frh)

[3.1. Xây dựng chức năng Hệ hỗ trợ dự báo sản xuất trong nông nghiệp 34](#_heading=h.206ipza)

[3.1.1. Công cụ và kỹ thuật sử dụng 34](#_heading=h.4k668n3)

[3.1.2. Xây dựng mô hình 34](#_heading=h.1egqt2p)

[3.1.2. Xây dựng chương trình 37](#_heading=h.2dlolyb)

[3.2. Áp dụng trong thực tế 40](#_heading=h.4bvk7pj)

[3.2.1. Môi trường áp dụng 40](#_heading=h.2r0uhxc)

[3.2.2. Đánh giá nhanh về hiệu quả đạt được sau khi áp dụng Hệ hỗ trợ 40](#_heading=h.1664s55)

[TÀI LIỆU THAM KHẢO 42](#_heading=h.25b2l0r)

# **CHƯƠNG 1: GIỚI THIỆU VỀ HỆ THỐNG DSS DỰ BÁO SẢN XUẤT TRONG NÔNG NGHIỆP**

## **1.1.Hệ hỗ trợ ra quyết định**

### **1.1.1 Hệ hỗ trợ ra quyết định là gì?**

#### **1.1.1.1. Khái niệm**

Trong thập niên 1970, Scott Morton đưa ra khái niệm đầu tiên về hệ hỗ trợ ra quyết định (Decision Support Systems – DSS). Ông định nghĩa DSS là hệ thống dựa trên máy tính, có tính tương tác, giúp các nhà ra quyết định dùng dữ liệu và mô hình để giải quyết các bài toán phi cấu trúc.

DSS kết hợp trí lực của con người với năng lực của máy tính để cải tiến chất lượng của quyết định. Đây là các hệ dựa vào máy tính hỗ trợ cho người ra quyết định giải các bài toán nửa cấu trúc (Keen and Scott Morton, 1978).

DSS là tập các thủ tục dựa trên mô hình nhằm xử lý dữ liệu và phán đoán của con người để giúp nhà quản lý ra quyết định (Little, 1970).

Nhu cầu về DSS: Những năm 1980, 1990 điều tra các công ty lớn cho thấy:

* Kinh tế thiếu ổn định
* Khó theo dõi vận hành của doanh nghiệp
* Cạnh tranh gay gắt
* Bộ phận IT quá bận, không giải quyết được các yêu cầu quản lý
* Cần phân tích lợi nhuận, hiệu quả và thông tin chính xác, mới, kịp thời
* Giảm giá phí hoạt động
* Xu hướng tính toán của người dùng

#### **1.1.1.2. Vai trò hệ hỗ trợ ra quyết định**

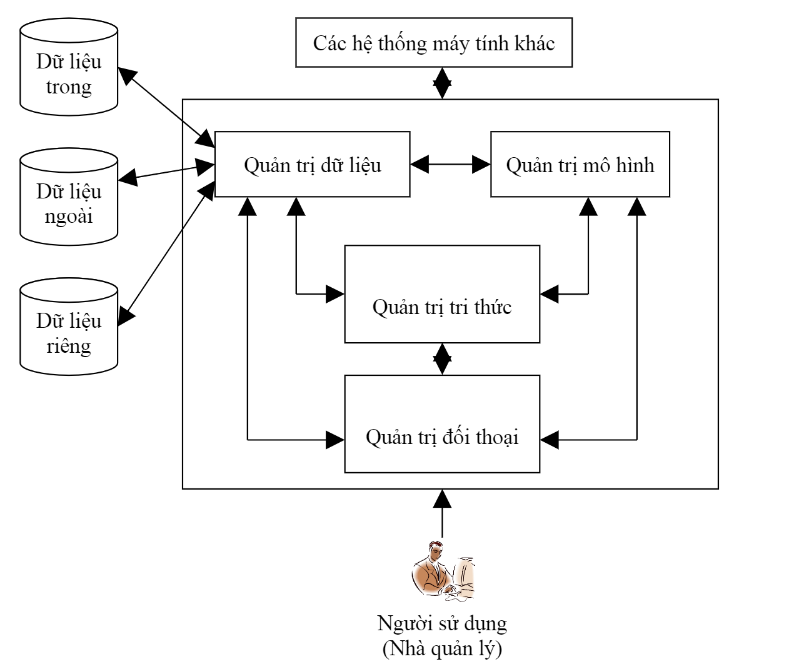
*Lý do sử dụng DSS:*

* Cải thiện tốc độ tính toán;
* Tăng năng suất của cá nhân liên đới;
* Cải tiến kỹ thuật trong việc lưu trữ, tìm kiếm, trao đổi dữ liệu trong và ngoài tổ chức theo hướng nhanh và kinh tế;
* Nâng cao chất lượng của các quyết định đưa ra;
* Tăng cường năng lực cạnh tranh của tổ chức;
* Khắc phục khả năng hạn chế của con người trong việc xử lý và lưu chứa thông tin.

*Các hỗ trợ từ DSS:*

* Cung cấp thông tin trạng thái và dữ liệu thô
* Khả năng phân tích tổng quát
* Mô hình biểu diễn (cân đối tài chính), mô hình nhân quả (dự báo, chuẩn đoán)
* Đề nghị giải pháp, đánh giá
* Chọn lựa giải pháp

### **1.1.2. Kiến trúc chung hệ hỗ trợ ra quyết định**



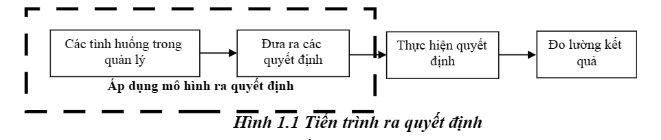
Trong đó:

* Quản trị dữ liệu: bao gồm các CSDL chứa dữ liệu liên quan đến một tình huống và được quản lý bởi phần mềm là hệ quản trị CSDL (quản lý và khai thác).
* Quản trị mô hình: cho phép khai thác và quản lý các mô hình định lượng (xử lý) khác nhau, cung cấp khả năng phân tích cho hệ thống.
* Quản trị đối thoại: cung cấp giao diện cho người dùng để liên lạc và ra lệnh cho Hệ hỗ trợ quyết định.
* Quản trị tri thức: hoạt động như 1 thành phần độc lập, hoặc có thể trợ giúp cho bất kỳ 1 hệ thống nào trong 3 hệ thống nói trên.

### **1.1.3 Mô hình ra quyết định**

Mô hình ra quyết định là một lĩnh vực khoa học quản trị nhằm tìm ra phương pháp tối ưu hoặc hiệu quả nhất của việc sử dụng các nguồn lực có hạn để có thể đạt được các mục tiêu của một cá nhân hoặc một doanh nghiệp đưa ra. Vì lý do này, mô hình ra quyết định thường được hiểu với một nghĩa khác là *Tối ưu hóa.*

Mô hình ra quyết định thường chỉ áp dụng hai giai đoạn đầu của tiến trình ra quyết định đó là các tình huống quản lý và đưa ra các quyết định còn lại các bước thực hiện quyết định và đo lường kết quả đạt được khi ra quyết định thì không được đề cập đến.



Triết lý về ra quyết định mỗi cách tiếp cận đối với các vấn đề khác nhau, phụ thuộc rất nhiều vào môi trường xung quanh, nền tảng kiến thức, kinh nghiệm, trạng thái tâm lý.

*Có nhiều mô hình ra quyết định, nổi bật nhất là:*

* Mô hình tỷ lệ: xem ra quyết định là một quá trình có cấu trúc, rút gọn bài toán thành lập các tham số đo được.
* Mô hình tổ chức: quan tâm nhiều đến các chính sách, định hướng tiếp theo, quan tâm nhiều tới việc định tính.
* Mô hình chính trị: kết quả của liên kết nhóm, thể hiện các khả năng cá nhân.

### **1.1.4. Mục đích xây dựng hệ hỗ trợ ra quyết định**

Mục đích chính của việc sử dụng DSS là trình bày thông tin cho khách hàng một cách dễ hiểu. Một DSS rất hữu ích vì nó có thể được lập trình để tạo nhiều loại báo cáo, tất cả dựa trên thông số kỹ thuật của người dùng.

DSS trợ giúp các hoạt động ra quyết định. Hệ thống này có rất nhiều ưu điểm nổi trội, cung cấp cho người dùng những thông tin hữu ích một cách linh hoạt.

Cung cấp các công cụ trợ giúp việc phát triển và cải thiện các mô hình nhận thức (về nhân và quả) của người ra quyết định bằng cách cung cấp dữ liệu nhanh, đúng & áp dụng các mô hình toán học.

### **1.1.5. Tính cần thiết và phạm vi ứng dụng của hệ hỗ trợ ra quyết định**

#### **1.1.5.1. Tính cần thiết của hệ hỗ trợ ra quyết định**

* Nghiên cứu và hoạch định tiếp thị: Chính sách giá cho khách hàng, dự báo sản phẩm tiêu thụ…
* Hoạch định chiến lược và vận hành: Theo dõi, phân tích và báo cáo về xu hướng thị trường…
* Hỗ trợ bán hàng: Chi tiết và tổng hợp tình hình bán hàng, so sánh và phân tích xu hướng bán hàng…

#### **1.1.5.2. Phạm vi ứng dụng của hệ hỗ trợ ra quyết định**

*Ở đâu****:***

* Ở các quyết định hàng ngày.
* Các quyết định có tính chiến thuật: chọn 1 phương sách thích hợp để đáp ứng 1 mục tiêu nào đó như chuẩn bị ngân sách, quản lý tài chính.
* Các quyết định có tính chiến lược: quyết định chính sách dài hạn, đầu tư dài hạn, tổ chức lại cơ quan, chiến lược tiếp thị.

*Khi nào:*

* Tài chính: mua trang thiết bị, phần mềm.
* Tổ chức: thích hợp.
* Hiệu quả và rủi ro: giảm rủi ro, không phải là hoàn vốn

*Các loại bài toán áp dụng:*

* Đơn giản.
* Có nhiều phương án chọn hơn, hàm đánh giá đơn giản.
* Phức tạp: đa chỉ tiêu.
* Rất phức tạp: vượt quá khả năng xử lý của con người.

*Như thế nào:*

* Thân thiện với người sử dụng về giao diện.
* Tính cập nhật thường xuyên cập nhật trong môi trường biến động.
* Độ chi tiết đảm bảo yêu cầu người sử dụng.

Tần suất sử dụng cao, tính thuận tiện: mềm dẻo, thiết kế tốt

### **1.1.6. Môi trường ra quyết định và các yếu tố ảnh hưởng**

#### **1.1.6.1 Môi trường ra quyết định**

*Các yếu tố tác động*: trực tiếp dễ nhìn ra, dễ lượng hoá, dễ nhìn, dễ hiểu, dễ cấu trúc hoá.

*Các yếu tố về tổ chức tác động đến mọi bài toán quyết định:*

* Chính sách: luật, mệnh lệnh, quan hệ, vay, trả.
* Cấu trúc tổ chức: cách quản lý, cách điều hành, quy chế.
* Uy tín của tổ chức có thể lượng hoá.
* Con người trong tổ chức: hành vi văn hoá, thái độ nhân cách.

*Các yếu tố ngoại cảnh:*

* Tình hình kinh tế.
* Tình hình thị trường.
* Tình hình môi trường.
* Pháp luật: chính sách có thể thay đổi còn pháp luật khó thay đổi.
* Sự chấp nhận của khách hàng.

*Các yếu tố thông tin:*

* Khả năng liên lạc.
* Độ bảo mật của thông tin.
* Độ tin cậy của thông tin: sự chính xác, mức độ cập nhật của thông tin.
* Các giải pháp thông tin tổng hợp, đa dạng, tỉ mỉ chi tiết.
* Giá cả của thông tin: thu nhập, truyền, xử lý.

*Các mục tiêu về quản lý:*

* Sự vận hành: kinh tế, hiệu quả, chất lượng, an toàn.
* Định lượng hay định tính.
* Mức độ rõ ràng của môi trường ra quyết định: rõ, ẩn hay mờ.

*Phân loại môi trường ra quyết định:*

* Môi trường truy nhập được và không truy nhập được.
* Môi trường tất định, không tất định.
* Môi trường tĩnh, động.
* Môi trường chắc chắn, không chắc chắn.

#### **1.1.6.2 Các yếu tố ảnh hưởng đến ra quyết định**

*Trực tiếp, dễ nhìn thấy, dễ hiểu như là:*

* Hạn chế về tài nguyên (ràng buộc ngân sách, khả năng phát triển sản xuất,...)
* Điều kiện vật lý (các tham số đo, các khoảng cách liên hệ,...)
* Các tham số chức năng ảnh hưởng đến hiệu quả (tỷ lệ sản xuất, độ tin cậy của nhà cung cấp, chất lượng sản phẩm, tay nghề nhân công, tỷ lệ sai hỏng, chỉ số kinh tế,...

*Các yếu tố tổ chức:*

* Chính sách: các luật mệnh lệnh, các quan hệ, sự vay trả, sự định hướng thực hiện.
* Cấu trúc: vị trí địa lý, cách quản lý, nghiên cứu – phát triển, cách điều hành (tập trung, phân cấp, ma trận).
* Hình ảnh (uy tín): uy tín xã hội (công cộng), uy tín kinh doanh (thỏa mãn khách hàng, đảm bảo tài chính), uy tín trong nội bộ (không mâu thuẫn).
* Con người: điều kiện xã hội, ý thức chấp hành, hành vi văn hóa, thái độ, nhân cách.

*Các yếu tố ngoại cảnh:*

* Pháp luật, các quy định cần tuân thủ (thời gian sản xuất, môi trường sản xuất, luật chất thải, giới hạn giá cả, xuất khẩu,...)
* Kinh tế: sự đầu tư tài chính, giá cả, thuế.
* Môi trường: thời tiết, yếu tố địa lý, thiên tai.
* Thị trường: sự cạnh tranh, phát triển công nghệ mới.
* Đòi hỏi của khách hàng, nhu cầu của khách hàng.

*Yếu tố thông tin:*

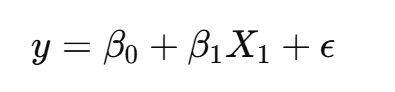
* Khả năng thông tin: độ bảo mật, khả năng truyền thông, phương tiện lưu trữ, nguồn, kênh thông tin (bên trong, bên ngoài).
* Độ tin cậy: sự chính xác, cập nhật, chuẩn mực.
* Giải pháp: thông tin tỉ mỉ, tổng hợp, đa dạng.
* Giá cả: cho thu thập, chuẩn bị, kiểm tra, bảo hành thông tin, cả về thời gian tính và đòi hỏi phần cứng.

### **1.1.7. Mô hình Hồi Quy Tuyến Tính (Linear Regression)**

#### **1.1.7.1. Khái niệm cơ bản**

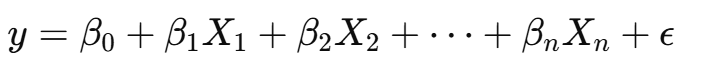
Hồi quy tuyến tính mô hình hóa mối quan hệ giữa biến phụ thuộc y và một hoặc nhiều biến độc lập X bằng cách tìm một hàm số có dạng tuyến tính.

Mô hình hồi quy tuyến tính đơn giản có thể được biểu diễn bằng công thức:



* y là biến phụ thuộc (hay còn gọi là biến mục tiêu).
* X1 là biến độc lập (hay còn gọi là đặc trưng).
* β0 là hệ số chệch (intercept), giá trị của y khi X1=0.
* β1 là hệ số góc (slope), đại diện cho sự thay đổi của y khi X1 thay đổi.
* ϵ là sai số ngẫu nhiên, phản ánh những yếu tố không thể giải thích được bởi mô hình.

Trong trường hợp hồi quy tuyến tính đa biến, công thức sẽ mở rộng như sau:



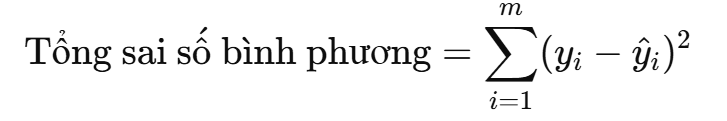
Ở đây, X1,X2,...,Xn​ là các biến độc lập (đặc trưng) và β1,β2,...,βn​ là các hệ số tương ứng của chúng.

#### **1.1.7.2. Mục đích của hồi quy tuyến tính**

* *Dự đoán*: Hồi quy tuyến tính được sử dụng để dự đoán giá trị của biến phụ thuộc y dựa trên các giá trị của các biến độc lập X1,X2,...,Xn
* *Giải thích mối quan hệ*: Hồi quy tuyến tính cũng giúp chúng ta hiểu mối quan hệ giữa các biến độc lập và biến phụ thuộc, ví dụ như biến nào có ảnh hưởng mạnh nhất đến biến phụ thuộc.

#### **1.1.7.3. Ước lượng hệ số hồi quy (fitting the model)**

Hệ số β0,β1,…,βn​ trong mô hình hồi quy được ước lượng bằng cách sử dụng phương pháp tối thiểu hóa sai số bình phương (Least Squares). Mục tiêu của phương pháp này là tìm các giá trị của các hệ số sao cho tổng của bình phương sai số (sai số giữa giá trị dự đoán và giá trị thực tế) là nhỏ nhất.



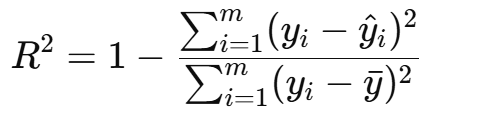
Trong đó:

* m là số lượng quan sát (dữ liệu).
* yi​ là giá trị thực tế của biến y tại quan sát thứ i.
* y^i là giá trị dự đoán của mô hình tại quan sát thứ i.

#### **1.1.7.4 . Đánh giá mô hình hồi quy tuyến tính**

Sau khi huấn luyện mô hình, chúng ta thường sử dụng một số chỉ số để đánh giá độ chính xác của mô hình:

*R-squared (R²):* Đây là chỉ số đo lường mức độ phù hợp của mô hình. Nó cho biết tỷ lệ biến động của biến phụ thuộc được giải thích bởi các biến độc lập trong mô hình. Giá trị R² dao động từ 0 đến 1, với giá trị càng gần 1 thì mô hình càng tốt.



* *Mean Absolute Error (MAE)*: Lỗi tuyệt đối trung bình, là trung bình của các sai số tuyệt đối giữa giá trị thực tế và giá trị dự đoán.
* *Mean Squared Error (MSE)*: Lỗi bình phương trung bình, là trung bình của các sai số bình phương giữa giá trị thực tế và giá trị dự đoán.
* *Root Mean Squared Error (RMSE)*: Căn bậc hai của MSE, cung cấp một chỉ số sai số có cùng đơn vị với dữ liệu gốc.

## **1.2. Mục tiêu của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**

### **1.2.1. Mục tiêu chính của dự án**

Mục tiêu chính của dự án là xây dựng một hệ thống hỗ trợ quyết định nhằm dự báo năng suất cây trồng trong nông nghiệp. Dự án nhằm mục đích tối ưu hóa việc sản xuất nông nghiệp thông qua việc cung cấp các dự báo chính xác về năng suất cây trồng. Hệ thống này sẽ sử dụng dữ liệu môi trường, đặc tính cây trồng, và phương pháp canh tác để đưa ra dự báo về năng suất cây trồng trong những điều kiện thời tiết và môi trường cụ thể.

Các mục tiêu cụ thể của hệ thống bao gồm:

* *Dự báo năng suất cây trồng*: Giúp nông dân hoặc các tổ chức nghiên cứu dự báo năng suất cây trồng dựa trên các yếu tố như loại cây trồng, loại đất, kỹ thuật canh tác, nhiệt độ, lượng mưa, độ ẩm, lượng phân bón, và các yếu tố môi trường khác.
* *Quản lý tài nguyên hiệu quả*: Dựa trên dự báo năng suất, người sử dụng có thể điều chỉnh và tối ưu hóa việc sử dụng tài nguyên như phân bón, nước tưới, và các yếu tố khác, giúp giảm thiểu chi phí và tăng hiệu quả sản xuất.
* *Hỗ trợ ra quyết định*: Cung cấp các công cụ phân tích và báo cáo giúp các nhà nghiên cứu và nông dân đưa ra các quyết định kịp thời và chính xác về việc quản lý cây trồng, điều chỉnh phương pháp canh tác và dự đoán các xu hướng năng suất theo mùa vụ.

### **1.2.2. Vấn đề chính cần giải quyết**

Nông nghiệp là một ngành phụ thuộc nhiều vào các yếu tố môi trường, như thời tiết, khí hậu, và đất đai. Tuy nhiên, việc dự báo chính xác năng suất cây trồng trong các điều kiện này là một thách thức lớn. Các phương pháp truyền thống dựa vào kinh nghiệm của nông dân hoặc các thống kê không đủ chính xác và không thể phản ánh được sự thay đổi nhanh chóng của các yếu tố như nhiệt độ, lượng mưa, độ ẩm, và các yếu tố khác.

Vì vậy, vấn đề chính cần giải quyết là *tối ưu hóa quá trình dự báo năng suất* thông qua việc sử dụng các kỹ thuật học máy và dữ liệu môi trường để cung cấp một dự báo năng suất chính xác và kịp thời. Điều này sẽ giúp giảm thiểu rủi ro trong việc sản xuất, đảm bảo hiệu quả sử dụng tài nguyên, và giúp nông dân có thể lên kế hoạch tốt hơn cho mùa vụ.

## **1.3. Phạm vi và giới hạn Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**

### **1.3.1. Phạm vi hỗ trợ**

1. Loại cây trồng: Hệ thống hỗ trợ dự báo năng suất cho nhiều loại cây trồng khác nhau, bao gồm nhưng không giới hạn ở các loại cây như: lúa, ngô, rau củ, cây ăn trái (như cam, xoài, chuối). Mỗi loại cây trồng có những đặc điểm riêng, do đó cần phải có các phương pháp và mô hình dự báo khác nhau.
2. Các yếu tố đầu vào:
   * Thông tin về môi trường: Dự báo năng suất dựa trên các yếu tố như nhiệt độ, lượng mưa, độ ẩm và lượng phân bón. Hệ thống sẽ thu thập và phân tích các thông tin này để đưa ra dự báo chính xác nhất.
   * Thông tin về đất đai: Bao gồm các loại đất khác nhau như đất thịt, đất cát, đất phèn, và khả năng phù hợp của từng loại đất với từng loại cây trồng.
   * Kỹ thuật canh tác: Hệ thống sẽ tính toán ảnh hưởng của các kỹ thuật canh tác (truyền thống, công nghệ cao) đối với năng suất cây trồng.
3. Đối tượng sử dụng:
   * Các nông dân sẽ sử dụng hệ thống này để lên kế hoạch sản xuất, tối ưu hóa việc sử dụng tài nguyên (nước, phân bón), và giảm thiểu các yếu tố rủi ro.
   * Các tổ chức nghiên cứu nông nghiệp có thể sử dụng hệ thống này để phân tích, khảo sát và đưa ra các kết luận khoa học về năng suất cây trồng.
   * Các chính phủ và cơ quan quản lý nông nghiệp có thể sử dụng hệ thống này để theo dõi tình hình năng suất và lập kế hoạch phát triển sản xuất nông nghiệp quốc gia.

### **1.3.2. Giới hạn và giả định của hệ hỗ trợ**

1. Giới hạn về dữ liệu:
   * Hệ thống chỉ có thể hoạt động hiệu quả khi có dữ liệu đầy đủ và chính xác. Dữ liệu đầu vào về các yếu tố như nhiệt độ, mưa, độ ẩm, và phân bón phải được thu thập từ các nguồn đáng tin cậy.
   * Dữ liệu lịch sử: Hệ thống yêu cầu có dữ liệu lịch sử về năng suất cây trồng trong các năm trước đó để huấn luyện mô hình và đưa ra các dự báo.
2. Giới hạn về mô hình dự báo:
   * Hệ thống sử dụng mô hình hồi quy tuyến tính, điều này có thể không phù hợp với một số loại cây trồng hoặc điều kiện môi trường cực đoan.
   * Các yếu tố không được tính đến trong mô hình như sâu bệnh, biến đổi khí hậu, và tai biến tự nhiên có thể ảnh hưởng đến độ chính xác của dự báo.
3. Giả định về tính ổn định của các yếu tố đầu vào:
   * Hệ thống giả định rằng các yếu tố môi trường (như nhiệt độ, lượng mưa, độ ẩm) có sự biến động ổn định theo thời gian, và các yếu tố này có thể dự đoán được dựa trên các dữ liệu lịch sử.

## **1.4. Phát biểu bài toán**

### **1.4.1. Một số yếu tố ảnh hưởng đến năng suất khi lựa chọn các yếu tố đầu vào**

Có rất nhiều yếu tố có thể ảnh hưởng đến việc quyết định năng suất cây trồng. Việc dự báo năng suất cây trồng là một quá trình phức tạp, chịu ảnh hưởng của nhiều yếu tố khác nhau. Tuy nhiên, những yếu tố sau là những yếu tố quan trọng nhất có thể ảnh hưởng đến năng suất cây trồng

* + Nhiệt độ (°C): Biến số đại diện cho mức nhiệt độ của môi trường, ảnh hưởng trực tiếp đến sự phát triển của cây trồng.
  + Lượng mưa (mm): Mức độ mưa trong khu vực canh tác, ảnh hưởng đến khả năng cung cấp nước cho cây.
  + Độ ẩm (%): Độ ẩm không khí hoặc đất, đóng vai trò quan trọng trong sự phát triển của cây trồng.
  + Loại đất: Đặc điểm của đất như độ pH, độ phì nhiêu, ảnh hưởng trực tiếp đến sự phát triển và năng suất cây trồng.
  + Phân bón (kg/ha): Lượng phân bón được sử dụng, ảnh hưởng đến sự tăng trưởng và năng suất của cây trồng.
  + Kỹ thuật canh tác: Các phương pháp canh tác (ví dụ: tưới tiêu, kỹ thuật gieo trồng, bảo vệ cây trồng khỏi sâu bệnh) mà người nông dân áp dụng, có thể tác động đến năng suất.

### **1.4.2. Bài toán**

#### **1.4.2.1. Phát biểu bài toán**

Mục tiêu của bài toán là xây dựng một hệ thống hỗ trợ quyết định (DSS) dự báo năng suất cây trồng dựa trên các yếu tố ảnh hưởng như thời tiết, loại đất, phân bón, và kỹ thuật canh tác. Hệ thống sẽ sử dụng mô hình hồi quy tuyến tính để dự báo năng suất cây trồng theo từng yếu tố đầu vào và đưa ra các khuyến nghị cải thiện năng suất.

Các yếu tố đầu vào sẽ bao gồm:

* Môi trường: Các yếu tố như nhiệt độ (°C), lượng mưa (mm), độ ẩm (%).
* Loại đất: Đặc điểm của đất như loại đất (đất cát, đất thịt, đất phù sa…).
* Phân bón: Lượng phân bón được sử dụng cho cây trồng (kg/ha).
* Kỹ thuật canh tác: Các phương pháp canh tác cụ thể mà người nông dân sử dụng, như truyền thống, công nghệ cao.

Hệ thống sẽ trả về năng suất cây trồng dự báo (tấn/ha) và cung cấp các khuyến nghị cải thiện năng suất ở các mức độ khác nhau.

#### **1.4.2.2. Thuật toán áp dụng: Hồi quy tuyến tính**

Hệ thống sử dụng hồi quy tuyến tính để dự báo năng suất cây trồng dựa trên các yếu tố đầu vào. Cụ thể:

* *Biến đầu vào (X)* bao gồm các yếu tố như nhiệt độ, lượng mưa, độ ẩm, loại đất, phân bón, và kỹ thuật canh tác.
* *Biến mục tiêu (y)* là năng suất cây trồng (tấn/ha).

Mô hình hồi quy tuyến tính sẽ có dạng:



Trong đó:

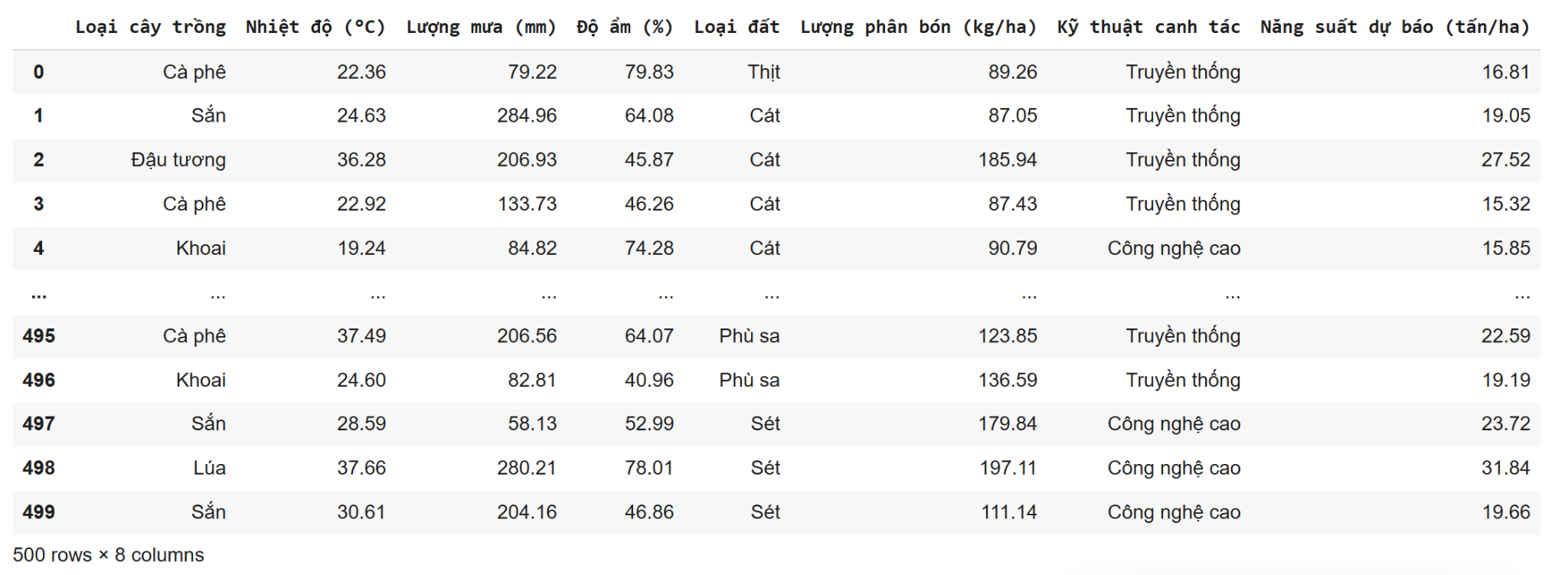
* y là năng suất cây trồng.
* x1,x2,…,xn là các yếu tố đầu vào (thời tiết, loại đất, phân bón, kỹ thuật canh tác).
* β0,β1,…,βn​ là các hệ số hồi quy cần tìm, phản ánh ảnh hưởng của mỗi yếu tố đầu vào đến năng suất cây trồng.
* ϵ là sai số ngẫu nhiên.

#### **1.4.2.3. Quy trình thực hiện**

1. Thu thập và chuẩn bị dữ liệu:
   * Thu thập dữ liệu từ các nguồn về thời tiết, loại đất, phân bón, kỹ thuật canh tác và năng suất cây trồng trong các mùa vụ trước.
   * Dữ liệu sẽ được xử lý và chuẩn hóa để đảm bảo tính đồng nhất.
2. Chuyển đổi dữ liệu:
   * Xử lý dữ liệu danh mục: Các dữ liệu về loại đất và kỹ thuật canh tác có thể là các biến danh mục. Cần thực hiện One-Hot Encoding để chuyển đổi thành dạng số hóa.
   * Chuẩn hóa dữ liệu: Dữ liệu đầu vào được chuẩn hóa đảm bảo rằng tất cả các yếu tố có cùng một thang đo giúp mô hình học tốt hơn.
3. Xây dựng mô hình hồi quy tuyến tính:
   * Mô hình hồi quy tuyến tính sẽ được huấn luyện với bộ dữ liệu đã chuẩn hóa. Mục tiêu là tìm ra các hệ số hồi quy sao cho dự báo năng suất cây trồng càng gần với giá trị thực tế càng tốt.
4. Dự báo năng suất:
   * Sau khi huấn luyện mô hình, hệ thống có thể nhận các yếu tố đầu vào từ người dùng (nhiệt độ, lượng mưa, độ ẩm, phân bón, loại đất, kỹ thuật canh tác), tính toán và dự báo năng suất cây trồng.
5. Khuyến nghị cải thiện năng suất:
   * Hệ thống sẽ đánh giá năng suất dự báo. Nếu năng suất thấp hơn mức trung bình, hệ thống sẽ đưa ra các khuyến nghị về việc cải thiện các yếu tố như phân bón, điều chỉnh kỹ thuật canh tác hoặc thay đổi các điều kiện môi trường (nhiệt độ, lượng mưa).
6. Đánh giá mô hình:
   * Đánh giá hiệu quả của mô hình bằng cách sử dụng các chỉ số đánh giá như MAE (Lỗi tuyệt đối trung bình), MSE (Lỗi bình phương trung bình) và RMSE (Căn bậc hai lỗi bình phương trung bình).

#### **1.4.2.4. Cơ sở dữ liệu**

Cơ sở dữ liệu sẽ được lưu trữ dưới dạng file *.csv*, bao gồm các cột dữ liệu sau:



# **CHƯƠNG 2: THIẾT KẾ HỆ HỖ TRỢ DỰ BÁO TRONG NÔNG NGHIỆP**

## **2.1. Phân tích thực trạng**

### **2.1.1. Thực trạng khi chưa có hệ hỗ trợ quyết định (DSS)**

* *Quyết định thủ công*: Nông dân và các nhà quản lý nông nghiệp chủ yếu dựa vào kinh nghiệm cá nhân và thông tin không đồng bộ, dẫn đến việc ra quyết định thiếu chính xác. Quy trình này thiếu dữ liệu khoa học để hỗ trợ trong việc chọn lựa thời điểm gieo trồng, sử dụng phân bón và chăm sóc cây trồng.
* *Quản lý tài nguyên không hiệu quả*: Việc sử dụng tài nguyên như phân bón, thuốc trừ sâu và nước không tối ưu, gây lãng phí và tăng chi phí sản xuất. Không có công cụ phân tích giúp nông dân tính toán lượng tài nguyên cần sử dụng dựa trên tình trạng đất đai và cây trồng cụ thể.
* *Thiếu công cụ dự báo*: Mô hình dự báo năng suất cây trồng, thay đổi khí hậu và sâu bệnh không tồn tại hoặc không được áp dụng, dẫn đến thiếu chuẩn bị cho các thay đổi đột ngột như thiên tai hay dịch bệnh. Điều này khiến nông dân không thể chủ động đối phó với rủi ro.
* *Khó khăn trong việc tổng hợp và phân tích thông tin*: Dữ liệu về các yếu tố như thời tiết, đất đai, phân bón, v.v. không được tích hợp, làm cho việc ra quyết định trở nên thiếu chính xác và không hiệu quả. Việc phân tích và theo dõi các yếu tố tác động đến sản xuất nông nghiệp gặp nhiều khó khăn.

### **2.1.2. Ưu và nhược điểm khi chưa có hệ hỗ trợ quyết định**

* **Ưu điểm**: Quy trình ra quyết định có thể linh hoạt và tùy chỉnh theo tình hình thực tế và đặc điểm địa phương. Nông dân có thể điều chỉnh chiến lược canh tác theo cảm nhận và kinh nghiệm cá nhân.
* **Nhược điểm**:
  + *Thiếu chính xác*: Quyết định chủ yếu dựa vào cảm tính và thiếu thông tin khoa học, dẫn đến sai sót trong việc chọn lựa phương pháp canh tác, sử dụng phân bón, hay quản lý cây trồng.
  + *Tài nguyên không tối ưu*: Việc sử dụng tài nguyên như phân bón và nước không hiệu quả, gây ra lãng phí và gia tăng chi phí sản xuất.
  + *Không có dự báo*: Thiếu các công cụ dự báo về thời tiết, sâu bệnh, hay thay đổi khí hậu khiến nông dân không thể chuẩn bị cho các tình huống bất ngờ, dẫn đến thiệt hại lớn.
  + *Khó khăn trong tổng hợp dữ liệu*: Việc thiếu hệ thống thông tin tích hợp khiến quá trình phân tích và ra quyết định trở nên khó khăn và không có cơ sở vững chắc.

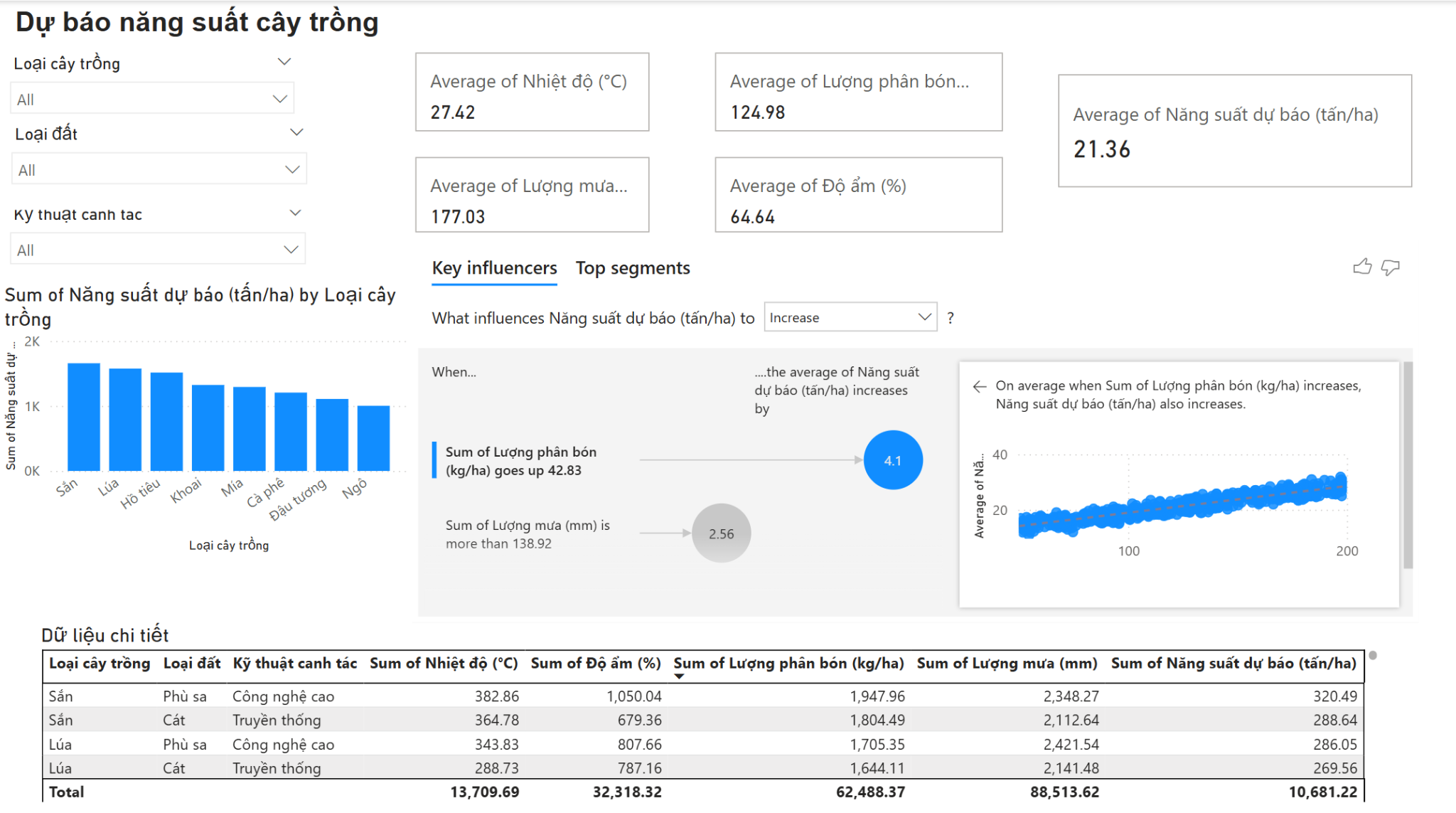
### **2.1.3. Đề xuất xây dựng hệ thống DSS dự báo sản xuất nông nghiệp**

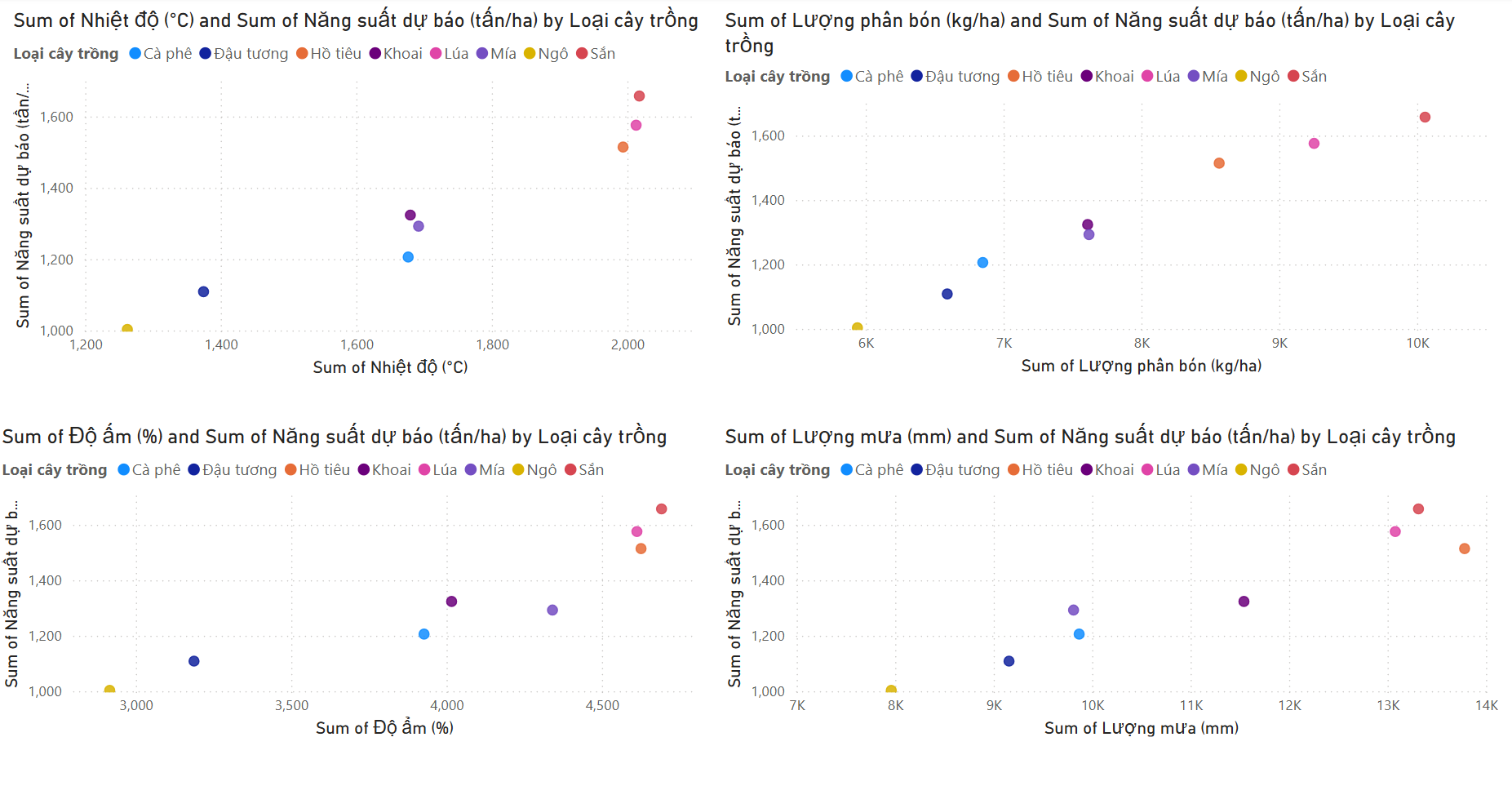
* *Tích hợp và phân tích dữ liệu*: Hệ thống DSS sẽ thu thập và tích hợp dữ liệu từ nhiều nguồn như thời tiết, tình trạng đất, loại cây trồng, phân bón và các yếu tố môi trường. Dữ liệu này sẽ được phân tích để tạo ra những dự báo chính xác về năng suất và các yếu tố ảnh hưởng đến sản xuất.
* *Dự báo và tối ưu hóa quyết định*: Hệ thống sẽ dự báo về thời tiết, các yếu tố nguy cơ như sâu bệnh và các điều kiện môi trường có thể ảnh hưởng đến mùa vụ, giúp nông dân chuẩn bị sẵn sàng. Dựa trên đó, hệ thống sẽ cung cấp các khuyến nghị về việc gieo trồng, sử dụng phân bón và thuốc bảo vệ thực vật phù hợp.
* *Tối ưu hóa tài nguyên*: Hệ thống DSS sẽ giúp tối ưu hóa việc sử dụng tài nguyên (như nước và phân bón) thông qua việc đưa ra các khuyến nghị cụ thể và chính xác. Điều này giúp giảm chi phí sản xuất và cải thiện hiệu quả canh tác.
* *Quản lý thông tin và ra quyết định dựa trên dữ liệu*: Hệ thống sẽ cung cấp cho nông dân một nền tảng dễ sử dụng để theo dõi và quản lý thông tin về sản xuất nông nghiệp, từ đó đưa ra quyết định chính xác và kịp thời. Các quyết định sẽ không còn chỉ dựa vào cảm tính mà sẽ được hỗ trợ bởi dữ liệu thực tế và các dự báo khoa học.

## **2.2. Thiết kế Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**

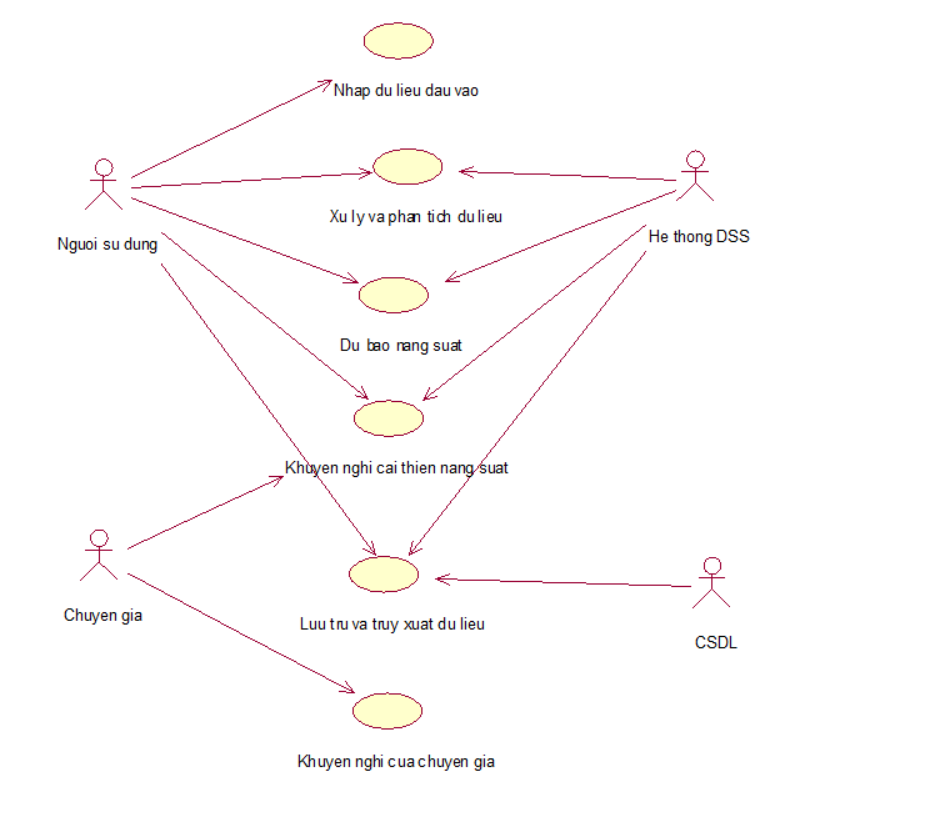
### **2.2.1. Dashboard tổng hợp dữ liệu**

Với sự phát triển của công nghệ phân tích dữ liệu, việc sử dụng Power BI để xây dựng một Dashboard trực quan giúp theo dõi, phân tích và dự báo năng suất cây trồng trở thành một giải pháp hiệu quả. Dashboard này sẽ giúp tổng hợp dữ liệu từ nhiều nguồn, hiển thị các chỉ số quan trọng và hỗ trợ ra quyết định dựa trên dữ liệu thực tế.



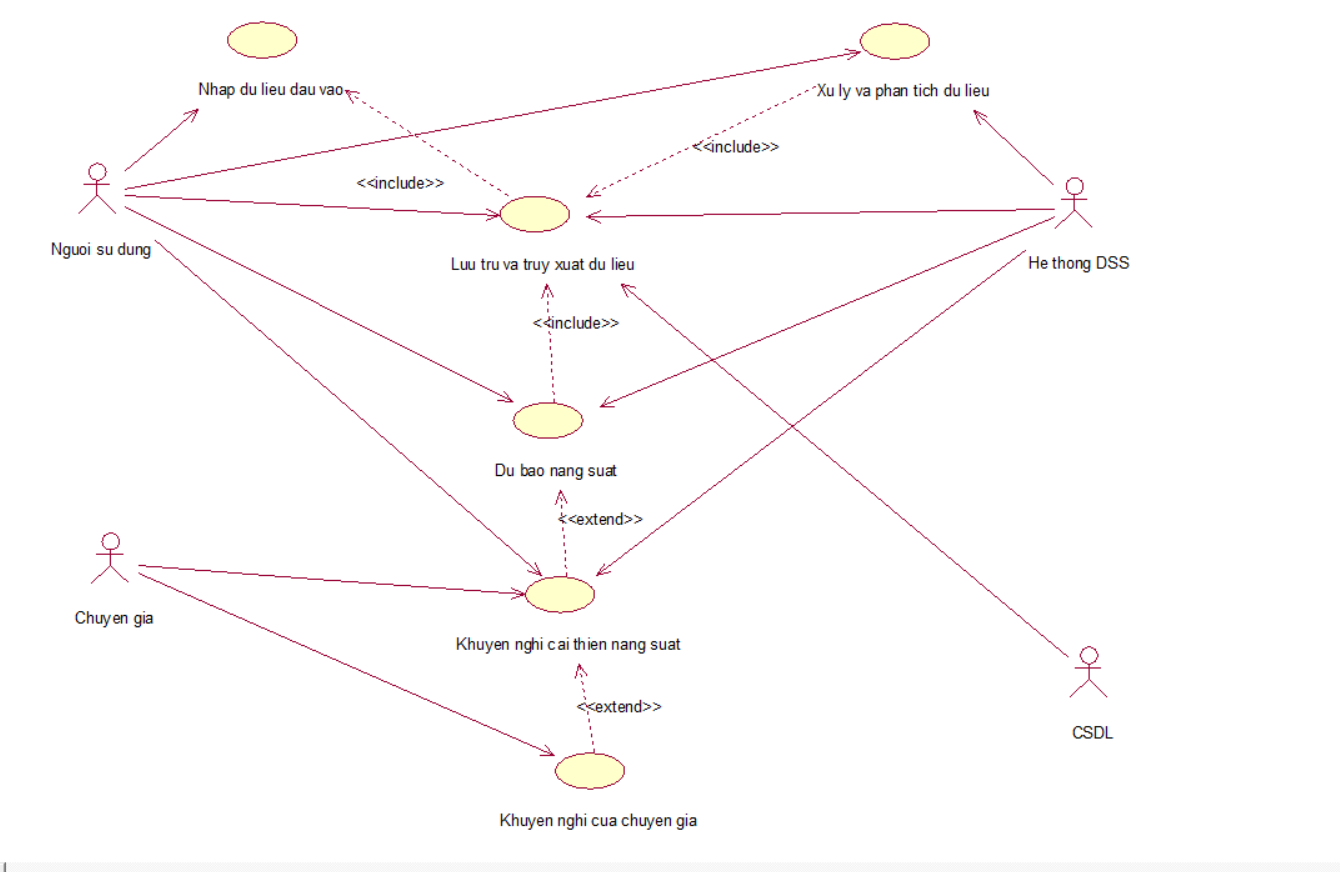


### **2.2.2. Thiết kế mô hình hoạt động tổng thể của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**



1. **Nhập dữ liệu đầu vào**: Cho phép người sử dụng nhập thông tin về loại cây trồng, điều kiện môi trường (nhiệt độ, độ ẩm, lượng mưa), loại đất, phân bón và kỹ thuật canh tác vào hệ thống để xử lý và phân tích.
2. **Xử lý và phân tích dữ liệu**: Hệ thống sẽ xử lý và phân tích các dữ liệu đầu vào mà người sử dụng cung cấp, sử dụng các thuật toán và mô hình để tính toán và đưa ra kết quả dự báo năng suất.
3. **Dự báo năng suất**: Hệ thống dự báo năng suất cây trồng dựa trên các yếu tố đầu vào như loại cây trồng, điều kiện môi trường, lượng phân bón và kỹ thuật canh tác.
4. **Khuyến nghị cải thiện năng suất**: Hệ thống sẽ đưa ra các khuyến nghị về cách cải thiện năng suất cây trồng, bao gồm các kỹ thuật canh tác, loại phân bón và các điều kiện môi trường cần điều chỉnh.
5. **Lưu trữ và truy xuất dữ liệu**: Hệ thống sẽ lưu trữ dữ liệu đầu vào và kết quả phân tích để có thể truy xuất lại khi cần thiết, hỗ trợ người sử dụng trong việc theo dõi lịch sử và hiệu quả của các biện pháp canh tác.
6. **Khuyến nghị của chuyên gia**: Chuyên gia nông nghiệp có thể tham gia và đưa ra những khuyến nghị chuyên sâu để cải thiện hiệu quả năng suất dựa trên kết quả phân tích của hệ thống, giúp người nông dân tối ưu hóa quá trình canh tác.

### **2.2.3. Thiết kế mô hình chức năng của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**



### **2.2.4. Phân tích chi tiết từng chức năng của Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**

#### **2.2.4.1. Mô tả use case Nhập dữ liệu đầu vào**

* **Tên use case**: Nhập dữ liệu đầu vào
* **Mô tả vắn tắt**: Use case này cho phép người sử dụng nhập các thông tin cần thiết về loại cây trồng, điều kiện môi trường (nhiệt độ, độ ẩm, lượng mưa), loại đất, phân bón và kỹ thuật canh tác vào hệ thống để xử lý và phân tích.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu khi người sử dụng đăng nhập vào hệ thống.
    2. Người sử dụng điền thông tin về loại cây trồng, điều kiện môi trường (nhiệt độ, độ ẩm, lượng mưa), loại đất, phân bón và phương pháp canh tác vào các trường dữ liệu của hệ thống.
    3. Hệ thống kiểm tra tính hợp lệ của các thông tin nhập vào.
    4. Nếu tất cả thông tin hợp lệ, hệ thống lưu trữ dữ liệu đầu vào và chuyển sang bước xử lý và phân tích.
    5. Use case kết thúc khi dữ liệu đã được nhập và lưu trữ thành công.
  + **Luồng rẽ nhánh**:
    1. Nếu thông tin đầu vào không hợp lệ (thiếu thông tin, sai định dạng), hệ thống yêu cầu người sử dụng sửa lại và không tiến hành lưu trữ.

2. Tại bất kỳ thời điểm nào trong quá trình thực hiện use case nếu không kết nối được với cơ sở dữ liệu thì hệ thống sẽ hiển thị một thông báo lỗi và use case kết thúc.

* **Các yêu cầu đặc biệt**: Không
* **Tiền điều kiện**: Người sử dụng phải đăng nhập vào hệ thống.
* **Hậu điều kiện**: Không
* **Điểm mở rộng**: Không

#### **2.2.4.2. Mô tả use case Xử lý và phân tích dữ liệu**

* **Tên use case**: Xử lý và phân tích dữ liệu
* **Mô tả vắn tắt**: Use case này cho phép hệ thống xử lý và phân tích các dữ liệu đầu vào mà người sử dụng đã cung cấp, sử dụng các thuật toán và mô hình để tính toán và đưa ra kết quả dự báo năng suất.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu khi dữ liệu đầu vào đã được người sử dụng cung cấp và lưu trữ trong hệ thống.
    2. Hệ thống bắt đầu xử lý các dữ liệu đầu vào như: loại cây trồng, điều kiện môi trường, phân bón và kỹ thuật canh tác.
    3. Hệ thống sử dụng các thuật toán và mô hình phân tích để tính toán các yếu tố ảnh hưởng đến năng suất cây trồng.
    4. Hệ thống tạo ra kết quả phân tích và chuyển tiếp sang bước dự báo năng suất.
    5. Use case kết thúc khi dữ liệu đã được phân tích xong và kết quả sẵn sàng cho bước tiếp theo.
  + **Luồng rẽ nhánh**:
    1. Nếu dữ liệu đầu vào không đủ hoặc không hợp lệ, hệ thống sẽ hiển thị thông báo lỗi và yêu cầu người sử dụng cung cấp thông tin đúng.

2. Tại bất kỳ thời điểm nào trong quá trình thực hiện use case nếu không kết nối được với cơ sở dữ liệu thì hệ thống sẽ hiển thị một thông báo lỗi và use case kết thúc.

* **Các yêu cầu đặc biệt**: Hệ thống cần đảm bảo rằng các mô hình phân tích và thuật toán được tối ưu để mang lại kết quả chính xác và nhanh chóng.
* **Tiền điều kiện**: Dữ liệu đầu vào phải được lưu trữ thành công.
* **Hậu điều kiện**: Kết quả phân tích đã sẵn sàng và hệ thống chuyển sang bước tiếp theo.
* **Điểm mở rộng**: Hệ thống có thể sử dụng các mô hình học máy để cải thiện độ chính xác của kết quả phân tích theo thời gian.

#### **2.2.4.3. Mô tả use case Dự báo năng suất**

* **Tên use case**: Dự báo năng suất
* **Mô tả vắn tắt**: Use case này cho phép hệ thống dự báo năng suất cây trồng dựa trên các yếu tố đầu vào như loại cây trồng, điều kiện môi trường, lượng phân bón và kỹ thuật canh tác.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu khi dữ liệu đã được xử lý và phân tích xong.
    2. Hệ thống sử dụng các mô hình dự báo năng suất để tính toán dựa trên các yếu tố đã nhập.
    3. Hệ thống hiển thị kết quả dự báo năng suất cây trồng lên màn hình cho người sử dụng.
    4. Use case kết thúc khi người sử dụng nhận được kết quả dự báo năng suất.
  + **Luồng rẽ nhánh**:
    1. Nếu hệ thống không thể tính toán được kết quả dự báo năng suất do thiếu dữ liệu hoặc lỗi trong mô hình, hệ thống sẽ thông báo lỗi và yêu cầu người sử dụng kiểm tra lại thông tin đầu vào.
    2. Tại bất kỳ thời điểm nào trong quá trình thực hiện use case nếu không kết nối được với cơ sở dữ liệu thì hệ thống sẽ hiển thị một thông báo lỗi và use case kết thúc.
* **Các yêu cầu đặc biệt**: Hệ thống phải sử dụng các mô hình dự báo đã được kiểm nghiệm và tối ưu để đảm bảo độ chính xác cao.
* **Tiền điều kiện**: Dữ liệu đầu vào đã được xử lý và phân tích thành công.
* **Hậu điều kiện**: Kết quả dự báo năng suất được hiển thị cho người sử dụng
* **Điểm mở rộng**: Không

#### **2.2.4.4. Mô tả use case Khuyến nghị cải thiện năng suất**

* **Tên use case**: Khuyến nghị cải thiện năng suất
* **Mô tả vắn tắt**: Use case này cho phép hệ thống đưa ra các khuyến nghị về cách cải thiện năng suất cây trồng, bao gồm các kỹ thuật canh tác, loại phân bón và các điều kiện môi trường cần điều chỉnh.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu sau khi hệ thống đã hoàn thành việc dự báo năng suất.
    2. Hệ thống phân tích kết quả dự báo năng suất và đưa ra các khuyến nghị để cải thiện năng suất.
    3. Các khuyến nghị sẽ bao gồm các phương pháp canh tác, điều chỉnh phân bón và cải thiện điều kiện môi trường.
    4. Hệ thống hiển thị các khuyến nghị cho người sử dụng.
    5. Use case kết thúc khi người sử dụng nhận được các khuyến nghị.
  + **Luồng rẽ nhánh**:
    1. Nếu không có khuyến nghị nào có thể đưa ra (do dữ liệu quá ít hoặc không có cải tiến rõ ràng), hệ thống sẽ thông báo cho người sử dụng.
    2. Tại bất kỳ thời điểm nào trong quá trình thực hiện use case nếu không kết nối được với cơ sở dữ liệu thì hệ thống sẽ hiển thị một thông báo lỗi và use case kết thúc.
* **Các yêu cầu đặc biệt**: Khuyến nghị phải được cá nhân hóa dựa trên điều kiện cụ thể của người sử dụng và loại cây trồng.
* **Tiền điều kiện**: Dự báo năng suất phải được hoàn thành.
* **Hậu điều kiện**: Các khuyến nghị sẽ hiển thị cho người sử dụng.
* **Điểm mở rộng**: Hệ thống có thể liên kết với các chuyên gia để đưa ra các khuyến nghị chi tiết hơn dựa trên phân tích dữ liệu.

#### **2.2.4.5. Mô tả use case Lưu trữ và truy xuất dữ liệu**

* **Tên use case**: Lưu trữ và truy xuất dữ liệu
* **Mô tả vắn tắt**: Use case này cho phép hệ thống lưu trữ dữ liệu đầu vào và kết quả phân tích để có thể truy xuất lại khi cần thiết, hỗ trợ người sử dụng trong việc theo dõi lịch sử và hiệu quả của các biện pháp canh tác.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu khi dữ liệu đầu vào hoặc kết quả phân tích được lưu trữ trong hệ thống.
    2. Người sử dụng có thể yêu cầu truy xuất lại dữ liệu đã lưu để tham khảo hoặc kiểm tra.
    3. Hệ thống cung cấp giao diện để người sử dụng tìm kiếm và truy xuất lại các thông tin đã lưu trữ.
    4. Use case kết thúc khi người sử dụng đã truy xuất thành công dữ liệu cần thiết.
  + **Luồng rẽ nhánh**:
    1. Nếu dữ liệu không có sẵn hoặc đã bị xóa, hệ thống sẽ hiển thị thông báo lỗi và yêu cầu người sử dụng nhập lại dữ liệu.
    2. Nếu có sự cố trong quá trình truy xuất dữ liệu (chẳng hạn lỗi kết nối), hệ thống sẽ thông báo lỗi.
* **Các yêu cầu đặc biệt**: Dữ liệu phải được bảo mật và lưu trữ lâu dài để người sử dụng có thể tham khảo trong tương lai.
* **Tiền điều kiện**: Dữ liệu phải được nhập và xử lý thành công.
* **Hậu điều kiện**: Dữ liệu đã được truy xuất thành công.
* **Điểm mở rộng**: Hệ thống có thể hỗ trợ việc xuất dữ liệu ra các định dạng khác như CSV hoặc Excel.

#### **2.2.4.6. Mô tả use case Khuyến nghị của chuyên gia**

* **Tên use case**: Khuyến nghị của chuyên gia
* **Mô tả vắn tắt**: Use case này cho phép hệ thống cung cấp các khuyến nghị chi tiết và chuyên sâu từ các chuyên gia nông nghiệp về các vấn đề liên quan đến cây trồng, kỹ thuật canh tác, phân bón, và điều kiện môi trường. Những khuyến nghị này dựa trên phân tích dữ liệu đầu vào của người sử dụng và các mô hình dự báo năng suất.
* **Luồng sự kiện**:
  + **Luồng cơ bản**:
    1. Use case bắt đầu khi hệ thống nhận được yêu cầu từ người sử dụng về các khuyến nghị từ chuyên gia.
    2. Hệ thống sẽ phân tích dữ liệu và kết quả dự báo năng suất, đồng thời đánh giá các yếu tố có thể cải thiện năng suất cây trồng.
    3. Hệ thống kết nối với các chuyên gia (có thể qua một nền tảng trực tuyến hoặc cơ sở dữ liệu chứa thông tin chuyên gia) để nhận các khuyến nghị chi tiết.
    4. Các chuyên gia sẽ đưa ra khuyến nghị dựa trên dữ liệu phân tích và thông tin đã được cung cấp.
    5. Hệ thống hiển thị các khuyến nghị từ chuyên gia cho người sử dụng.
    6. Use case kết thúc khi người sử dụng nhận được khuyến nghị.
  + **Luồng rẽ nhánh**:
    1. Nếu không thể kết nối với chuyên gia (do lỗi mạng hoặc chuyên gia không khả dụng), hệ thống sẽ hiển thị thông báo lỗi và yêu cầu người sử dụng thử lại sau.
    2. Nếu không có khuyến nghị chuyên gia được cung cấp (do thiếu thông tin hoặc dữ liệu không đủ), hệ thống sẽ thông báo rằng không có khuyến nghị khả thi.
    3. Tại bất kỳ thời điểm nào trong quá trình thực hiện use case nếu không kết nối được với cơ sở dữ liệu thì hệ thống sẽ hiển thị một thông báo lỗi và use case kết thúc.
* **Các yêu cầu đặc biệt**:
  + Hệ thống cần phải có cơ chế để kết nối với chuyên gia nông nghiệp, có thể là qua một mạng lưới chuyên gia hoặc dịch vụ tư vấn trực tuyến.
  + Khuyến nghị phải được cung cấp với mức độ chi tiết và có thể thực hiện ngay lập tức, giúp người sử dụng đưa ra quyết định nhanh chóng.
  + Hệ thống phải đảm bảo tính bảo mật khi chia sẻ dữ liệu với chuyên gia (nếu cần thiết).
* **Tiền điều kiện**: Dữ liệu đầu vào và kết quả dự báo năng suất phải được hệ thống xử lý và phân tích xong.
* **Hậu điều kiện**: Khuyến nghị của chuyên gia được hiển thị cho người sử dụng và có thể được lưu trữ để tham khảo sau.
* **Điểm mở rộng**: Hệ thống có thể cung cấp một giao diện người dùng để người sử dụng có thể tương tác với các chuyên gia trong thời gian thực (chat hoặc video call).

# **CHƯƠNG 3: XÂY DỰNG HỆ HỖ TRỢ DỰ BÁO SẢN XUẤT TRONG NÔNG NGHIỆP VÀ ÁP DỤNG THỰC TẾ**

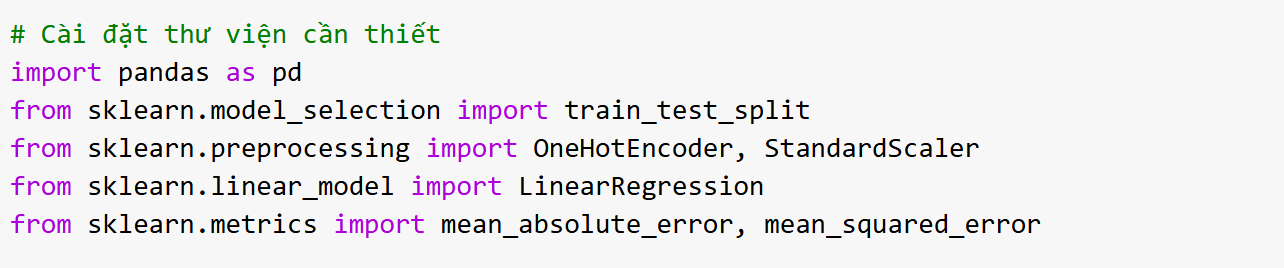
## **3.1. Xây dựng chức năng Hệ hỗ trợ dự báo sản xuất trong nông nghiệp**

### **3.1.1. Công cụ và kỹ thuật sử dụng**

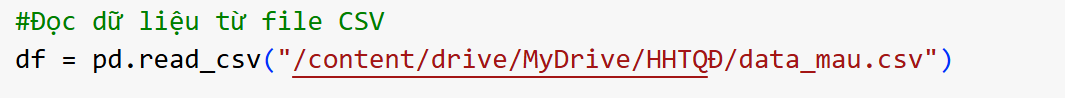
* *Công cụ lập trình*: Hệ thống được xây dựng trên nền tảng Python, sử dụng các thư viện phổ biến trong học máy và phân tích dữ liệu như Pandas, Numpy, Streamlit, và Scikit-learn. Các thư viện này giúp chúng ta xử lý dữ liệu, huấn luyện mô hình và xây dựng giao diện người dùng.
* *Mô hình học máy*: Mô hình Hồi quy tuyến tính (Linear Regression) để dự báo năng suất cây trồng dựa trên các yếu tố như thời tiết, loại đất, phân bón, và kỹ thuật canh tác.
* *Xử lý dữ liệu*: Các dữ liệu đầu vào bao gồm thông tin về loại cây trồng, loại đất, kỹ thuật canh tác và các yếu tố môi trường (nhiệt độ, lượng mưa, độ ẩm, lượng phân bón). Để xử lý dữ liệu này, chúng tôi sử dụng:
* Xử lý các biến danh mục như "Loại cây trồng", "Loại đất" và "Kỹ thuật canh tác".
* Chuẩn hóa dữ liệu (Standardization): Đảm bảo các biến số như nhiệt độ, lượng mưa, độ ẩm, lượng phân bón có cùng thang đo để tránh ảnh hưởng đến kết quả của mô hình.
* *Tạo giao diện người dùng*: Để người dùng có thể dễ dàng sử dụng hệ thống sử dụng Streamlit để tạo giao diện web tương tác. Người dùng có thể chọn loại cây trồng, loại đất, kỹ thuật canh tác và nhập các yếu tố như nhiệt độ, lượng mưa, độ ẩm, lượng phân bón. Dựa trên các lựa chọn và dữ liệu nhập vào hệ thống sẽ cung cấp dự báo năng suất cây trồng.

### **3.1.2. Xây dựng mô hình**

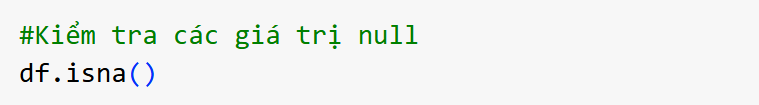
* ***Import các thư viện Python cần dùng:***



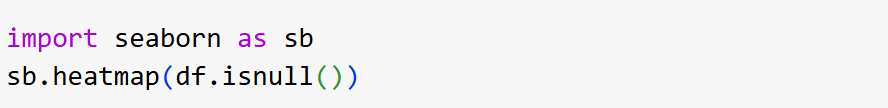
* ***Đọc dữ liệu từ file CSV***



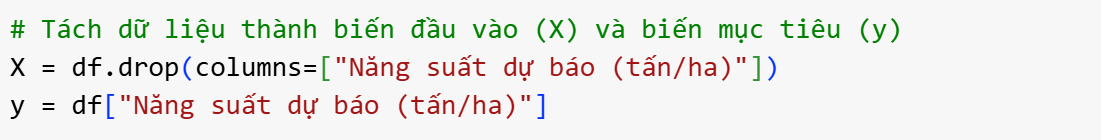
* ***Kiểm tra các giá trị null trong bộ dữ liệu***



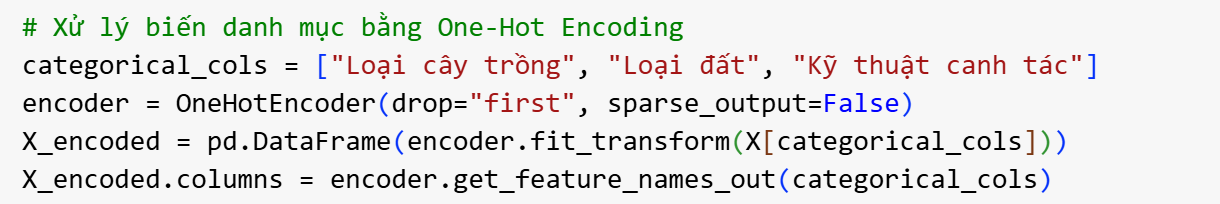
* ***Trực quan hóa dữ liệu bị thiếu***



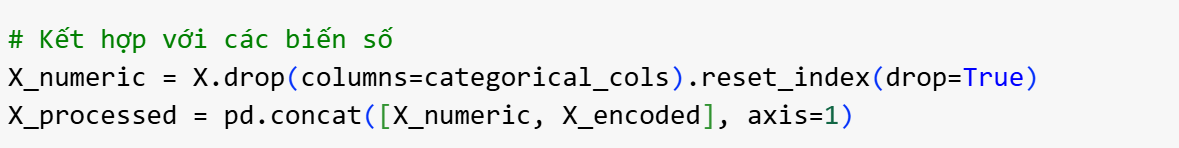
* ***Tách dữ liệu thành 2 biến***



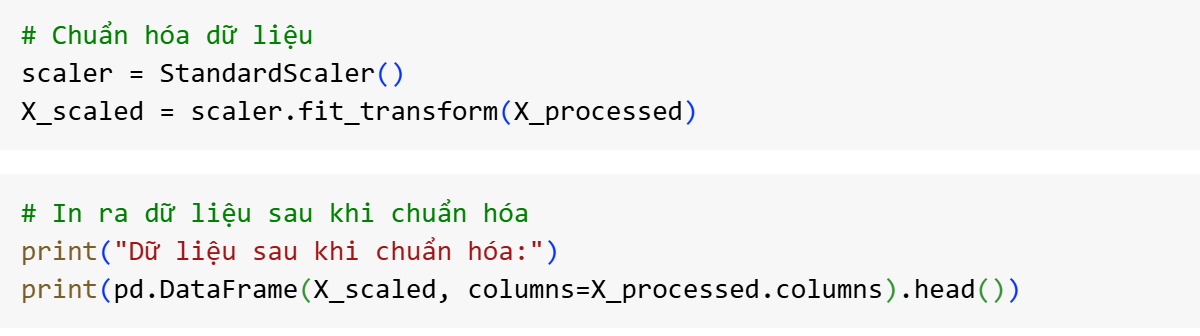
* ***Xử lý biến bằng One-Hot Encoding***



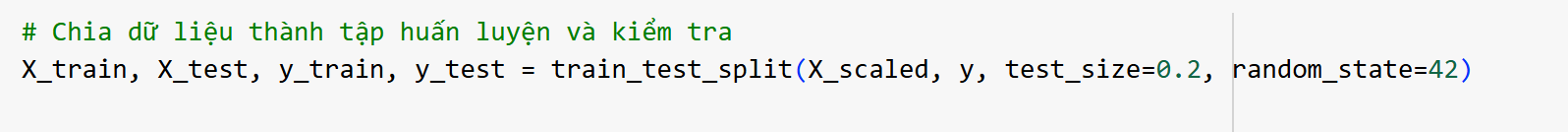
* ***Kết hợp với các biến số***



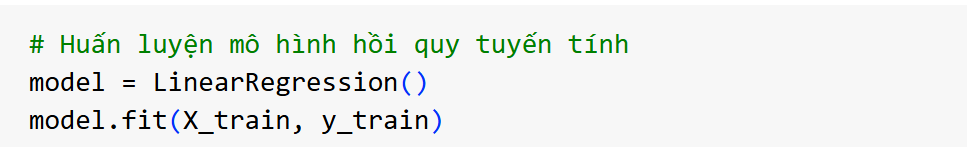
* ***Chuẩn hóa và in ra dữ liệu***



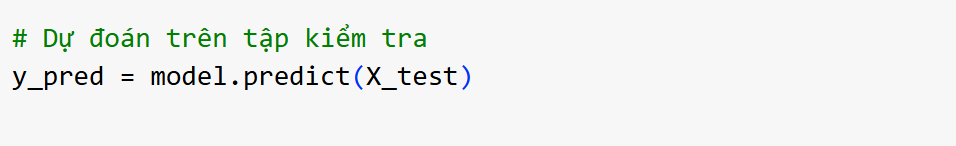
* ***Chia dữ liệu thành tập huấn luyện và kiểm tra***



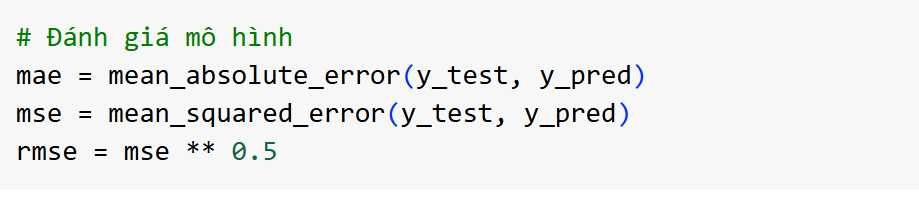
* ***Huấn luyện mô hình***



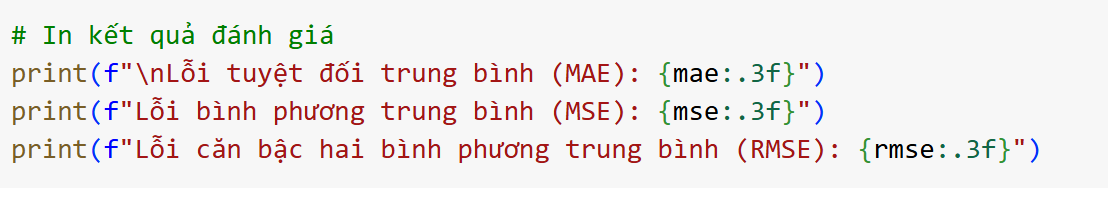
* ***Dự đoán trên tập kiểm tra***



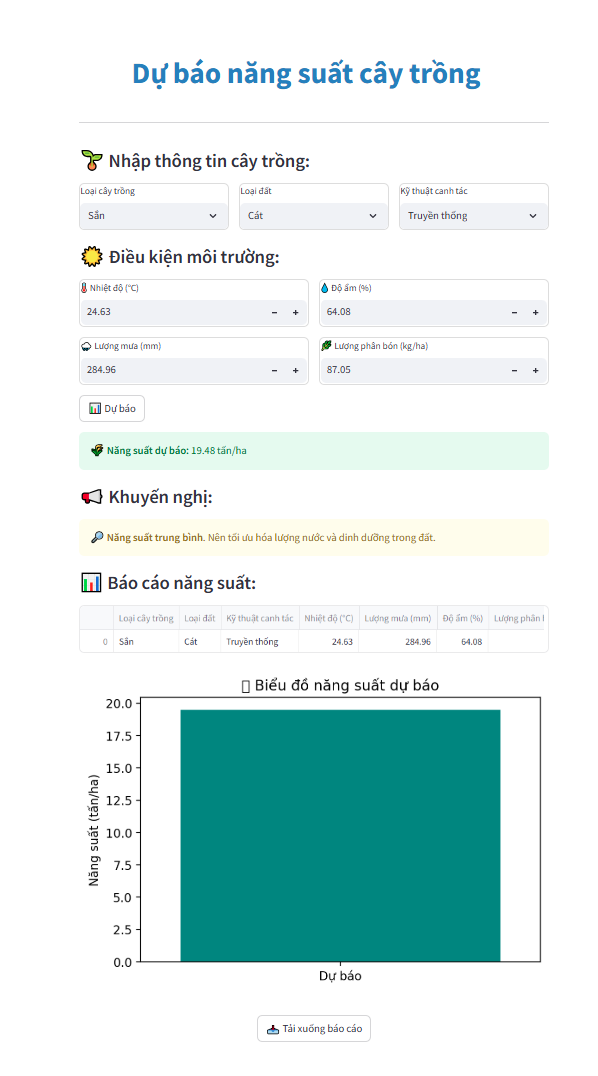
* ***Đánh giá mô hình***



* ***In kết quả đánh giá***



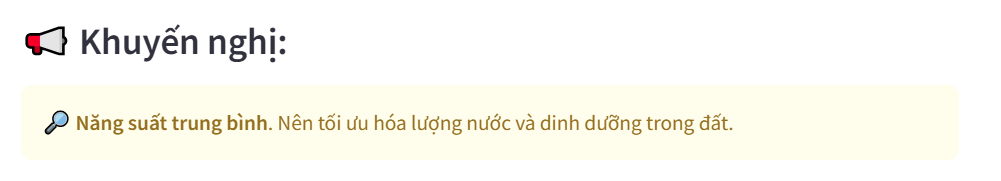
### **3.1.2. Xây dựng chương trình**



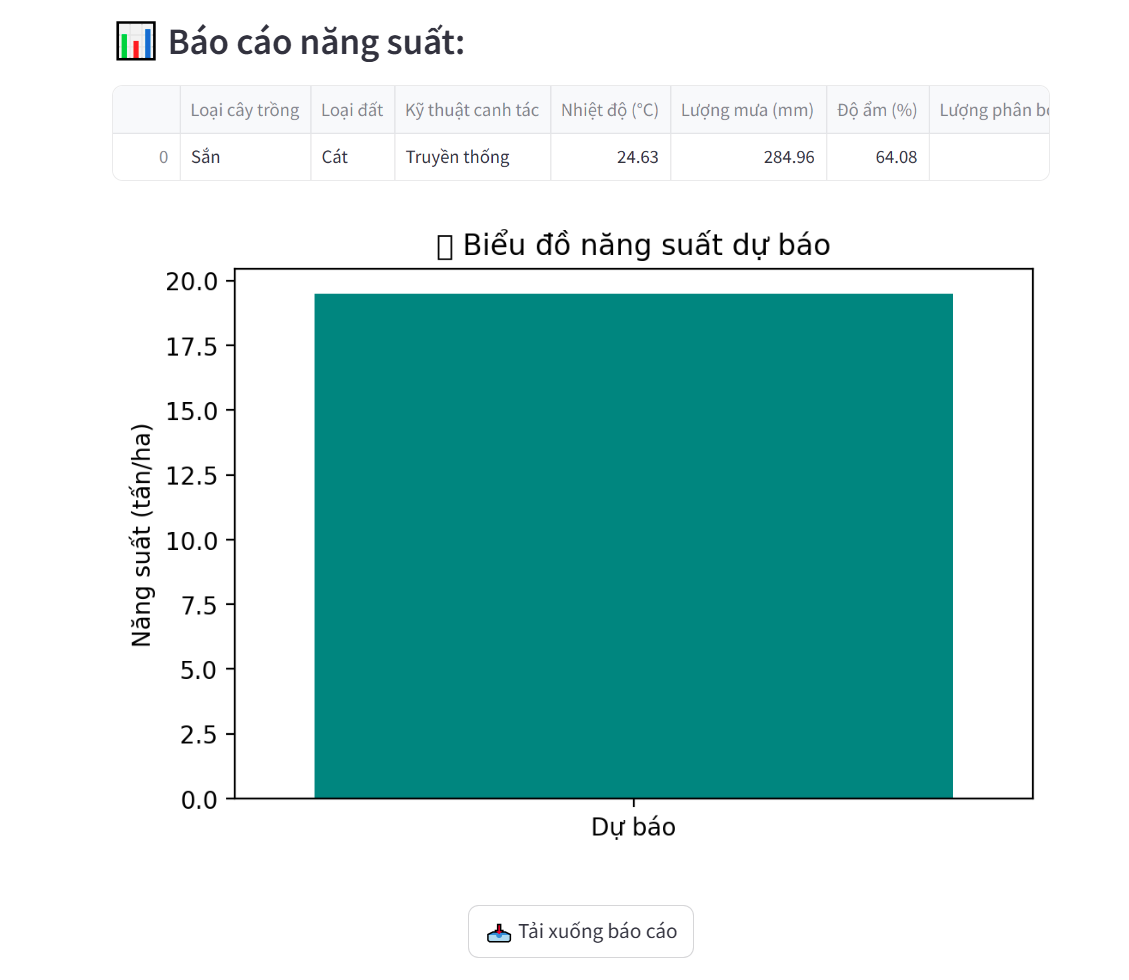
1. *Dự báo năng suất cây trồng*: Hệ thống cho phép người dùng nhập các thông tin liên quan đến loại cây trồng, loại đất, kỹ thuật canh tác và các yếu tố môi trường để dự báo năng suất cây trồng. Mô hình hồi quy tuyến tính sẽ sử dụng các yếu tố này để tính toán và đưa ra kết quả dự báo năng suất.



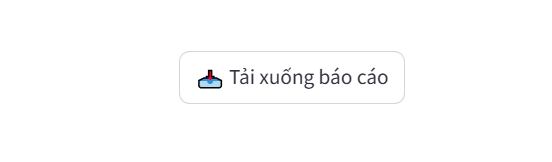
1. *Đưa ra khuyến nghị*: Hệ thống đưa ra khuyến nghị giúp cải thiện hoặc duy trì năng suất giúp người làm nông nghiệp có chiến lược canh tác hiệu quả hơn, tiết kiệm chi phí và tối ưu sản lượng.

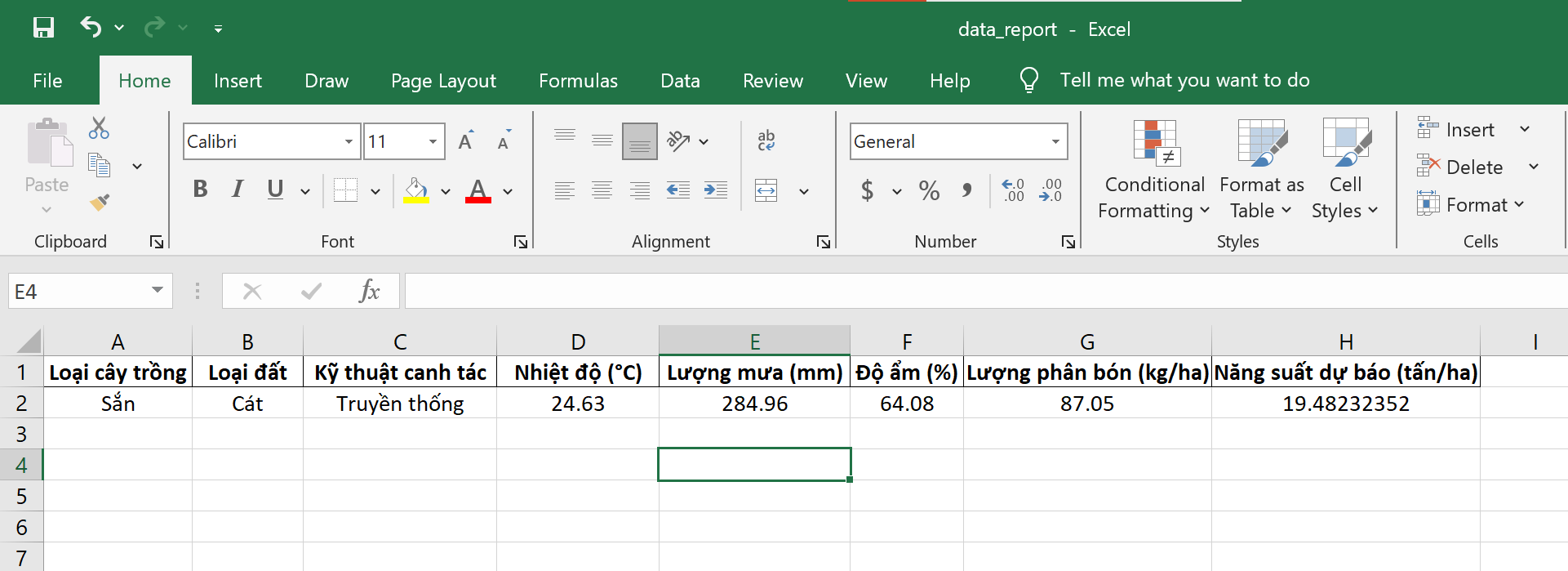


1. *Báo cáo dự báo năng suất*: Hệ thống còn cung cấp báo cáo dự báo năng suất cho từng loại cây trồng, giúp người dùng nắm bắt được xu hướng năng suất trong suốt các mùa vụ. Các báo cáo này được tính toán từ dữ liệu lịch sử và sẽ được hiển thị trong giao diện Streamlit.



1. *Xuất báo cáo dự báo năng suất*: Hỗ trợ người dùng lưu trữ, phân tích và chia sẻ dữ liệu dự báo năng suất cây trồng. Báo cáo này hỗ trợ nông dân và nhà quản lý nông nghiệp trong việc theo dõi xu hướng sản xuất, đánh giá hiệu quả canh tác và điều chỉnh chiến lược phù hợp.





## **3.2. Áp dụng trong thực tế**

### **3.2.1. Môi trường áp dụng**

Hệ thống dự báo năng suất cây trồng có thể áp dụng trong các môi trường nông nghiệp. Các nông dân có thể sử dụng hệ thống để dự báo năng suất cây trồng của mình, từ đó đưa ra các quyết định chính xác hơn về việc sử dụng phân bón, tưới tiêu và các biện pháp chăm sóc cây trồng.

### **3.2.2. Đánh giá nhanh về hiệu quả đạt được sau khi áp dụng Hệ hỗ trợ**

* *Tính chính xác của dự báo*: Mô hình hồi quy tuyến tính đã cho kết quả khá chính xác trong việc dự báo năng suất cây trồng. Đặc biệt, với các yếu tố như nhiệt độ, lượng mưa và kỹ thuật canh tác, mô hình có thể dự báo khá sát với năng suất thực tế trong các điều kiện kiểm tra.
* *Tiện ích cho người dùng*: Giao diện Streamlit cung cấp sự đơn giản và dễ sử dụng, giúp người dùng nhanh chóng nhận được kết quả mà không cần phải có kiến thức chuyên sâu về dữ liệu hoặc học máy.
* *Quyết định chính xác hơn*: Các nông dân và các tổ chức nghiên cứu đã phản hồi rằng họ có thể sử dụng dự báo năng suất từ hệ thống để đưa ra các quyết định về việc sử dụng tài nguyên (như phân bón, nước) một cách tối ưu hơn, từ đó tăng hiệu quả sản xuất.
* *Khả năng mở rộng*: Hệ thống có thể mở rộng để áp dụng cho nhiều loại cây trồng khác nhau và có thể tích hợp với các hệ thống quản lý nông nghiệp khác để nâng cao hiệu quả và tính tự động hóa trong việc dự báo và quản lý năng suất.

# **TÀI LIỆU THAM KHẢO**

**Tài liệu**

1.Nguyễn Minh Tuấn (2018). *Giáo trình Hệ thống hỗ trợ ra quyết định trong nông nghiệp*. NXB Khoa học Tự nhiên và Công nghệ.

**Nghiên cứu khoa học và báo cáo ngành nông nghiệp**

1. Smith, J., & Doe, A. (2021). *Crop Yield Prediction Using Machine Learning Techniques*. International Journal of Agricultural Data Science, 15(4), 56-72.

2. FAO (2023). *FAOSTAT: Crop Production Data*. Food and Agriculture Organization of the United Nations. Truy cập tại: [*https://www.fao.org/faostat*](https://www.fao.org/faostat)

**Công cụ và thư viện phát triển**

1.Python Software Foundation. (2024). *Python Documentation*. Truy cập tại: [*https://docs.python.org*](https://docs.python.org)

2. Streamlit Team. (2024). *Streamlit Documentation: The Fastest Way to Build Data Apps*. Truy cập tại: *https://docs.streamlit.io*