



Tập dữ liệu hoa *Iris* hay tập dữ liệu **Fisher's *Iris*** là một **tập dữ liệu đa biến** được nhà **thống kê** và **nhà sinh học người Anh Ronald Fisher** giới thiệu trong bài báo năm 1936 của ông. Việc sử dụng nhiều phép đo trong các bài toán phân loại như một ví dụ về **phân tích phân biệt tuyến tính**.^[1] Đôi khi nó được gọi là **tập dữ liệu *Iris* của Anderson** vì **Edgar Anderson** đã thu thập dữ liệu để định lượng sự biến đổi **hình thái** của hoa *Iris* của ba loài có liên quan.^[2] Hai trong số ba loài được thu thập ở **Bán đảo Gaspé** "tất cả từ cùng một đồng cỏ, và được hái trong cùng một ngày và được đo cùng một lúc bởi cùng một người với cùng một thiết bị".^[3]

Bộ dữ liệu bao gồm 50 mẫu của mỗi loài trong số ba loài *Iris* (*Iris setosa* , *Iris virginica* và *Iris versicolor*). Bốn **đặc điểm** được đo từ mỗi mẫu: chiều dài và chiều rộng của các **lá đài** và **cánh hoa** , tính bằng cm. Dựa trên sự kết hợp của bốn đặc điểm này, Fisher đã phát triển một mô hình phân biệt tuyến tính để phân biệt các loài với nhau.