

# **Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus (Sử dụng thuật toán học ADAP để dự báo bệnh tiểu đường)**

Hannah Morales

# Nội dung trình bày



- Các mục chính:
  - Bối cảnh và Tầm quan trọng của vấn đề
  - Tổng quan về tập dữ liệu (Data Summary & EDA)
  - Phân tích các yếu tố nguy cơ (Univariate & Bivariate Analysis)
  - Giới thiệu thuật toán học ADAP (Methodology)
  - Kết quả thực nghiệm (Results)
  - Kết luận và Hướng phát triển (Conclusion & Future Work)

# Table of

## Bối cảnh và Vấn đề (Introduction & Problem Statement)

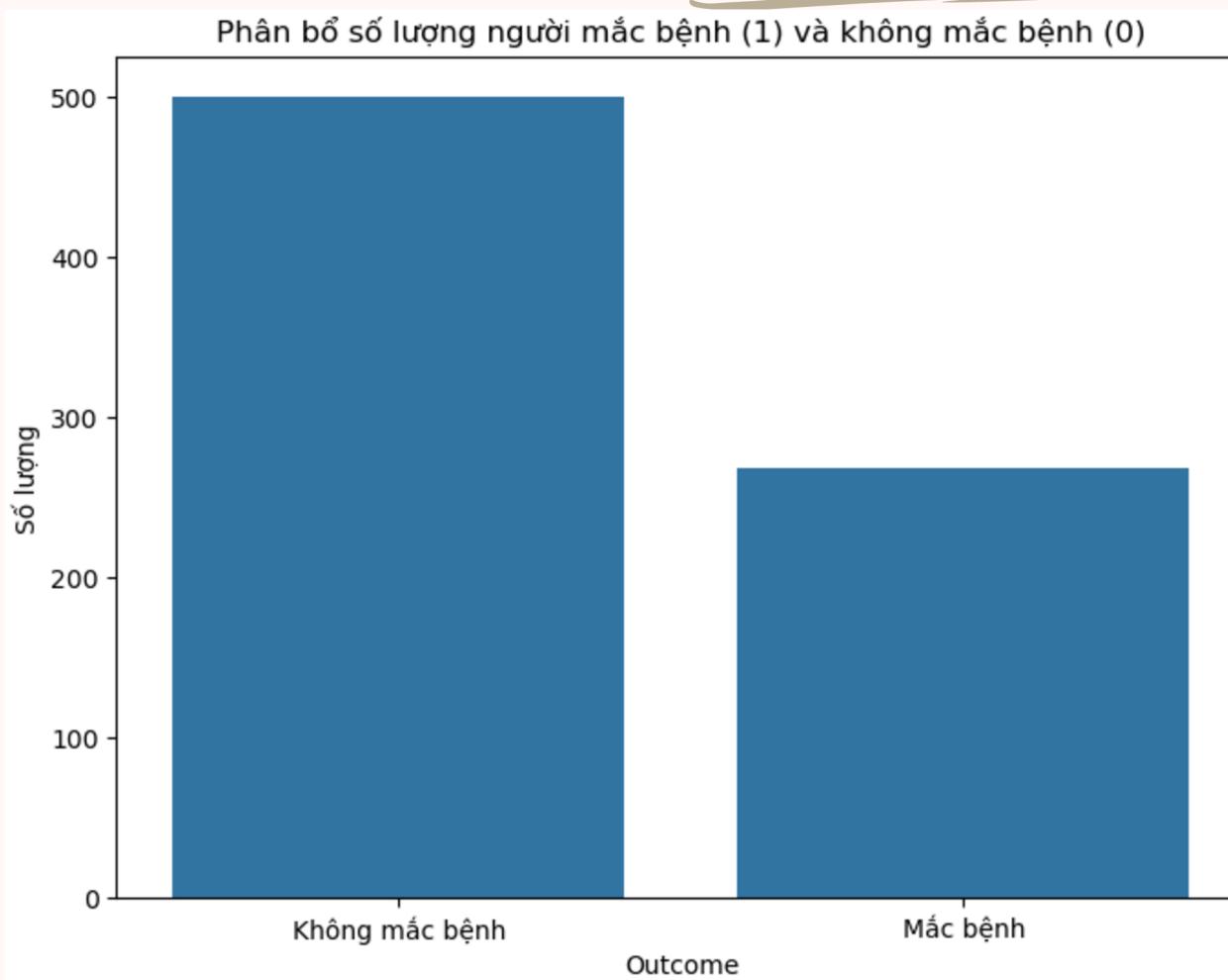
- Khả năng dự báo và chẩn đoán là yếu tố cốt lõi trong y học.
- Các phương pháp thống kê truyền thống (hồi quy, phân tích phân biệt) có thể không hiệu quả khi:
  - Kích thước mẫu nhỏ.
  - Không biết dạng của mối quan hệ chức năng.
  - Mối quan hệ phức tạp và có nhiều tương quan chéo giữa các biến.
- Mạng nơ-ron (Neural Networks) mang lại một hướng tiếp cận đầy hứa hẹn cho các bài toán nhận dạng mẫu trong y tế.
- Mục tiêu nghiên cứu: Thử nghiệm khả năng của mô hình mạng nơ-ron sơ khai (ADAP) để dự báo sự khởi phát của bệnh tiểu đường trong vòng 5 năm ở nhóm dân số có nguy cơ cao (Pima Indians).
-

# Tổng quan về Tập dữ liệu (Data Summary)

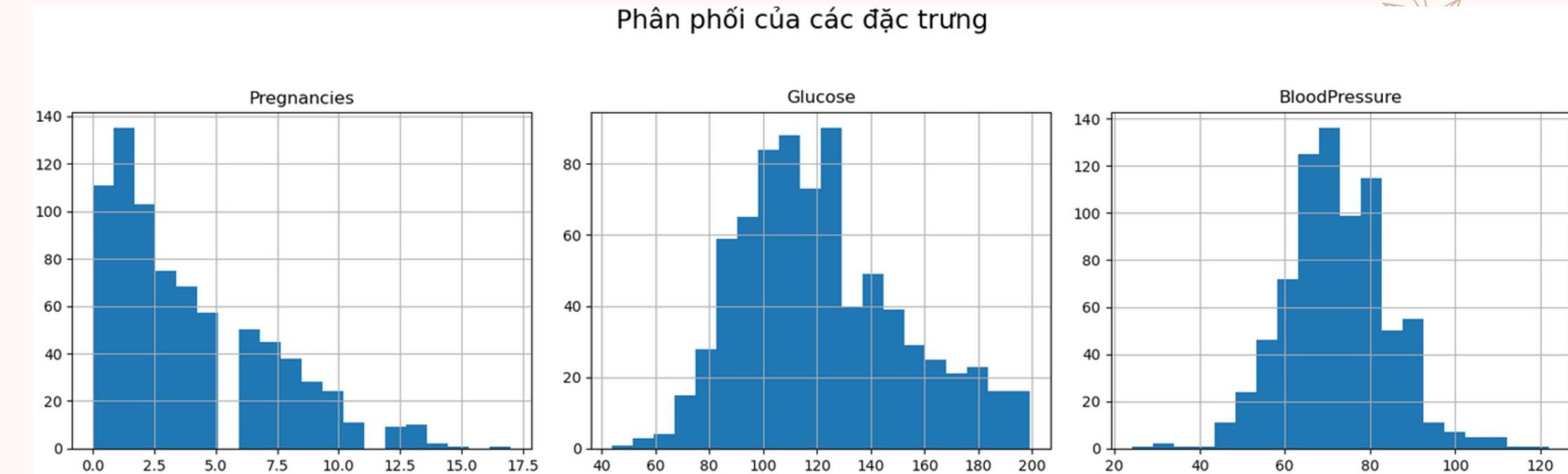
- **Nguồn dữ liệu:** Dân số người Pima Indian gần Phoenix, Arizona, được nghiên cứu liên tục từ năm 1965 do tỷ lệ mắc bệnh tiểu đường rất cao.
- **Đối tượng:** 768 phụ nữ từ 21 tuổi trở lên.
- **Các biến đầu vào (Input Variables):** Liệt kê 8 biến được chọn từ bài báo.
  - Number of times pregnant (Số lần mang thai)
  - Plasma Glucose Concentration (Nồng độ glucose huyết tương)
  - Diastolic Blood Pressure (Huyết áp tâm trương)
  - Triceps Skin Fold Thickness (Độ dày nếp gấp da cơ tam đầu)
  - 2-Hour Serum Insulin (Insulin huyết thanh sau 2 giờ)
  - Body Mass Index (Chỉ số khối cơ thể - BMI)
  - Diabetes Pedigree Function (Hàm phả hệ tiểu đường)
  - Age (Tuổi)
- **Biến mục tiêu (Outcome):** Chẩn đoán mắc bệnh tiểu đường trong vòng 5 năm (1: Có, 0: Không).

# Phân tích Dữ liệu Khám phá (EDA)

## - Phân bổ dữ liệu



**Biểu đồ 1 (Count Plot):**  
Tỷ lệ mắc bệnh

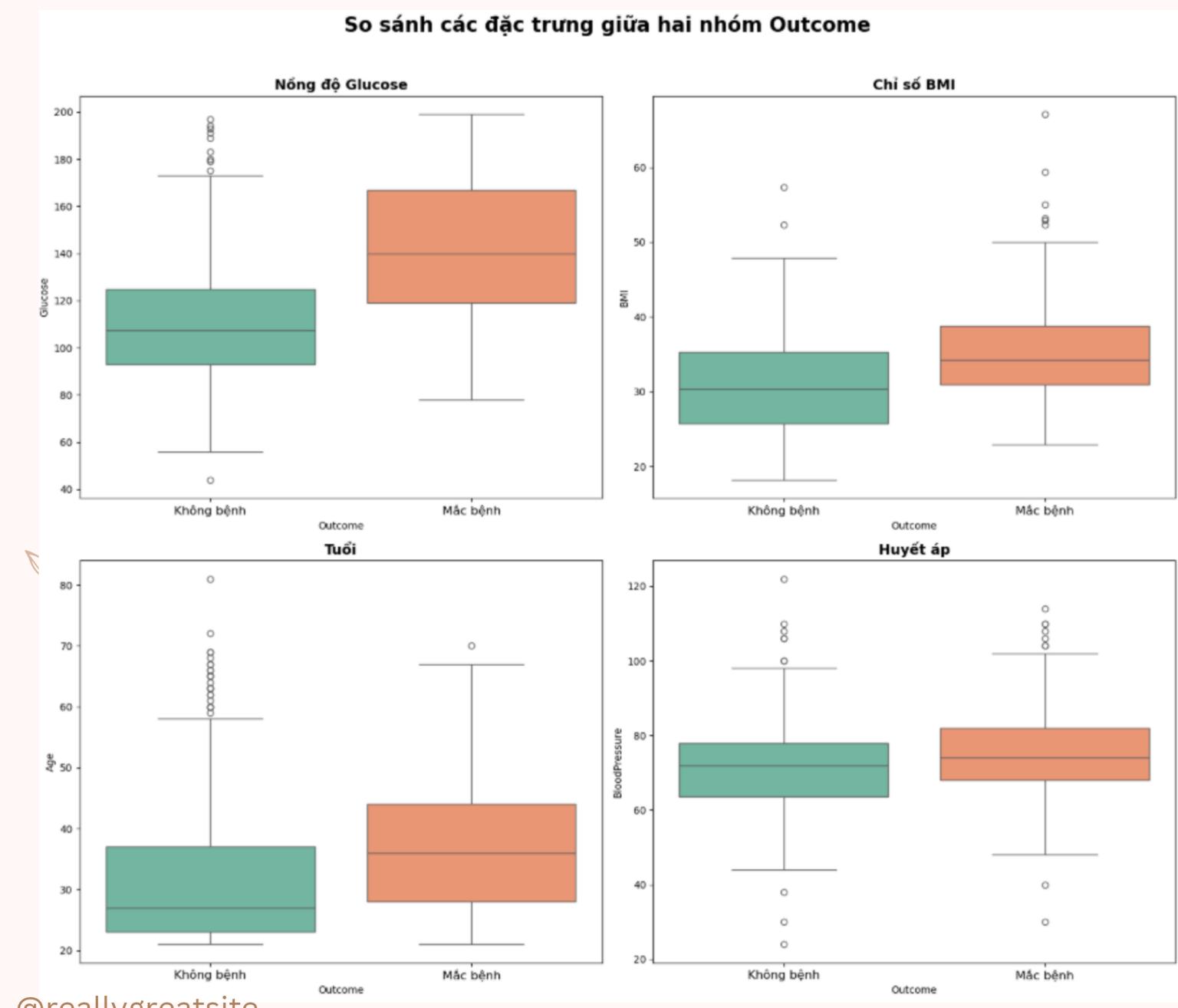


@reallygreatsite

**Biểu đồ 2 (Histogram):**  
Phân phối của các đặc trưng chính

# Phân tích các yếu tố nguy cơ

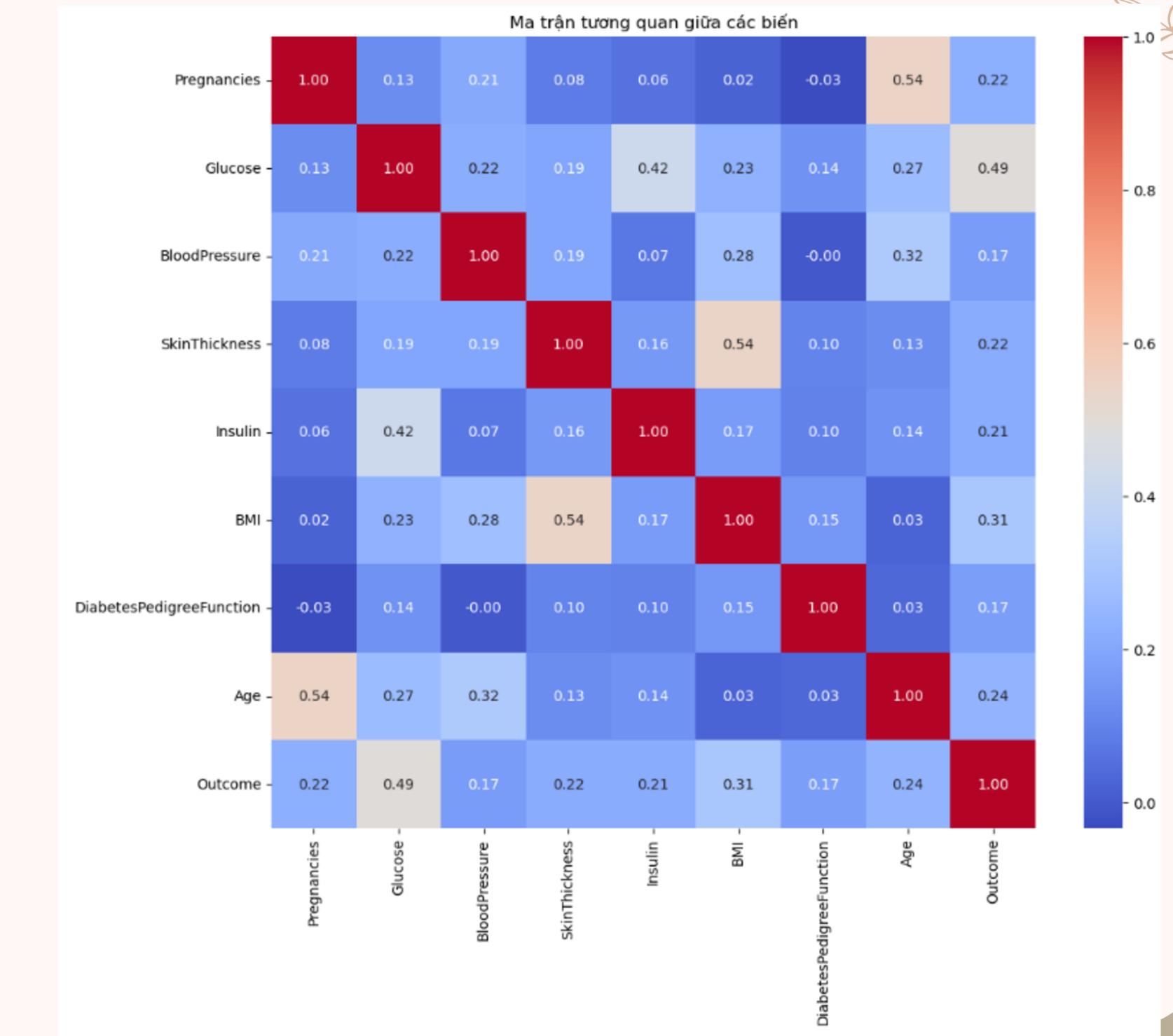
- Glucose: Mức glucose là yếu tố khác biệt rõ rệt nhất. Nhóm mắc bệnh có nồng độ glucose trung bình cao hơn đáng kể.
- BMI & Tuổi: Người mắc bệnh có xu hướng có chỉ số BMI và tuổi cao hơn.



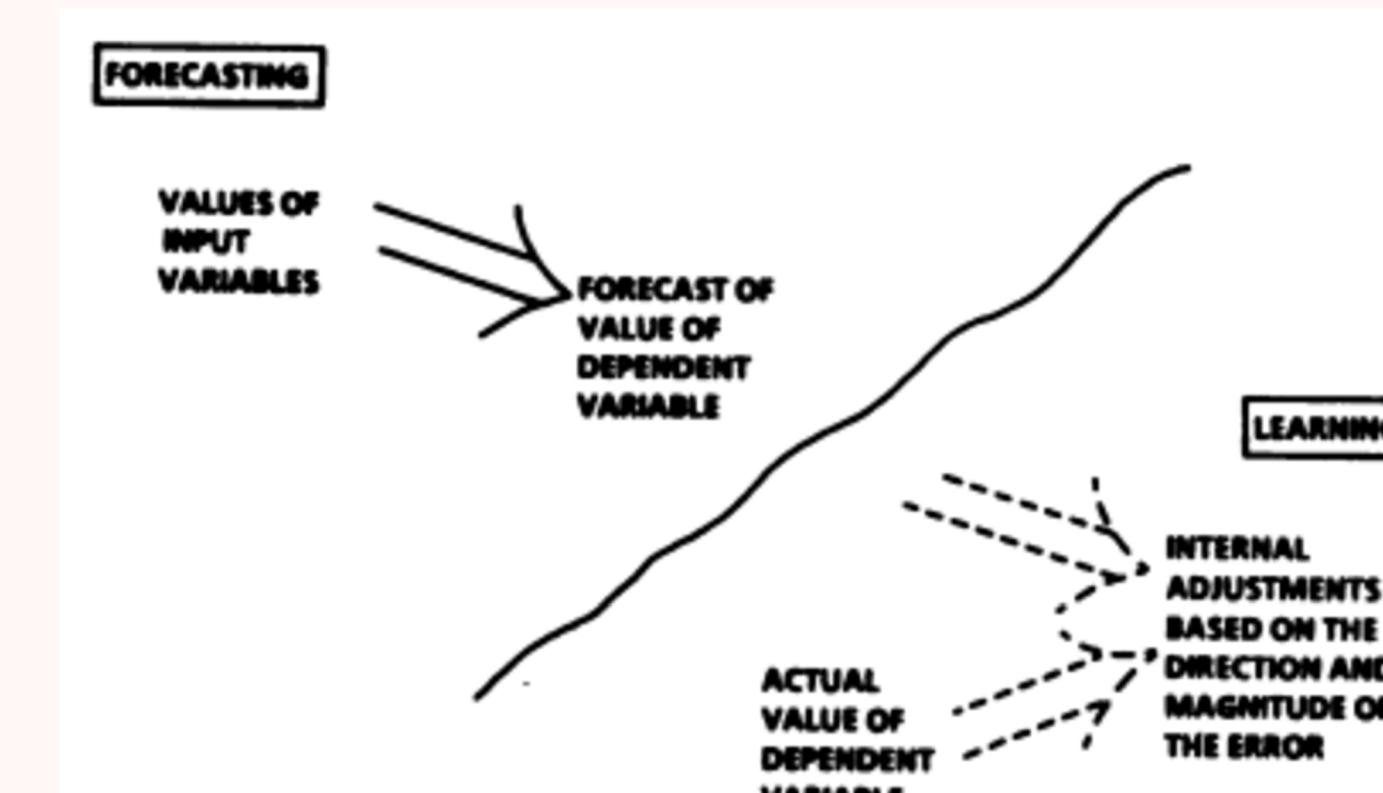
## So sánh chỉ số giữa hai nhóm

# Ma trận tương quan (Correlation Heatmap)

- "Glucose có độ tương quan dương cao nhất (0.49) với Outcome, xác nhận đây là yếu tố dự báo quan trọng nhất."
- "BMI (0.31) và Tuổi (0.24) cũng cho thấy mối tương quan đáng kể."
- "Các biến như Tuổi và Số lần mang thai có tương quan mạnh với nhau (0.54)."

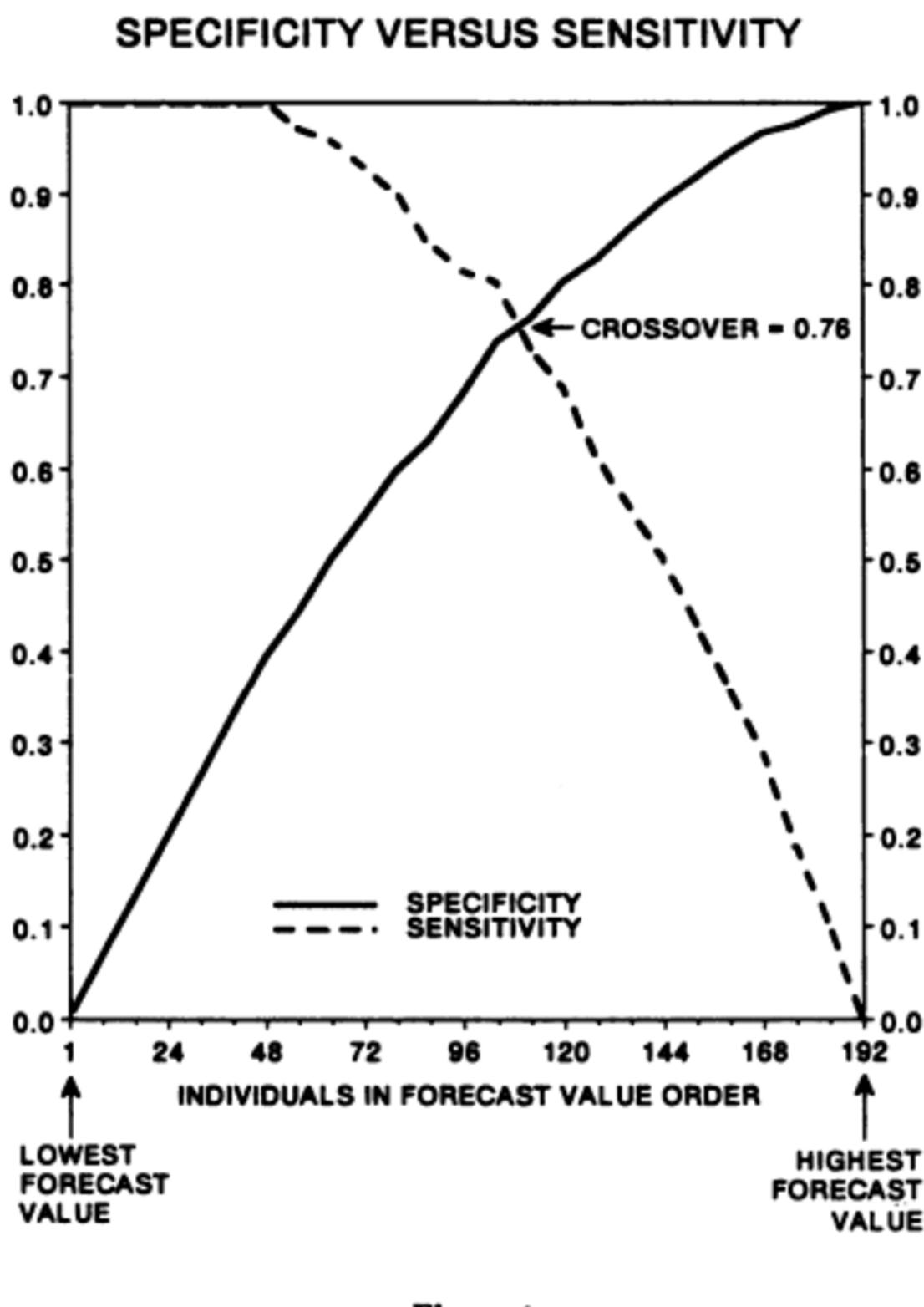


# Phương pháp luận - Thuật toán ADAP



- ADAP là một quy trình học thích ứng, tạo ra các "analogues" giống perceptron.
- Cơ chế học: ADAP học bằng cách thực hiện các điều chỉnh nội bộ khi dự báo sai.
- Kiến trúc:
  - Sensors (Cảm biến): Đại diện cho một giá trị hoặc một khoảng giá trị của biến đầu vào.
  - Partitions (Phân vùng): Tập hợp các sensor cho một biến.
  - Association Units (Đơn vị liên kết): Kết nối các sensor thông qua các trọng số có thể điều chỉnh.
  - Responder (Đơn vị phản hồi): Tổng hợp tín hiệu từ các đơn vị liên kết để đưa ra dự báo cuối cùng.

# Kết quả thực nghiệm (Results)



- Tập dữ liệu: Được chia thành 576 mẫu để huấn luyện (training) và 192 mẫu để dự báo (forecasting).
- Ngưỡng phân loại: Giá trị 0.448 được chọn để phân biệt giữa hai lớp.
- Hiệu suất mô hình:
  - Độ nhạy (Sensitivity): 76%
  - Độ đặc hiệu (Specificity): 76%
  - Điểm giao nhau (Crossover point) là 0.76.
- Đường cong ROC (Receiver Operating Characteristic): Được sử dụng để đánh giá hiệu suất của mô hình ở các ngưỡng khác nhau.

## Kết luận và Hướng phát triển (Conclusion & Future Work)

- Kết luận:
  - Thuật toán mạng nơ-ron ADAP có khả năng dự báo sự khởi phát của bệnh tiểu đường với độ chính xác khá tốt (76%).
  - EDA cho thấy các yếu tố như Glucose, BMI, và Tuổi là những chỉ báo quan trọng, phù hợp với kết quả của mô hình.
- Hướng phát triển (Future Work):
  - Tiếp tục kiểm tra và so sánh kết quả của ADAP với các mô hình thống kê khác như Hồi quy Logistic (Logistic Regression) và Perceptron tuyến tính (Linear Perceptron) trên cùng một tập dữ liệu.

*Thank  
You*

@reallygreatsite