# Tóm tắt: Dự đoán bệnh tiểu đường bằng thuật toán FreeSpan

## 1. Giới thiệu

- Mục tiêu: Phát hiện các mẫu tuần tự trong dữ liệu bệnh nhân tiểu đường để hỗ trợ dự báo nguy cơ mắc bệnh.  
- Thuật toán sử dụng: FreeSpan, Apriori, Decision Tree.  
- Công cụ: Python (pandas, NumPy, scikit-learn, Matplotlib, Seaborn).

## 2. Tiền xử lý dữ liệu

- Lọc bỏ dữ liệu không hợp lệ, chuẩn hóa tuổi, BMI, HbA1c, mức đường huyết theo ngưỡng y khoa.  
- Chuyển dữ liệu số sang dạng danh mục (categorical mapping).  
- Kiểm tra lại dữ liệu sau xử lý để tránh mất thông tin quan trọng.

## 3. Phân tích dữ liệu với FreeSpan & Apriori

- FreeSpan: Tìm ra các mẫu tuần tự phổ biến trong dữ liệu bệnh nhân.  
- Apriori: Xác định các luật kết hợp giữa các yếu tố nguy cơ.  
- Trực quan hóa kết quả: Biểu đồ tần suất mẫu phổ biến, ma trận lift đánh giá mối quan hệ giữa các yếu tố.

## 4. So sánh với các thuật toán khác

|  |  |  |
| --- | --- | --- |
| Thuật toán | Ưu điểm | Nhược điểm |
| FreeSpan | Tìm mẫu tuần tự hiệu quả | Chậm với dữ liệu lớn |
| FP-Growth | Xử lý nhanh hơn FreeSpan | Phức tạp khi triển khai |
| LSTM (Deep Learning) | Dự báo mẫu dựa trên dữ liệu lớn | Cần nhiều tài nguyên tính toán |
| Random Forest | Phân loại tốt với dữ liệu y tế | Không tối ưu cho dữ liệu tuần tự |

## 5. Kết luận & Ứng dụng

- Mô hình giúp xác định bệnh nhân có nguy cơ cao mắc tiểu đường.  
- Có thể áp dụng vào hệ thống hỗ trợ quyết định y khoa, dự báo bệnh sớm.  
- Mở rộng:  
 Kết hợp với Machine Learning (LSTM) để cải thiện độ chính xác.  
 Ứng dụng thuật toán Random Forest để phân loại bệnh nhân.  
 Thử nghiệm trên tập dữ liệu lớn hơn để kiểm tra độ ổn định.