TRƯỜNG ĐẠI HỌC DUY TÂN

KHOA SAU ĐẠI HỌC

*Tiểu luận môn*

**AN TOÀN VÀ BẢO MẬT THÔNG TIN**

**CÁCH TẤN CÔNG VÀ PHÒNG CHỐNG**

**LỖ HỔNG SQL INJECTION**

Hướng dẫn : PGS.TS Lê Đắc Nhường

Thực hiện : Phạm Minh Tuấn

Võ Đình Hiếu

Nguyễn Anh Quân

Lớp : K22MCS (Khoa học máy tính)

Đà nẵng, 06/2021

**MỤC LỤC**

[LỜI MỞ ĐẦU 1](#_Toc66660841)

[CHƯƠNG 1:](#_Toc66660842) [CÂY QUYẾT ĐỊNH VÀ THUẬT TOÁN C4.5 3](#_Toc66660843)

[1.1. KHÁI NIỆM CÂY QUYẾT ĐỊNH 3](#_Toc66660844)

[1.2. CÁC LOẠI CÂY QUYẾT ĐỊNH 3](#_Toc66660845)

[1.3. TẠO CÂY QUYẾT ĐỊNH 4](#_Toc66660846)

[1.4. SỬ DỤNG CÂY QUYẾT ĐỊNH 6](#_Toc66660847)

[1.5. DUYỆT CÂY VÀ PHÂN LỚP DỮ LIỆU 7](#_Toc66660848)

[1.5.1. Lựa chọn tiêu chuẩn phân lớp 7](#_Toc66660849)

[1.5.2. Điều kiện để dừng việc phân chia 7](#_Toc66660850)

[1.5.3. Độ lợi thông tin (Information Gain) 7](#_Toc66660851)

[1.5.4. Vấn đề quá khớp trong phân lớp dữ liệu 10](#_Toc66660852)

[1.5.5. Cây quyết định với cơ sở dữ liệu lớn 11](#_Toc66660853)

[1.6. THUẬT TOÁN DUYỆT CÂY C4.5 11](#_Toc66660854)

[CHƯƠNG 2:](#_Toc66660855) [DỮ LIỆU BỆNH TIỂU ĐƯỜNG VÀ THUẬT TOÁN C4.5 14](#_Toc66660856)

[2.1. GIỚI THIỆU BỆNH TIỂU ĐƯỜNG 14](#_Toc66660857)

[2.2. THÔNG TIN DỮ LIỆU 14](#_Toc66660858)

[2.3. MÔ TẢ THUỘC TÍNH 15](#_Toc66660859)

[2.4. CÀI ĐẶT THUẬT TOÁN TRÊN TẬP DỮ LIỆU 16](#_Toc66660860)

[2.5. DANH SÁCH LUẬT 37](#_Toc66660861)

[CHƯƠNG 3:](#_Toc66660862) [ỨNG DỤNG THỰC TẾ 42](#_Toc66660863)

[3.1. GIỚI THIỆU ỨNG DỤNG 42](#_Toc66660864)

[3.2. TRUY CẬP ỨNG DỤNG 42](#_Toc66660865)

[3.3. GIAO DIỆN VÀ TÍNH NĂNG 42](#_Toc66660866)

[Chương 4:](#_Toc66660867) [KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 47](#_Toc66660868)

[TÀI LIỆU THAM KHẢO 48](#_Toc66660869)

# 

**DANH MỤC HÌNH VẼ**

|  |  |  |
| --- | --- | --- |
| **Ký hiệu** | **Nội dung** | **Trang** |
| Hình 1.1 | Cấu trúc cây quyết định | 3 |
| Hình 1.2 | Cây quyết định dữ liệu mua máy tính | 6 |
| Hình 2.1 | Cây quyết định tại thuộc tính Polyuria | 28 |
| Hình 2.2 | Cây quyết định của một phần dữ liệu mẫu | 37 |
| Hình 2.3 | Cây quyết định dữ liệu bệnh tiểu đường | 41 |
| Hình 3.1 | Cách truy cập ứng dụng Hỗ trợ chuẩn đoán bệnh tiểu đường | 42 |
| Hình 3.2 | Giao diện trang chủ website | 43 |
| Hình 3.3 | Giao diện trang kết quả nếu không mắc bệnh | 44 |
| Hình 3.4 | Giao diện trang kết quả người dùng mắc bệnh | 45 |
| Hình 3.5 | Giao diện trang giới thiệu | 46 |

**DANH MỤC BẢNG**

|  |  |  |
| --- | --- | --- |
| **Ký hiệu** | **Nội dung** | **Trang** |
| Bảng 1.1 | Dữ liệu mua máy tính | 5 |
| Bảng 2.1 | Các thuộc tính và giá trị của dữ liệu bệnh tiểu đường | 16 |
| Bảng 2.2 | Dữ liệu mẫu bệnh tiểu đường | 17 |

# LỜI MỞ ĐẦU

1. **Lý do chọn đề tài**

Ngày nay, sự phát triển vượt bậc của công nghệ đã đem lại rất nhiều thuận lợi cho người sử dụng cũng như các nhà phát triển. Nhưng cùng với sự phát triển này thì các ứng dụng và dữ liệu cũng trở thành mục tiêu ưu thích của những kẻ tấn công. Các hình thức tấn công rất đa dạng như thay đổi nội dung, tấn công từ chối dịch vụ hoặc chiếm quyền điều khiển… Mục đích của các hacker khác nhau, có thể xuất phát từ thiện chí muốn tấn công để tìm ra những điểm yếu và cảnh báo nhưng nghiêm trọng hơn là tấn công để phục vụ cho các mục đích xấu như thay đổi, thêm bớt, bán dữ liệu cho các bên tội phạm khác, sử dụng thông tin có được cho các hình thức tấn công xa hơn, hoặc chỉ đơn giản xóa toàn bộ cơ sở dữ liệu để khiến hệ thống không còn hoạt động chính xác nữa.

Các hình thức tấn công vô cùng đa dạng nhưng có thể chia làm 2 loại chính là tìm cách truy cập vào tài khoản quản trị (Access Control) hoặc tấn công vào lỗ hổng bảo mật trên phần mềm. Tiêu biểu nhất cho hình thức tấn công thứ 2 là tấn công bằng cách khai thác lỗ hổng SQL Injection (SQLi).

Mức độ nguy hiểm của lỗ hổng SQL Injection trong quá khứ có thể kể đến các vụ bị rò rĩ dữ liệu nổi tiếng như: Sony, Microsoft UK, Vietnamwork, Visa International và MasterCard International đã để lại hậu quả cực kỳ nghiêm trọng. Vì lý do đó, Inject (Không chỉ SQL mà OS và LDAP) nằm chễm chệ ở vị trí đầu bảng trong top 10 lỗ hỗng bảo mật của OWASP.

Ngày nay các kiểu tấn công của các hacker cũng ngày càng tinh vi, phức tạp và khó ngăn chặn. Hậu quả để lại thì vô cùng nặng nề nên xuất phát từ những lý do đó, chúng em chọn đề tài “Cách tấn công và phòng chống lỗ hổng SQL Injection” làm đề tài nghiên cứu nhằm phổ biến kiến thức về phòng chống khi xây dựng các ứng dụng.

1. **Mục tiêu nghiên cứu**

Giúp hiểu hơn về các mối đe dọa an toàn thông tin trên môi trường mạng và trong các ứng dụng.

Hiểu rõ hơn khái niệm SQL Injection và phương thức hoạt động của các hacker thông qua lỗ hổng này nhằm đưa ra các giải pháp phòng chống.

Biết sử dụng kiến thức và công cụ cơ bản để kiểm tra bảo mật trên các ứng dụng đã được triển khai để khuyến cáo hoặc bịt lại các lỗ hổng kịp thời.

1. **Phương pháp nghiên cứu**

Tìm hiểu về ngôn ngữ truy vấn có cấu trúc SQL.

Nghiên cứu các kỹ thuật tấn công thông qua các lỗ hổng bảo mật SQL Injection từ đó đưa ra các giải pháp phòng chống có hiệu quả.

1. **Bố cục tiểu luận**

Nội dung của bài tiểu luận được trình bày với bố cục gồm 04 chương như sau:

Chương 1: Giới thiệu về lỗ hổng SQL Injection.

Chương 2: Cách tấn công bằng cách khai thác lỗ hổng SQL Injection.

Chương 3: Cách phòng chống tấn công SQL Injection.

Chương 4: Xây dựng ứng dụng demo một vài cách tấn công SQL Injection.

# Chương 1

# GIỚI THIỆU VỀ LỖ HỔNG SQL INJECTION

* 1. **ĐẶC TRƯNG ỨNG DỤNG CƠ SỞ DỮ LIỆU**

Ngày nay cơ sở dữ liệu hầu như hiện diện ở tất cả các ứng dụng có liên quan đến công nghệ thông tin. Cơ sở dữ liệu là hệ thống thông tin có cấu trúc, có tính nhất quán được lưu trữ tại các thiết bị lưu trữ, nhằm phục vụ nhu cầu khai thác và sử dụng cho nhiều người, nhiều chương trình khác nhau. Cơ sở dữ liệu có thể truy cập, chỉnh sửa, bổ sung, truy xuất tùy theo mục đích sử dụng khác nhau.

Ngày nay hầu hết các ứng dụng đều được phát triển theo mô hình 3 lớp. Lớp đầu tiên lưu cơ sở dữ liệu, lớp thứ 2 là giao diện để người dùng thao tác và lớp còn lại là cầu nối xử lý thông tin người dùng đưa vào trước khi lưu trữ vào cơ sở dữ liệu.

* 1. **TỔNG QUAN SQL INJECTION**
     1. **Khái niệm:**

SQL Injection là một kỹ thuật cho phép những kẻ tấn công lợi dụng lỗ hổng của việc kiểm tra dữ liệu đầu vào trong các ứng dụng và các thông báo lỗi của hệ quản trị cơ sở dữ liệu trả về để inject (tiêm vào) và thi hành các câu lệnh SQL bất hợp pháp nhằm chiếm đoạt thông tin hoặc phá huỷ dữ liệu của hệ thống.

* + 1. **Nguyên nhân:**

Dữ liệu đầu vào từ người dùng hoặc từ các nguồn khác không được kiểm tra hoặc kiểm tra không kỹ lưỡng.

Ứng dụng sử dụng các câu lệnh SQL động, trong đó dữ liệu được kết nối với mã SQL gốc để tạo câu lệnh SQL hoàn chỉnh.

* + 1. **Mức độ nguy hiểm**

Tùy vào mức độ tinh vi, SQL Injection có thể cho phép kẻ tấn công:

Vượt qua các khâu xác thực người dùng

Chèn, xóa hoặc sửa đổi dữ liệu

Đánh cắp các thông tin trong CSDL

Chiếm quyền điều khiển hệ thống

* + 1. **Phân loại SQL Injection**

Tùy vào mức độ tinh vi, SQL Injection có thể cho phép kẻ tấn công:

# Chương 2

# DỮ LIỆU BỆNH TIỂU ĐƯỜNG VÀ THUẬT TOÁN C4.5

* 1. **GIỚI THIỆU BỆNH TIỂU ĐƯỜNG**

Theo báo cáo của Tổ chức Y tế Thế giới (WHO), bệnh tiểu đường là một trong những căn bệnh mãn tính đe dọa tính mạng phát triển nhanh nhất, đã ảnh hưởng đến 422 triệu người trên toàn thế giới, theo báo cáo của Tổ chức Y tế Thế giới (WHO), vào năm 2018. Do sự hiện diện của giai đoạn không có triệu chứng tương đối dài, nên việc phát hiện sớm bệnh tiểu đường là luôn mong muốn cho một kết quả có ý nghĩa về mặt lâm sàng. Khoảng 50% tất cả những người mắc bệnh tiểu đường không được chẩn đoán vì giai đoạn không có triệu chứng kéo dài của nó.

* 1. **THÔNG TIN DỮ LIỆU**

Dữ liệu được lấy từ website:

* 1. **MÔ TẢ THUỘC TÍNH**

Đặc điểm của tập dữ liệu: Đa biến

Số lượng bản ghi: 520

Số thuộc tính: 17

# Chương 3

# ỨNG DỤNG THỰC TẾ

* 1. **GIỚI THIỆU ỨNG DỤNG**

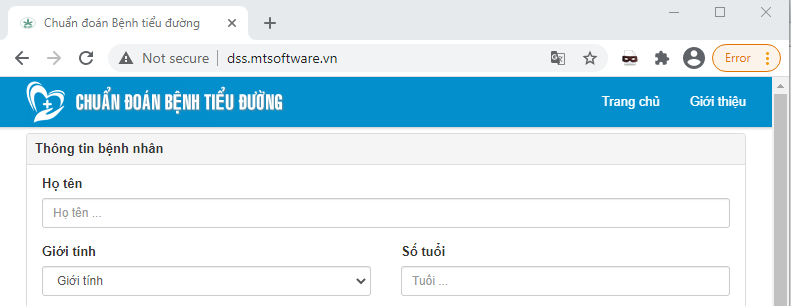
Chương trình “Hỗ trợ chuẩn đoán bệnh tiểu đường” được phát triển dạng website để tăng độ hiệu quả khi triển khai và dễ dàng nâng cấp ứng dụng khi có như cầu.

Website phát triển trên nền tảng Microsoft .Net Framework 4.0, sử dụng ngôn ngữ C#, JavaScript để lập trình. Về mặt giao diện sử dụng Bootstrap 3.7 và các thành phần phụ thuộc khác để triển khai ứng dụng.

* 1. **TRUY CẬP ỨNG DỤNG**

Ứng dụng đã được triển khai trên internet nên để truy cập trên máy tính hoặc điện thoại chỉ cần mở trình duyệt và truy cập đường dẫn sau để vào ứng dụng.

<http://dss.mtsoftware.vn/>



*Hình 3.1: Cách truy cập ứng dụng Hỗ trợ chuẩn đoán bệnh tiểu đường*

* 1. **GIAO DIỆN VÀ TÍNH NĂNG**

Để thực hiện tra cứu có mắc bệnh tiểu đường hay không, sau khi truy cập vào website sẽ hiển thị giao diện trang chủ như hình sau (Hình 3.2):



*Hình 3.2: Giao diện trang chủ website*

# Chương 4

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bệnh tiểu đường rất nguy hiểm nên việc có một ứng dụng để phổ cập và hỗ trợ chuẩn đoán là vô cùng thiết thực. Việc thực hiện đề tài mang ý nghĩa nhằm giúp cho người sử dụng hiểu biết hơn về bệnh tiểu đường, tiết kiệm được nhiều thời gian cũng như tự đánh giá được tình trạng sức khỏe để đi khám và điều trị kịp thời.

Thông qua quá trình nghiên cứu về mô hình cây quyết định và kiến thức môn học Hệ hỗ trợ ra quyết định, tiểu luận đã tiến hành giải quyết bài toán thực tế về hỗ trợ chuẩn đoán bệnh tiểu đường. Cụ thể, tiểu luận đã đi sâu nghiên cứu và làm rõ những nội dung sau:

Đưa ra cơ sở lý thuyết về mô hình cây quyết định và thuật toán C4.5 để ứng dụng vào việc phân tích dữ liệu để chuẩn đoán bệnh.

Phát triển website dựa trên tập luật sinh ra từ dữ liệu được tổng hợp và khai thác tại trang web <https://archive.ics.uci.edu/ml/index.php> để triển khai và vận hành thực tế.

Tiểu luận đã cho thấy sự hữu ích của việc phân tích dữ liệu để áp dụng, giải quyết các bài toán thực tế về bệnh tiểu đường. Tuy nhiên, do một số nguyên nhân khách quan và chủ quan, tiểu luận vẫn còn tồn tại một số hạn chế sau:

Dữ liệu thu thập được còn ít nên công tác dự báo mới chỉ dừng lại ở phạm vi hỗ trợ, nhiều trường hợp còn sai số.

Chưa tìm hiểu hết tất cả các thuật toán về cây quyết định để áp dụng linh hoạt tùy theo từng trường hợp dữ liệu để tăng độ chính xác kết quả trả về.

Để khắc phục những hạn chế nêu trên, trong thời gian tới, hướng nghiên cứu sẽ tiếp tục mở rộng phạm vi thu thập dữ liệu, nghiên cứu sâu hơn về các thuật toán và các công cụ hỗ trợ khác để tiến hành nâng cấp ứng dụng và triển khai rộng rãi hơn.

# TÀI LIỆU THAM KHẢO

1. Giáo trình Hệ hỗ trợ ra quyết định – PGS.TS Lê Đắc Nhường
2. Damanik, I. S., Windarto, A. P., Wanto, A., Andani, S. R., & Saputra, W. (2019). Decision Tree Optimization in C4. 5 Algorithm Using Genetic Algorithm. Paper presented at the Journal of Physics: Conference Series.
3. Lakshmi, B., Indumathi, T., & Ravi, N. (2016). A Study on C. 5 decision tree classification algorithm for risk predictions during pregnancy. Procedia Technology, 24, 1542-1549.
4. Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily, H., . . . Ghayour-Mobarhan, M. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree
5. Trương, T. Q. (2018). Hướng xây dựng cây quyết định với chi phí hiệu quả. Trường Đại học Bách khoa-Đại học Đà Nẵng,
6. Hoan, Nguyen Quang, et al. "MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION, PREDICTION." UTEHY Journal of Science and Technology 17 (2018): 62-66.
7. <https://archive.ics.uci.edu/ml/index.php>