**KỸ THUẬT PHÂN TÍCH DỮ LIỆU BẢO TOÀN TÍNH RIÊNG TƯ CHO**

**KHẢO SÁT THÔNG TIN BỆNH TRẦM CẢM**

NGUYỄN HỒNG TUYẾT QUỲNH 1 (\*), NGUYỄN HOÀNG ÁI LINH1, NGUYỄN THỊ VIÊN1,

NGUYỄN MINH PHÚC1, LÊ HÀ TÚ MY1, NGUYỄN QUANG MẠNH 1,

HUỲNH THỊ KIM QUỲNH2, LÊ PHÚC LỮ1, NGUYỄN HỮU TÌNH1

*1Khoa công nghệ Thông tin, trường Đại học Công nghiệp Thành phố Hồ Chí Minh*

*2Khoa công nghệ Thông tin, trường Đại học Việt Đức*

*quiana2k4@gmail.com, hoangailinh11062004@gmail.com, 22003025.vien@student.iuh.edu.vn, csephuc@gmail.com, lehatmy@gmail.com, nguyenquangmanh160304@gmail.com, kimquynh15@gmail.com*, [*lephuclu@iuh.edu.vn*](mailto:lephuclu@iuh.edu.vn)*,* [*nguyenhuutinh@iuh.edu.vn*](mailto:nguyenhuutinh@iuh.edu.vn)

**Tóm tắt:** Bệnh trầm cảm là một vấn đề nhức nhối của xã hội hiện đại, gây hại nhiều đến sức khỏe và chiếm tỷ lệ cao trong dân số. Xu hướng trẻ hóa của căn bệnh này đến từ việc nó có xuất hiện nhiều trong lứa tuổi học sinh, sinh viên và nguyên nhân chính là do các căng thẳng trong công việc, áp lực học tập, chuyện gia đình, tình cảm, ... Việc phát hiện, khoanh vùng sớm thông qua các các câu hỏi khảo sát cá nhân sẽ hỗ trợ, giúp chữa trị tốt, kịp thời và tránh được điều đáng tiếc xảy ra. Tuy nhiên, một điểm khó là các bệnh nhân thường ngại chia sẻ thông tin hoặc trả lời sai lệch các câu hỏi có phần nhạy cảm, dẫn đến việc khảo sát không như mong muốn. Vì thế, nghiên cứu này sẽ tìm hiểu các kỹ thuật phân tích dữ liệu bảo toàn tính riêng tư, từ đó ứng dụng vào việc thêm nhiễu vào câu trả lời cho bảng câu hỏi khảo sát, giúp người trả lời thoải mái hơn trong việc phản hồi, chịu chia sẻ thông tin mà dữ liệu cá nhân vẫn được đảm bảo. Đề tài cũng dùng các phương pháp thống kê để đánh giá tính chính xác và so sánh với mức độ riêng tư đạt được để chọn ra các tham số cho mô hình một cách phù hợp.

**Từ khóa:** bệnh trầm cảm, an toàn thông tin, phân tích dữ liệu bảo toàn tính riêng tư, bảng câu hỏi, phân phối Laplace, phân phối chuẩn, ma trận nhầm lẫn.

**PRIVACY PRESERVING DATA ANALYSIS TECHNIQUES FOR**

**DEPRESSION INFORMATION SURVEY**

**Abstract:** Depression is a painful problem in modern society, causing great harm to health and accounting for a high proportion of the population. The rejuvenation trend of this disease comes from the fact that it appears a lot in the school age group, and the main cause is due to work stress, study pressure, family issues, love affairs, etc. Early detection and localization through personal survey questions will support and help with good and timely treatment, avoid unfortunate events from happening. However, a difficult point is that patients are often afraid to share information or answer sensitive questions incorrectly, leading to the survey not being as expected. Therefore, this study will explore privacy-preserving data analysis techniques, thereby applying noise to survey questionnaire answers, helping respondents to be more comfortable in answering questions, respond and share information while personal data is still guaranteed. We also use statistical methods to balance the trade-off between accuracy and the achieved level of privacy then select appropriate parameters for the model.

**Keyword:** depression, information security, privacy-preserving data analysis, questionnaire, Laplace distribution, normal distribution, confusion matrix.

**Lĩnh vực:** Điện – Điện tử – Công nghệ thông tin.