[**I.** **DATA MINING IN HEALCARE: CURRENT APPLICATIONS AND ISSUES** (By Ruben D. Canlas Jr., MSIT, MBA 5 August 2009) 2](#_Toc336457043)

[1. Trích dẫn: 2](#_Toc336457044)

[2. Giới thiệu và các nhân tố cơ bản: 2](#_Toc336457045)

[3. Phương pháp luận: 2](#_Toc336457046)

[4. Các nghiên cứu: 2](#_Toc336457047)

[5. Các vấn đề và thử thách: 4](#_Toc336457048)

[6. Kết luận và khuyến nghị: 4](#_Toc336457049)

[**II.** **NEXT** 4](#_Toc336457050)

1. **DATA MINING IN HEALCARE: CURRENT APPLICATIONS AND ISSUES** (By Ruben D. Canlas Jr., MSIT, MBA 5 August 2009)
2. Trích dẫn:

* Với sự thành công của các ứng dụng được tích hợp Data mining trong các lĩnh vực rõ rệt như thương mại điện tử, marketing và bán lẻ đã cho ta thấy tính phổ biến của việc sử dụng những tri thức được lấy từ cơ sở dữ liệu ( Knowledge Discovery in Database – KDD) trong tất cả các lĩnh vực. Một trong các lĩnh vực mới nhất vừa được áp dụng Data mining đó chính là y học và sức khỏe cộng đồng.
* Bài báo ngày cho thấy một cuộc khảo sát các kĩ thuật hiện tại của KDD, sử dụng các công cụ đã tích hợp Data mining trong y tế và sức khỏa công đồng. Ngoài ra còn bao gồm các vấn đề quan trọng và thử thách trong việc kết hợp Data mining với y tế nói chung.
* Bài nghiên cứu còn cho thấy sự phát triển của một lượng lớn các ứng dụng có Data mining, bao gồm các ứng dụng như phân tích dữ liệu tại các bệnh viện để đưa ra các chính sách hoạch định sức khỏe tốt hơn, phát hiện dịch bệnh, ngăn ngừa các trường hợp tử vong trong bệnh viện và phát hiện các gian lận trong quyền bản hiểm.

1. Giới thiệu và các nhân tố cơ bản:

* Mục tiêu:
* Liệt kê các lợi ích và nổi bật tầm quan trọng của Data mining trong y học vào sức khỏe cộng đồng.
* Tìm kiếm các kỹ thuật Data mining đã được sử dụng trong các lĩnh vực khác có thể áp dụng cho sức khỏe cộng đồng.
* Xác định các vấn đề và thử thách trong ứng dụng Data mining vào y khoa.
* Vạch ra những khuyến nghị cho việc thu thập các tri thức trong các cơ sở dữ liệu điện tử thông qua Data mining

1. Phương pháp luận:

* Dựa theo các phương pháp bị hạn chế, phương pháp luận chính được sử dụng trong bài nguyên cứu này thông qua các bài đã được xuất bản trong nhiều lĩnh vực của y học, khoa học và kỹ thuật máy tính. Bài nghiên cứu dựa trên những xuất bản gần đây nhất, từ năm 2000 trở về đây.

1. Các nghiên cứu:
   1. Data mining trong phát hiện dịch bệnh:

* Sử dụng dữ liệu thiết thực để hỗ trợ cho việc ra quyết định trong y học ( còn được biết là EBM – Evidence Based Medicine) đã tồn tại trong nhiều thế kỷ. John Snow, đã được biết đến là cha đẻ của dịch tễ học hiện đại, đã sử dụng các bản đồ đồ thị vào năm 1854 để phát hiện ra nguồn bệnh thổ tả và đã chứng minh rằng bệnh này lây lan qua các hệ thống cấp nước.
* Snow đếm số lượng các bệnh nhân và vẽ sơ đồ địa chỉ của các bệnh nhân trên bản đồng là các thanh màu đen. Ông đã phát hiện rằng hầu hết các ca tử vong đều nằm xung quanh một điểm bơm nước nhất định tại London.
* Florence Nightngale đã phát minh ra biểu đồ vùng phân cực vào năm 1855 để chi ra nhiều ca tử vong của binh lính là do sự mất vệ sinh trong khám lâm sàn và có thể ngăn ngừa được. Bà đã sử dụng những biểu đồ để thuyết phục các chinh sách thực hiện cải cách và đã giảm được số lượng các ca tử vong.
* Snow và Nightingale đã chính mình thực hiện thu thập dữ liệu, sang lọc và phân tích thông qua các dữ liệu về tỉ lệ tử vong trong suốt thời gian nghiên cứu vì số lượng bộ thông tin có thể quản lý được. Ngày nay, số lượng dân số, dữ liệu điện tử thu thập được cùng với sự toàn cầu hóa và tốc độ phát bệnh của các dịch bệnh làm cho việc thao tác dữ liệu gần như hoàn toàn không thể hoàn thành.
* Đó là nguyên nhân tại sao Data mining trở nên hữu ích trong việc chăm sóc sức khỏe. Tuy phát triển khác chậm nhưng vẫn được áp dụng để giải quyết các vấn đề khác nhau của việc thu thập tri thức trong lĩnh vực này.
* Data mining và các ứng dụng của nó trong y học và sức khỏe cộng đồng là một lĩnh vực học tập còn non trẻ. Vào năm 2003, Wilson et al bắt đầu thăm dò các trường hợp mà KDD và kỹ thuật Data mining đã được áp dụng trong cơ sở dữ liệu y tế. Người ta cũng đã phát hiện nhiều sự nhầm lẫn trong các lĩnh vực được tạo thành từ Data mining. “Một vài tác giả cho rằng Data mining chỉ là một quá trình thu thập, trong khi các tác giả khác cho rằng Data mining sử dụng kỹ thuật thống kế trong quá trình thu thập trị thức.” (Wilson et al 2003)
* Vì những quan niệm sai lầm vẫn tiếp tục trong cộng đồng y khoa về những gì mà Data mining chứa đựng, chúng ta sẽ định nghĩa lại những quan niệm đó. Một định nghĩa được chấp nhận nhiều nhất về Data mining ngày nay là một bộ các quy trình và kỹ thuật trong khai phá và mô tả kiểu và hướng trong dữ liệu (Witten and Frank 2005). Chúng ta sẽ sử dụng định nghĩa này xuyên suốt trong bài cáo
  1. Tầm quan trọng và lợi ích của Data mining trong y học và sức khỏe cộng đồng:
* Mặc dù sự khác biệt và xung đột trong hướng tiếp cận, lĩnh vực y tế hiện nay cần nhiều Data mining hơn. Có 1 vài đối số có thể được nâng cao để hỗ trợ trong việc sử dụng Data mining trong lĩnh việc y tế. Không chỉ trong y tế cộng đồ mà còn trong lĩnh vực y tế tư nhân.
* Tràn dữ liệu (*Data overload*). Nhiều tri thức có giá trị được thu bởi việc tin họa hóa các dữ liệu y tế. Tuy nhiên, một lượng lớn dữ liệu được lưu trữ sẽ làm cho việc này trở nên khó khăn và gần như là không thể nếu người ta phải sàng lọc và khai phá tri thức (Cheng, et al 2006).
* Trong thực tế, một số chuyên gia tin rằng những đột phá trong lĩnh vực y tế đã chậm lại, thêm vào đó là độ phức tạp của các thông tin y tế hiện nay. Máy tính và Data mining là thích hợp nhất cho mục đích trên. (Shillabeer and Roddick 2007).
* Y học dựa trên thông tin và ngăn chặn các sai sót trong bệnh viên. Khi các tổ chức y tế ứng dụng Data mining vào dữ liệu, họ có thể khai phá được các tri thức hữu ích và khả năng cứu sống nhiều mạng sống vẫn còn nằm trong cơ sở dữ liệu của họ. Ví dụ, một nghiên cứu đang được thực hiện trong các bệnh viên đã cho thấy rằng 87% ca tử vong trong các bệnh viên tại Mỹ có thể ngăn ngừa nếu các nhân viên bệnh viện (kể cả bác sĩ) cẩn thận hơn trong ngăn ngừa sai sót ( HeathGrades Hospitals Study 2007). Bằng việc khai phá các dữ liệu bệnh viên, các vấn đề về an toàn có thể được phát hiện và giải quyết bởi các quản lý bệnh viện và chính phủ.
* Hoạch định chính sách trong y tế cộng đồng (*Policy-making in public heath*). Lavrac et al. (2007)đã kết hợp GIS và Data mining với nhau, Weka với J48, để phân tích điểm giống nhau giữa các trung tâm y tế tại Slovenia. Sự dụng Data mining, họ có thể khai phá được các mô hình trong trung tâm y tế để có thể đưa ra các chính sách khuyến nghị đến Viện Y Tế Cộng Đồng. Họ đã kết luận rằng “Data mining và hệ thống hỗ trợ ra quyết định, kể cả các phương pháp trực quan mới, có thể nâng cao hệ thống ra quyết định.
* Các yếu tố trên đã nhắc chúng ta nhớ tới một sự cố tại trung tâm y tế Rizal của Pasig City của Indonexia vào tháng 10 năm 2006. Thất bại trong việc thực hiện vệ sinh nghiêm ngặt và các biện pháp khử trùng trong bệnh viên đã góp phần dẫn đến các ca tử vong của trẻ so sinh do bị nhiểm khuẩn. Không ai có thể biết được chính xác chuyện gì đang xảy ra cho tới khi các ca tư vong trở nên thường xuyên. Sau khi kiểm tra dữ liệu của bệnh viên, bộ phận y tế ( Department of Health – DOH) đã tìm thấy 12 trong 28 trẻ được sinh ra vào ngày 4 tháng 10 đã tử vong vì nhiểm khuẩn (Tandoc 2006). Với cơ sở dữ liệu đã được xử lý và một ứng dụng Data mining, DOH có thể các hiện các sự kiện không bình thường và hạn chế chúng diễn biến tệ hơn.

1. Các vấn đề và thử thách:
2. Kết luận và khuyến nghị:
3. **NEXT**