**BÁO CÁO GIAI ĐOẠN 1**

*Thời gian dự kiến 09/09/2012 – 09/10/2012*

**Nội dung công việc**

[**1.** **ỨNG DỤNG CỦA KHAI PHÁ DỮ LIỆU VÀO Y HỌC:** 2](#_Toc337025602)

[1.1. Mục tiêu: 2](#_Toc337025603)

[1.2. Các nghiên cứu và ứng dụng: 2](#_Toc337025604)

[1.3. Next: 2](#_Toc337025605)

[**2.** **BỆNH TIỂU ĐƯỜNG:** 2](#_Toc337025606)

[**3.** **CÁC THUẬT TOÁN ĐÃ ĐƯỢC ỨNG DỤNG VÀO KHAI PHÁ DỮ LIỆU CỦA BỆNH NHÂN MẮC BỆNH TIỂU ĐƯỜNG:** 2](#_Toc337025607)

1. **ỨNG DỤNG CỦA KHAI PHÁ DỮ LIỆU VÀO Y HỌC:**

Với những thành công của việc kết hợp khai phá dữ liệu vào xây dựng các ứng dụng trong thương mại điện tử, marketing và bán lẻ đã cho thấy một tiềm năng lớn của việc trích xuất tri thức từ cơ sở dữ liệu ( Knowledge Discovery in Database – KDD) trong tất cả các lĩnh vực hiện nay.

Trong chương này, nhóm nghiên cứu sẽ nêu ra những kỹ thuật KDD hiện tại, các ứng dụng được tích hợp khai phá dữ liệu trong y tế và sức khỏe cộng đồng, các vấn đề quan trọng và thử thách trong việc kết hợp khai phá dữ liệu với y học. Ngoài ra còn nêu lên sự phát triển của Khai phá dữ liệu tại Việt Nam.

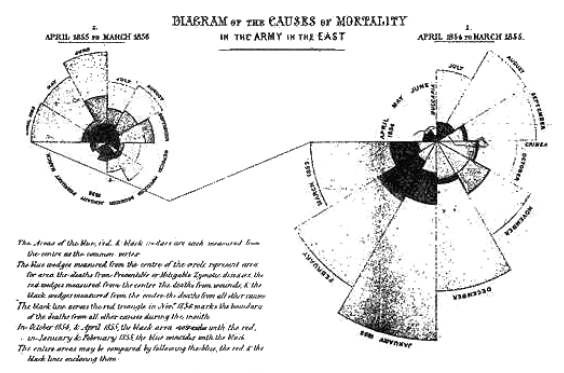
Phương pháp luận chính là dựa trên những bài báo đã được xuất bản từ năm 2000 trở lại đây trong nhiều lĩnh vực như y khoa, khoa học máy tính, kỹ thuật máy tính.

* 1. Mục tiêu:
* Liệt kệ các lợi ích và tầm quan trọng của khai phá dữ liệu trong y học.
* Tìm kiếm các kỹ thuật khai phá dữ liệu đã được sử dụng trong những lĩnh vực khác nhưng có thể áp dụng vào y học.
* Nếu lên sự phát triển của một lượng lớn các ứng dụng được kết hợp với khai phá dữ liệu.
* Xác định các vấn đề và thử thách trong ứng dụng khai phá dữ liệu vào y học.
* Đưa ra những khuyến nghị cho việc thu thập trị thức từ cơ sở dữ liệu thông qua khai phá dữ liệu.
  1. Các nghiên cứu:
     1. Khai pha dữ liệu trong y học:
* Có thể nói, việc sử dụng dữ liệu thật để hỗ trợ cho việc ra quyết định trong y học ( Evidence Bases Medicine – EBM) đã tồn tại trong nhiều thế kỷ.
* Năm 1854, John Snow, cha để của dịch tễ học hiện đại đã sử dụng các bản đồ đồ thị để phát hiện ra mầm bệnh thổ tả và đã chứng minh rằng bệnh này lây lan qua hệ thống bơm nước. Snow đã đếm số lượng bệnh nhân và vẽ sơ đồ địa chỉ của các bệnh nhân trên bản đồ bằng các thanh mà đen. Lúc đó ông đã phát hiện rằng hầu hết các ca tử vong đều nằm xung quanh một điểm bơm nước cố định tại London.



*Bản đồ mà John Snow dùng để phát hiện mầm bệnh*

* Năm 1855, Florence Nightingale đã phát minh ra biển đồ phân cực để tìm ra nhiều ca tử vong của binh lính do mất vệ sinh trong khám lâm sàn và có thể ngăn ngừa được. Bà đã sử dụng biểu đồ này để thuyết phục chính phủ cải cách các chính sách và việc này đã làm giảm đi đáng kể số lượng các ca tử vong.



*Biểu đồ phân cực do Nightinale phát minh*

* Snow và Nightingale đã tự mình thực hiện đầy đủ các bước: thu thập dữ liệu, sàng lọc và phân tích thông qua các dữ liệu về tỉ lệ tử vong trong suốt thời gian nghiên cứu vì số lượng dữ liệu có thể quản lý được. Nhưng hiện nay, việc bùng nổ dân số, sự toàn cầu hóa và tốc độ phát bệnh của bệnh dịch làm cho việc thao tác dữ liệu gần như không thể thực hiện.
* Đó cũng là nguyên nhân tại sao khai phá dữ liệu trở nên hữu ích trong y học. Tuy còn phát triển khá châm những vẫn được áp dụng để giải quyết các vấn đề khác nhau trong việc trích xuất tri thức. Có thể nói việc ứng dụng khai phá dữ liệu trong y học là một lĩnh vực còn rất non trẻ.
* Năm 2003, Wilson đã đưa ra một kết luận hết sức bất ngờ sau khi đã thực hiện một nghiên cứu trong các trường hợp mà KDD và các kỹ thuật khai phá dữ liệu đã được áp dụng trong y học. Cho đến thời điểm này, vẫn có nhiều tác giả cho rằng khai phá dữ liệu chỉ là một quá trình thu thập tri trức, còn một số khác thì cho rằng khai phá dữ liệu sử dụng kỹ thuật thống kê quá trình thu thập tri thức.
* Những khái niệm sai lầm này vẫn tồn tại trong cộng động y khoa. Trong đó có một định nghĩa tạm chấp nhận về khai phá dữ liệu ngày nay đó là một quy trình và kỹ thuật trong khai phá mô hình và hướng của dữ liệu (Witten and Frank 2005).
  + 1. Tầm quan trọng và lợi ích của khai phá dữ liệu mang lại:
  1. Khó khăn và thử thách:

1. **BỆNH TIỂU ĐƯỜNG:**
2. **CÁC THUẬT TOÁN ĐÃ ĐƯỢC ỨNG DỤNG VÀO KHAI PHÁ DỮ LIỆU CỦA BỆNH NHÂN MẮC BỆNH TIỂU ĐƯỜNG:**