



# **BỘ GIÁO DỤC ĐÀO TẠO** **TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Khoa Công nghệ Thông tin**

**PHÂN TÍCH DỮ LIỆU - BUỔI 01**  
**Tổng Quan Phân Tích Dữ Liệu**

**CBGD: Phan Thành Huấn**

✉️ ⚡ : [pthuan112358@gmail.com](mailto:pthuan112358@gmail.com)

📞 : 097 882 8842

## Lịch học

BUỔI	NGÀY	BUỔI	NGÀY
01	07/9/2023	07	19/10/2023
02	14/9/2023	08	26/10/2023
03	21/9/2023	09	02/11/2023
04	28/9/2023	10	09/11/2023
05	05/10/2023	11	16/11/2023
06	12/10/2023	12	23/11/2023
<b>BÁO CÁO ĐỒ ÁN (3 BUỔI 13-15)</b>			
30/11/2023		07/12/2023	14/12/2023

Chọn đề tài  
Đồ án môn học

## Đánh giá học phần

Bộ phận	Hình thức	
Quá trình 50%	Chuyên cần, thái độ học tập, ... 10%	
	Điểm bài tập ở nhà, trên lớp 40%	
Đồ án Cuối kỳ 50%	Nội dung 25%	Trả lời câu hỏi 5%
	Trình bày 15%	Tham dự đầy đủ 3 buổi báo cáo đồ án 5%

## Học liệu

### ❖ Tài liệu bắt buộc

1. Jiawei Han, Micheline Kamber, Jian Pei, “**Data Mining: Concepts and Techniques**”, Third Edition, Morgan Kaufmann Publishers, 2012.
2. David Hand, Heikki Mannila, Padhraic Smyth, “**Principles of Data Mining**”, MIT Press, 2001.



## Học liệu

### ❖ Tài liệu tham khảo

1. David L. Olson, Dursun Delen, “**Advanced Data Mining Techniques**”, Springer-Verlag, 2008.
2. Graham J. Williams, Simeon J. Simoff, “**Data Mining: Theory, Methodology, Techniques, and Applications**”, Springer-Verlag, 2006.
3. ZhaoHui Tang, Jamie MacLennan, “**Data Mining with SQL Server 2005**”, Wiley Publishing, 2005.
4. Oracle, “**Data Mining Concepts**”, B28129-01, 2008.



## Học liệu

### ❖ Tài liệu tham khảo

5. Oracle, “**Data Mining Application Developer’s Guide**”, B28131-01, 2008.
6. Ian H.Witten, Frank Eibe, Mark A. Hall, “**Data mining: practical machine learning tools and techniques**”, Third Edition, Elsevier Inc, 2011.
7. Florent Messegia, Pascal Poncelet & Maguelonne Teisseire, “**Successes and new directions in data mining**”, IGI Global, 2008.
8. Oded Maimon, Lior Rokach, “**Data Mining and Knowledge Discovery Handbook**”, Second Edition, Springer Science + Business Media, LLC 2005, 2010.

## Nội dung

1. Quá trình khám phá tri thức
2. Các khái niệm
3. Ý nghĩa và vai trò của phân tích dữ liệu
4. Ứng dụng của phân tích dữ liệu

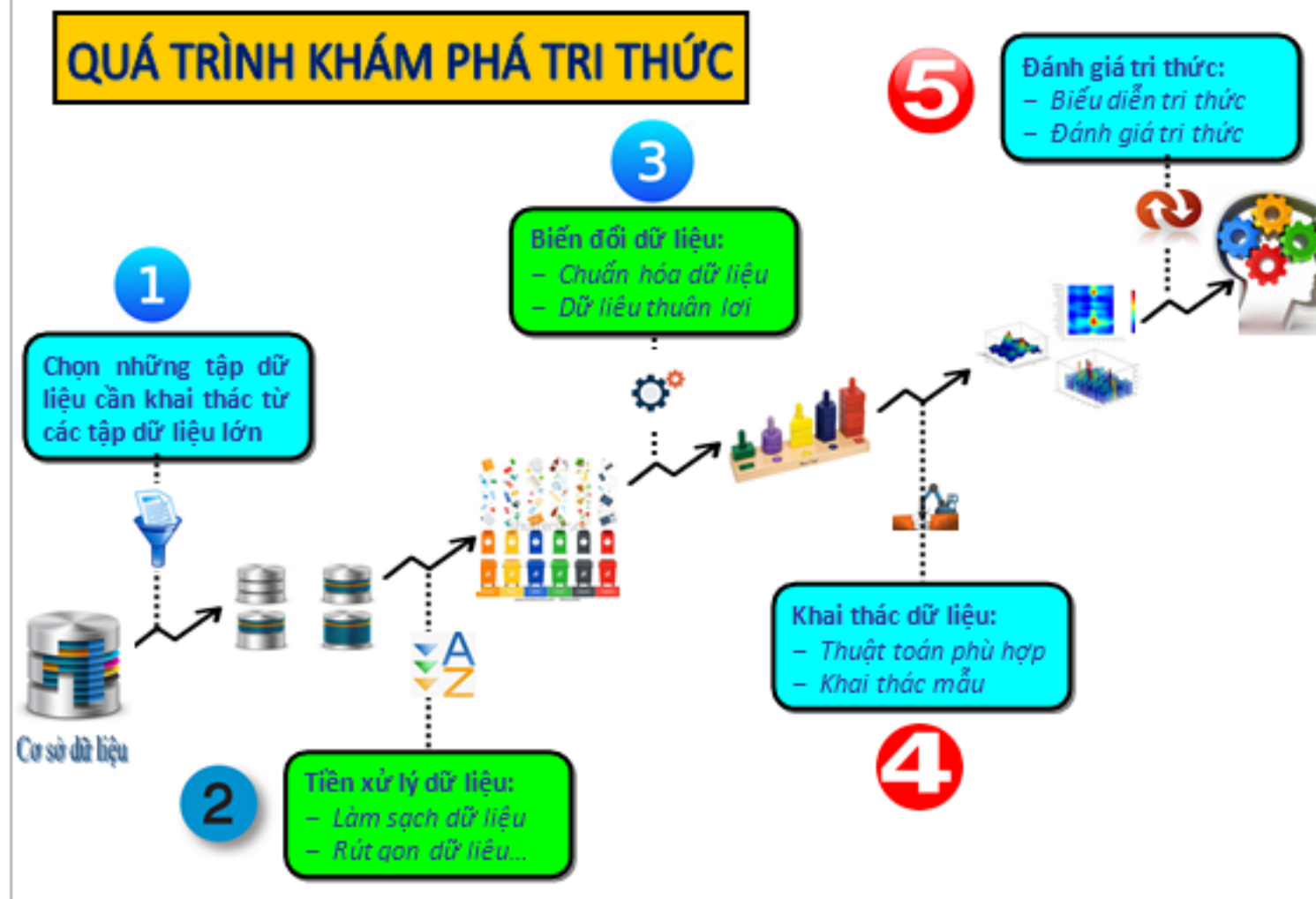
# 1-Quá trình khám phá tri thức

## (Knowledge Discovery in Databases - KDD)

1. Tìm hiểu và khám phá về các thông tin, kiến thức và sự hiểu biết mới trong một lĩnh vực hoặc chủ đề cụ thể;
2. Nghiên cứu, khám phá và tìm hiểu về các khía cạnh, nguyên tắc và quy luật trong một lĩnh vực tri thức nhằm hiểu rõ hơn về nó;
3. Khám phá và khai thác các nguồn tri thức có sẵn để tìm hiểu và nâng cao hiểu biết của mình về một chủ đề hoặc vấn đề;
4. Nghiên cứu, xây dựng và phát triển kiến thức mới thông qua việc phân tích, tổ chức và áp dụng thông tin đã thu thập được;



# 1-Quá trình khám phá tri thức (Knowledge Discovery in Databases - KDD)



“Khám phá tri thức là quá trình tìm ra những tri thức - *những mẫu tìm ẩn trước đó chưa biết và là thông tin hữu ích đáng tin cậy đang tồn tại trong hàng núi dữ liệu*”

# 1-Quá trình khám phá tri thức (Knowledge Discovery in Databases - KDD)

1. **Xác định mục tiêu:** Tìm hiểu mẫu, tạo mô hình dự đoán, phân loại, gom cụm, hay tìm hiểu thông tin mới;
2. **Thu thập dữ liệu:** Thu thập dữ liệu từ các nguồn khác nhau, bao gồm cả dữ liệu cấu trúc và không cấu trúc. Dữ liệu có thể được thu thập từ cơ sở dữ liệu, tệp tin, mạng xã hội, website, hoặc các nguồn khác;

# 1-Quá trình khám phá tri thức (Knowledge Discovery in Databases - KDD)

3. **Tiền xử lý dữ liệu:** Lọc dữ liệu không hợp lệ, xử lý dữ liệu khuyết, loại bỏ nhiễu, chuẩn hóa dữ liệu, và tạo các biểu diễn dữ liệu phù hợp cho quá trình phân tích;
4. **Phân tích dữ liệu:** Áp dụng các phương pháp và công cụ PTDL để khám phá thông tin ẩn chưa biết trong dữ liệu. Các phương pháp PTDL có thể bao gồm phân tích thống kê, học máy, khai phá dữ liệu, khai phá mạng, và các phương pháp khác;

# 1-Quá trình khám phá tri thức (Knowledge Discovery in Databases - KDD)

5. **Đánh giá mô hình:** Đánh giá hiệu suất của các mô hình và kết quả phân tích dữ liệu. Đánh giá có thể được thực hiện bằng cách sử dụng các phương pháp thống kê, so sánh với các mô hình khác, hoặc dựa trên phản hồi từ người dùng;
6. **Triển khai tri thức:** Áp dụng tri thức đã khám phá vào ứng dụng thực tế. Điều này có thể bao gồm việc xây dựng hệ thống dự đoán, tối ưu hoá quy trình kinh doanh, cung cấp thông tin hỗ trợ quyết định, hoặc phát triển sản phẩm và dịch vụ mới;



# 1-Quá trình khám phá tri thức (Knowledge Discovery in Databases - KDD)

7. **Giám sát và duy trì:** Theo dõi hiệu suất của ứng dụng tri thức và thực hiện các điều chỉnh cần thiết để đảm bảo rằng tri thức vẫn còn hiệu quả và đáng tin cậy trong thời gian dài.

Quá trình KDD không phải là một quá trình tuyến tính, mà thường là một quá trình lặp đi lặp lại, trong đó các bước có thể được thực hiện lại để cải thiện kết quả và tăng cường tri thức được khám phá.



## 2- Các khái niệm

Trong quá trình KPTT, việc “**chọn tập dữ liệu**” phù hợp để khai thác từ các tập dữ liệu lớn là rất quan trọng:

1. **Xác định mục tiêu:** Xác định mục tiêu của quá trình KPTT. Bạn cần biết rõ rằng bạn đang tìm kiếm thông tin gì từ dữ liệu và mục tiêu đó sẽ giúp bạn chọn tập dữ liệu phù hợp;
2. **Xem xét đặc điểm dữ liệu:** Kích thước, độ phức tạp, độ tin cậy và sẵn có của dữ liệu. Các tập dữ liệu lớn có thể chứa nhiều thông tin, nhưng cũng có thể khó xử lý  $\Rightarrow$  Cần đảm bảo rằng tập dữ liệu bạn chọn có đủ thông tin để đáp ứng mục tiêu;

## 2- Các khái niệm

3. **Xác định nguồn dữ liệu:** Có thể là cơ sở dữ liệu của tổ chức, dữ liệu trên mạng, dữ liệu từ các nguồn bên ngoài, hoặc các nguồn dữ liệu công cộng. Việc chọn nguồn dữ liệu đúng cũng đóng vai trò quan trọng trong việc đạt được mục tiêu;
4. **Áp dụng kỹ thuật lọc dữ liệu:** Trường hợp tập dữ liệu quá lớn  $\Rightarrow$  sử dụng kỹ thuật lọc dữ liệu để giảm kích thước của tập dữ liệu. Các kỹ thuật như lấy mẫu ngẫu nhiên, lấy mẫu theo thời gian hoặc lọc các thuộc tính không quan trọng có thể được áp dụng để giảm kích thước tập dữ liệu;

## 2- Các khái niệm

5. **Xác định quy trình KPTT:** Cần xác định quy trình KPTT được sử dụng để khai thác dữ liệu. Quy trình này bao gồm các bước như tiền xử lý dữ liệu, khai phá dữ liệu, phân tích kết quả và đánh giá  $\Rightarrow$  Cần đảm bảo rằng tập dữ liệu được chọn phù hợp với quy trình KPTT.

Tóm lại, việc chọn tập dữ liệu phù hợp là một phần quan trọng trong quá trình KPTT  $\Rightarrow$  Cần xác định mục tiêu, xem xét đặc điểm dữ liệu, xác định nguồn dữ liệu, áp dụng kỹ thuật lọc dữ liệu và xác định quy trình KDD để chọn tập dữ liệu phù hợp.

## 2- Các khái niệm

**“Tiền xử lý dữ liệu”** trong KPTT là giai đoạn quan trọng để chuẩn bị dữ liệu trước khi áp dụng các thuật toán KPTT:

- 1. Làm sạch dữ liệu:** Loại bỏ các dữ liệu không cần thiết, dữ liệu không chính xác, trùng lặp hoặc không hoàn chỉnh  $\Rightarrow$  Gồm việc kiểm tra tính hợp lệ của dữ liệu, xử lý các giá trị bị thiếu hoặc không hợp lệ và đồng nhất hóa dữ liệu;
- 2. Tiền xử lý văn bản:** Trường hợp dữ liệu là các tệp văn bản - tách từ, loại bỏ các stop word (từ không mang nhiều ý nghĩa), chuyển đổi chữ hoa thành chữ thường, và xử lý các ký tự đặc biệt;



## 2- Các khái niệm

3. **Xử lý ngôn ngữ tự nhiên:** Trường hợp dữ liệu chứa ngôn ngữ tự nhiên, có thể cần sử dụng các công cụ xử lý ngôn ngữ tự nhiên như xác định từ loại, phân tích cú pháp và rút trích thông tin để trích xuất thông tin quan trọng từ văn bản;
4. **Chuẩn hóa dữ liệu:** Chuẩn hóa dữ liệu để đảm bảo độ nhất quán và dễ dàng so sánh  $\Rightarrow$  việc chuyển đổi các đơn vị đo lường, định dạng ngày tháng hoặc số liệu;



## 2- Các khái niệm

5. **Rút trích đặc trưng:** Rút trích các đặc trưng quan trọng từ dữ liệu để sử dụng trong quá trình khám phá tri thức  $\Rightarrow$  việc chọn lọc các thuộc tính quan trọng, sử dụng các phương pháp như word embedding hoặc phân tích thành phần chính;
6. **Gom cụm:** Gom cụm các mẫu dữ liệu tương tự vào các nhóm để tạo ra cấu trúc dữ liệu phù hợp cho việc KPTT;

## 2- Các khái niệm

7. **Mã hóa dữ liệu:** Mã hóa dữ liệu thành một định dạng phù hợp để sử dụng trong các thuật toán KPTT  $\Rightarrow$  chuyển đổi dữ liệu thành ma trận hoặc biểu đồ.

Quá trình “**tiền xử lý dữ liệu**” trong KPTT đóng vai trò quan trọng để tạo ra dữ liệu sạch, chuẩn hóa và phù hợp cho việc áp dụng các thuật toán KPTT.

## 2- Các khái niệm

“**Biến đổi dữ liệu**” trong KPTT là giai đoạn tiếp theo sau quá trình tiền xử lý dữ liệu. Quá trình này nhằm biến đổi dữ liệu ban đầu thành dạng phù hợp để áp dụng các thuật toán KPTT:

1. **Vector hóa**: Dữ liệu thường được biểu diễn dưới dạng văn bản hoặc đặc trưng. Quá trình vector hóa giúp chuyển đổi dữ liệu thành các vectơ số hóa. Có nhiều phương pháp vector hóa như Bag-of-Words, TF-IDF, Word2Vec, hoặc GloVe;

## 2- Các khái niệm

2. **Biểu diễn đồ thị:** Nếu dữ liệu có cấu trúc đồ thị, quá trình biến đổi dữ liệu có thể liên quan đến biểu diễn đồ thị  $\Rightarrow$  Phương pháp như mạng Bayes, mạng Markov ẩn, hay mạng xã hội có thể được sử dụng để biểu diễn và biến đổi dữ liệu thành đồ thị;
3. **Mã hóa dữ liệu:** Dữ liệu có thể được mã hóa thành các định dạng phù hợp để sử dụng trong các thuật toán KPTT. Ví dụ, dữ liệu có thể được mã hóa thành ma trận hoặc biểu đồ;

## 2- Các khái niệm

4. **Gom cụm:** Dữ liệu có thể được gom cụm thành các nhóm dữ liệu tương tự nhau để tạo ra cấu trúc dữ liệu phù hợp cho việc KPTT  $\Rightarrow$  Phương pháp gom cụm như K-means, DBSCAN, hay hierarchical clustering có thể được sử dụng;
5. **Rút trích đặc trưng:** Quá trình này nhằm rút trích các đặc trưng quan trọng từ dữ liệu để sử dụng trong quá trình KPTT  $\Rightarrow$  Phương pháp như phân tích thành phần chính (PCA), Linear Discriminant Analysis (LDA), hoặc t-SNE có thể được áp dụng;



## 2- Các khái niệm

6. **Trích xuất thông tin:** Dữ liệu có thể được trích xuất thông tin quan trọng bằng cách sử dụng các phương pháp như trích xuất thực thể (entity extraction), trích xuất mối quan hệ (relation extraction), hoặc trích xuất tri thức (knowledge extraction).

Quá trình “**biến đổi dữ liệu**” trong KPTT giúp tạo ra dữ liệu phù hợp và chuẩn bị cho việc áp dụng các thuật toán KPTT để tìm ra các tri thức, mẫu và mối quan hệ tiềm ẩn trong dữ liệu.

## 2- Các khái niệm

Khai thác dữ liệu (Data Mining) là quá trình tìm kiếm, khám phá và phân tích thông tin ẩn, mẫu, và tri thức hữu ích từ một tập dữ liệu lớn. Nó bao gồm sử dụng các phương pháp và thuật toán để khám phá mối quan hệ, xu hướng, mẫu và tri thức tiềm ẩn trong dữ liệu.

1. **Áp dụng thuật toán và phương pháp KTDL:** Sử dụng các thuật toán và phương pháp KTDL như phân loại, gom cụm, hồi quy, khai phá luật kết hợp, phân tích chuỗi thời gian, và các phương pháp khác để khám phá mẫu, quy tắc, và tri thức tiềm ẩn trong dữ liệu;

## 2- Các khái niệm

2. **Đánh giá và lựa chọn mô hình:** Sau khi áp dụng các phương pháp KTDL, cần đánh giá hiệu suất của mô hình và lựa chọn mô hình tốt nhất dựa trên các tiêu chí như độ chính xác, độ tin cậy, độ phủ, và các yếu tố khác tương ứng với mục tiêu của dự án;
3. **Trích xuất tri thức:** Cuối cùng, quá trình KTDL dẫn đến việc trích xuất tri thức từ dữ liệu đã được khai thác. Tri thức có thể được biểu diễn dưới dạng các mô hình, quy tắc, mẫu, đồ thị, hoặc các dạng khác tùy thuộc vào loại dữ liệu và mục tiêu của dự án.

## 2- Các khái niệm

Quá trình KTDL là một phần quan trọng trong việc *tìm kiếm tri thức từ dữ liệu và đưa ra quyết định thông minh*. Nó được sử dụng rộng rãi trong nhiều lĩnh vực như kinh doanh, y tế, khoa học, marketing, và nhiều lĩnh vực khác để tạo ra tri thức mới, đưa ra dự đoán, tối ưu hóa quy trình, và tìm hiểu thêm về dữ liệu.



### 3- Ý nghĩa và vai trò của Phân tích dữ liệu

- 1. Tìm kiếm thông tin hữu ích:** PTDL giúp tìm kiếm thông tin, mẫu, quy luật, xu hướng hay tri thức ẩn trong dữ liệu  $\Rightarrow$  Hiểu rõ hơn về một vấn đề cụ thể và đưa ra quyết định thông minh;
- 2. Xác định mẫu và xu hướng:** PTDL giúp xác định các mẫu và xu hướng trong dữ liệu  $\Rightarrow$  Dự đoán, dự báo và đưa ra các quyết định dựa trên các xu hướng dữ liệu hiện tại;



### 3- Ý nghĩa và vai trò của Phân tích dữ liệu

3. **Phát hiện dữ liệu ngoại lệ:** PTDL có thể giúp phát hiện các dữ liệu ngoại lệ, tức là các giá trị dữ liệu không phù hợp hoặc không tuân theo các quy luật thông thường  $\Rightarrow$  Phát hiện sự cố, lỗi hoặc hành vi bất thường trong hệ thống;
4. **Tối ưu hóa quy trình và hiệu suất:** PTDL giúp tìm ra các cách tối ưu hóa quy trình và hiệu suất làm việc  $\Rightarrow$  Cải thiện năng suất, giảm thiểu lãng phí và tăng cường hiệu quả hoạt động;

### 3- Ý nghĩa và vai trò của Phân tích dữ liệu

5. **Hỗ trợ ra quyết định:** PTDL cung cấp thông tin và dữ liệu để hỗ trợ quyết định. Dựa trên kết quả phân tích, người quản lý và nhà quyết định có thể đưa ra các quyết định thông minh và dựa trên dữ liệu;
6. **Dự báo và dự đoán:** PTDL giúp dự báo và dự đoán các sự kiện tương lai dựa trên dữ liệu hiện tại và các mẫu đã xác định  $\Rightarrow$  Hỗ trợ quyết định kế hoạch và lập kế hoạch tương lai;
7. **Tạo ra tri thức mới:** PTDL giúp tạo ra tri thức mới từ dữ liệu. Bằng cách khám phá, phân tích và kết hợp các thông tin trong dữ liệu, người ta có thể tìm ra tri thức mới, hiểu rõ hơn về một vấn đề cụ thể và đưa ra các phát hiện mới.

## 4- Ứng dụng của Phân tích dữ liệu

- 1. Quản lý khách hàng:** Hiểu rõ hơn về *hành vi* và *sở thích* của khách hàng, từ đó tạo ra các chiến lược tiếp thị và dịch vụ tốt hơn  $\Rightarrow$  Xây dựng các hồ sơ khách hàng, dự đoán hành vi mua hàng và tạo ra các chiến dịch tiếp thị cá nhân hóa;
- 2. Phân tích thị trường:** Hiểu rõ hơn về thị trường, đối thủ cạnh tranh và các xu hướng tiêu dùng  $\Rightarrow$  Nắm bắt thông tin thị trường, dự đoán xu hướng và đánh giá hiệu quả các chiến lược kinh doanh;

## 4- Ứng dụng của Phân tích dữ liệu

- 3. Dự đoán và dự báo:** Dự đoán và dự báo các sự kiện tương lai. Ví dụ, ngành tài chính sử dụng phân tích dữ liệu để dự đoán giá cổ phiếu, tỷ giá hối đoái, và rủi ro tín dụng  $\Rightarrow$  Dự đoán rủi ro và định giá bảo hiểm;
- 4. Phân tích dữ liệu y tế:** Giúp phát hiện và dự đoán các bệnh, tối ưu hóa quy trình chăm sóc sức khỏe và cải thiện kết quả điều trị  $\Rightarrow$  Các bác sĩ và nhà nghiên cứu đánh giá mối quan hệ giữa các yếu tố y tế và dự đoán kết quả điều trị;



## 4- Ứng dụng của Phân tích dữ liệu

- 5. **An ninh mạng:** Phát hiện và ngăn chặn các cuộc tấn công mạng ⇒ Giám sát, phát hiện và phản ứng nhanh với các hoạt động đáng ngờ trên hệ thống mạng;
- 6. **Quản lý chuỗi cung ứng:** Tối ưu hóa quy trình chuỗi cung ứng, từ việc dự đoán nhu cầu sản phẩm, quản lý kho hàng, đến định giá và quản lý rủi ro ⇒ Cải thiện hiệu suất và giảm thiểu lãng phí trong chuỗi cung ứng;



## 4- Ứng dụng của Phân tích dữ liệu

- 7. Phân tích xã hội:** Hiểu rõ hơn về hành vi và tương tác của con người trong mạng xã hội  $\Rightarrow$  Đánh giá hiệu quả chiến dịch tiếp thị, nghiên cứu ý kiến khách hàng và phát hiện xu hướng xã hội;
- 8. Dự đoán và phân tích thời tiết:** Dự đoán và phân tích các biến đổi thời tiết  $\Rightarrow$  Lập kế hoạch vận chuyển, quản lý năng lượng và ứng phó với các tác động thời tiết.

...

## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**Quản lý khách hàng**”

- **Google Analytics**: Cung cấp thông tin về lưu lượng truy cập, hành vi người dùng và hiệu quả các chiến dịch tiếp thị trên trang web;
- **Tableau**: Phân tích dữ liệu mạnh mẽ, cho phép kết nối và phân tích dữ liệu từ nhiều nguồn khác nhau, tạo ra các biểu đồ và báo cáo tương tác để hiểu rõ hơn về khách hàng;
- **Microsoft Power BI**: Phân tích dữ liệu và trực quan hóa, cho phép kết nối và tổ chức dữ liệu từ nhiều nguồn khác nhau, tạo ra các báo cáo và biểu đồ để hiểu rõ hơn về khách hàng;

## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**Phân tích thị trường**”

- **SAS**: Phân tích và khai thác dữ liệu mạnh mẽ - công cụ và thuật toán phân tích phức tạp để hiểu và dự đoán xu hướng thị trường;
- **IBM SPSS**: Phân tích và khai thác dữ liệu phổ biến - công cụ và thuật toán phân tích thống kê để nghiên cứu và dự đoán thị trường;
- **QlikView**: Một công cụ phân tích dữ liệu trực quan, cho phép người dùng tạo ra các báo cáo và biểu đồ tương tác từ dữ liệu thị trường để hiểu và phân tích thông tin;

## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**Dự đoán và dự báo**”

- **RapidMiner**: Công cụ và thuật toán dự báo để tạo ra mô hình dự đoán và đưa ra dự báo cho các vấn đề phức tạp;
- **Microsoft Excel**: Cung cấp các chức năng dự đoán và dự báo như hồi quy, dự đoán chuỗi thời gian và phân tích phức tạp bằng cách sử dụng các công thức và chức năng tích hợp;
- **Python** và **R**: Hai ngôn ngữ lập trình phổ biến - cung cấp các thư viện và gói phân tích dữ liệu mạnh mẽ như *scikit-learn* (Python) và *forecast* (R) để xây dựng mô hình dự đoán và dự báo;



## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “Phân tích dữ liệu y tế”

- **Microsoft HealthVault**: Lưu trữ và quản lý dữ liệu y tế - người dùng tổ chức, phân tích và chia sẻ dữ liệu y tế cá nhân và gia đình;
- **OpenClinica**: Phần mềm mã nguồn mở - công cụ và chức năng phân tích dữ liệu để hỗ trợ việc thu thập và phân tích dữ liệu y tế trong các nghiên cứu y tế;
- **Python** và **R**: Hai ngôn ngữ lập trình phổ biến - cung cấp các thư viện và gói phân tích dữ liệu y tế mạnh mẽ như *scikit-learn* (Python) và *Bioconductor* (R) để phân tích dữ liệu y tế.

## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**An ninh mạng**”

- **IBM QRadar**: Giám sát và phân tích mạng - chức năng phân tích dữ liệu để phát hiện và phản ứng đối với các mối đe dọa an ninh;
- **McAfee Enterprise Security Manager**: Tổ chức dữ liệu và phân tích an ninh - người dùng thu thập và phân tích dữ liệu từ các nguồn mạng để phát hiện và giải quyết các mối đe dọa;
- **Elastic SIEM**: Mã nguồn mở, sử dụng nền tảng Elastic Stack - chức năng phân tích dữ liệu và giám sát an ninh để phát hiện và phản ứng đối với các mối đe dọa;

## 4- Ứng dụng của Phân tích dữ liệu

- ❖ Một số phần mềm hỗ trợ “**Phân tích chuỗi cung ứng**”
  - **Oracle Supply Chain Management**: Công cụ phân tích dữ liệu để tối ưu hóa các quy trình và quyết định trong chuỗi cung ứng;
  - **Blue Yonder (formerly JDA Luminate)**: PTDL và trí tuệ nhân tạo của Blue Yonder - phân tích dữ liệu để tối ưu hóa và dự đoán các quyết định trong chuỗi cung ứng;
  - **IBM Sterling Supply Chain Insights**: PTDL và trí tuệ nhân tạo - công cụ phân tích dữ liệu để hiểu và tối ưu hóa hoạt động trong chuỗi cung ứng;

## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**Phân tích xã hội**”

- **Meltwater**: PTDL xã hội và quản lý dư luận - thu thập, phân tích và theo dõi thông tin từ nguồn mạng xã hội, truyền thông trang web;
- **Quid**: PTDL và trí tuệ nhân tạo, cung cấp các công cụ và thuật toán phân tích để hiểu, khám phá thông tin từ các nguồn mạng xã hội & truyền thông;
- **Python** và **R**: thư viện và gói phân tích dữ liệu mạnh mẽ như tweepy (Python) và twitterR (R) để thu thập và PTDL từ các nền tảng mạng xã hội;



## 4- Ứng dụng của Phân tích dữ liệu

### ❖ Một số phần mềm hỗ trợ “**Dự đoán và phân tích thời tiết**”

- **IBM Weather Company Data:** Cung cấp dữ liệu thời tiết chi tiết và chính xác từ các nguồn khác nhau, cho phép người dùng phân tích và dự đoán thời tiết trong tương lai;
- **The Weather Channel API:** Một giao diện lập trình ứng dụng của The Weather Channel, cung cấp dữ liệu thời tiết và các công cụ phân tích để tích hợp vào ứng dụng và dự đoán thời tiết;
- **AccuWeather:** Cung cấp dữ liệu thời tiết và các công cụ phân tích, cho phép người dùng dự đoán và hiểu thông tin về thời tiết từ khắp nơi trên thế giới.