**GIỚI THIỆU CHUNG**

Khoa học dữ liệu (KHDL) là khoa học về việc quản trị và phân tích dữ liệu để tìm ra các hiểu biết, các tri thức hành động, các quyết định dẫn dắt hành động. KHDL gồm ba phần chính: Tạo ra và quản trị dữ liệu, phân tích dữ liệu, và chuyển kết quả phân tích thành giá trị của hành động. Nôm na bước thứ nhất là về số hóa và bước thứ hai là về dùng dữ liệu. Việc phân tích và dùng dữ liệu lại dựa vào ba nguồn tri thức: toán học (thống kê toán học), công nghệ thông tin (máy học) và tri thức của lĩnh vực ứng dụng cụ thể.

**ỨNG DỤNG CỦA KHOA HỌC DỮ LIỆU**

Nếu phân tích dữ liệu về nhu cầu thị trường ta có thể quyết định cần nuôi bao nhiêu lợn mỗi nơi mỗi lúc. Nếu có và phân tích được dữ liệu mô phỏng các phương án xả lũ vào mùa mưa ta có thể chọn được cách xả lũ ít thiệt hại nhất. Nếu có và phân tích được các bệnh án điện tử của người bệnh ta có thể tìm ra được phác đồ thích hợp hơn cả cho người bệnh. Amazon đã phân tích các lần mua hàng trước của bạn để dự đoán những món đồ bạn có thể sẽ thích mua và gửi quảng cáo tới, v.v. Khi nghe nói về các thành tựu đột phá gần đây của Trí tuệ nhân tạo người nghe có thể cũng chưa biết rằng phần lớn chúng đều dựa vào các phương pháp và đột phá của KHDL.

**Mạng xã hội và dữ liệu người dùng**

Facebook, mạng xã hội lớn nhất hành tinh, một trong những cái tên được nhắc tới nhiều nhất trong giới trẻ hiện nay, là một trong những ứng dụng nổi tiếng của khoa học dữ liệu.

Tại hội nghị các nhà phát triển F8 đầu năm 2016, CEO Mark Zurkerberg cũng đã thông báo về một lộ trình mười năm tới. Trong đó, sẽ tạo ra một hệ sinh thái với những sản phẩm và công nghệ tiên tiến như trí tuệ nhân tạo (Artificial Intelligence). Tất cả đều dựa trên toàn bộ nguồn dữ liệu từ người dùng và các thuật toán máy học (Machine Learning Algorithms).

**"Mỗi cú click chuột, mỗi cái like, mỗi bình luận và tất cả các kết nối đều được sử dụng để xây dựng một hồ sơ hoàn chỉnh cho mỗi người dùng."**

Đằng sau những trải nghiệm kết nối và tương tác giữa bạn bè và người thân, đó là sự vận hành của các thuật toán đánh giá người dùng được xây dựng bởi những kỹ sư hàng đầu thế giới.

Tính tới tháng 8 năm 2016, tổng số lượng người dùng trên trang này cán đến mốc 1,750,000 người, gấp 5 lần dân số nước Mỹ, tương đương với 1/3 dân số thế giới và lớn hơn tổng số dân của châu Âu, châu Úc và Nam Mỹ cộng lại.

**Và những nhà khoa học phân tích dữ liệu …**

Theo thống kê của Glassdoor, một trong những trang web việc làm lớn nhất thế giới, ngành khoa học dữ liệu đứng đầu trong số 25 nghề nghiệp tốt nhất, đứng thứ 16 về mức lương với trung bình hơn $116,000 và có nhiều vị trí được tìm kiếm tuyển dụng nhất trong năm 2015 ở Hoa Kỳ

Trong một nghiên cứu của O'Reilly, một trong những nhà phát hành chuyên về mảng công nghệ và khoa học máy tính, có 4 dạng nhà khoa học dữ liệu tiêu biểu.

**1. Doanh nhân (Data Businesspeople)**

Quan tâm vào sản phẩm và phát triển lợi nhuận, họ là các nhà lãnh đạo, nhà quản lý và doanh nhân có sự am hiểu về mặt kỹ thuật. Đa phần đều có nền tảng giáo dục xuất phát bằng kỹ sư kết hợp với một MBA.

**2. Nhà sáng tạo (Data Creatives)**

Có nhiều biệt tài và kinh nghiệm với nhiều dạng dữ liệu và công cụ, những nhà sáng tạo thường ví von mình như là một nghệ sĩ hoặc tin tặc. Điểm nhấn thường thấy là sự xuất sắc sử dụng các công nghệ minh họa (Visualization Techonology) và mã nguồn mở.

**3. Nhà phát triển (Data Developers)**

Nhà phát triển dữ liệu thường tập trung vào việc viết phần mềm để làm phân tích, thống kê, và nhiệm vụ học máy, thường xuyên trong môi trường sản xuất. Họ thường có trình độ khoa học máy tính, và thường xuyên làm việc với cái gọi là "dữ liệu lớn" (Big Data).

**4. Nhà nghiên cứu (Data Researchers)**

Đó là những người áp dụng những kỹ năng được đào tạo trong khoa học cùng với các công cụ và kỹ thuật, số liệu. Một số có bằng tiến sĩ, và các ứng dụng sáng tạo các công cụ toán học mang lại những hiểu biết và sản phẩm có giá trị.