MINING DATA

AND WHY IT MATTERS

FROM POLITICS TO BIOLOGY TO BIG DATA, THE EFFECTS OF DATA MINING ARE *EVERYWHERE*

ONE WAY DATA MINING CAN INFLUENCE OUR LIVES IS THROUGH PREDICTIONS

Genomics

Data mining has provided massive insights to understanding the human genome, illuminating the causes of ADHD (attention deficit/hyperactivity disorder), obesity, asthma, colorectal cancer, Crohn's disease, and providing clues to factors that have influenced human evolution.

Politics

Political campaigns are very early adopters of data mining techniques. Today, most campaigns make vast use of voter data to identify particular areas or demographics that need to be targeted. Fund raising has also greatly benefited by a deeper understanding of how effective different approaches to that are, a knowledge directly derived from detailed data analysis.

Education

Various school districts, colleges and universities today, leverage data collected as students progress through their education. That data helps identifying students that are in need of individual attention or who could benefit from teaching methods that are tailored to them, helping maximize their chance of succeeding.

EXPLODING DATA

THE POTENTIAL OF BIG DATA

WHAT IS BIG DATA?

"Big Data" refers to sets of data whose size surpasses that of what data storage tools can typically handle. It's something that grows concurrently with the development of technology and something that helps

The amount of digital data in our world has been exponentially growing in just a few short years. Big data has the potential to become the next frontier for innovation, competition, and profit.

A GROWING TORRENT

Just how big is big data? Huge; and with the potential to expand even more in the future.

4 Billion

PIECES OF CONTENT ARE SHARED EVERYDAY ON FACEBOOK

235 Terabytes

OF DATA HAS BEEN COLLECTED BY THE US LIBRARY OF CONGRESS IN APRIL 2011

40%

THE PROJECTED GROWTH OF GLOBAL DATA PER YEAR

5%

THE PROJECTED GROWTH OF GLOBAL IT SPENDING PER YEAR

DATA SECTORS

15 OUT OF 17

SECTORS IN THE UNITED STATES HAVE MORE DATA PER COMPANY THAN THE US LIBRARY OF CONGRESS

AMOUNT OF STORED STATE BY SECTOR: (IN PETABYTES, 2009)

- DISCRETE MANUFACTURING: 966
- GOVERNMENT: 848
- COMMUNICATIONS: 715
- PROCESS MANUFACTURING: 694
- BANK: 619
- HEALTHCARE: 434
- SECURITIES & INVESTMENTS: 429
- PROFESSIONAL SERVICES: 411
- RETAIL: 364
- EDUCATION: 269
- INSURANCE: 243
- TRANSPORTATION: 227
- WHOLESALE: 202
- UTILITIES: 194
- RESOURCE INDUSTRIES: 116
- CONSUMER RECREATIONAL: 106
- CONSTRUCTION: 51

3.8 PETABYTES

ON AVERAGE, A SECURITIES AND INVESTMENT FIRM WITH UNDER 1,000 EMPLOYEES WILL HAVE 3.8 PETABYTES OF DATA STORED.

WHAT IS THIS DATA WORTH?

Simply knowing that all this data is out there is one thing. Utilizing this data to turn a profit is another issue. Using big data has the potential to increase profitability across all sectors. However there are five sectors that stand to benefit the most.

HEALTH

Healthcare has lagged behind many other industries when it comes to improving operational performance and adopting technology-enabled process improvements.

$300 Billion

IN TEN YEARS, THE IMPLEMENTATION OF BIG DATA IN THE HEALTH INDUSTRY COULD CAPTURE $300 BILLION ANNUALLY

AND BRING CURRENT HEALTHCARE EXPENDITURES BY THE US GOVERNMENT DOWN 8%

- 2010: 17%
- 2020: 9%

HEALTHCARE EXPENDITURES AS A PERCENTAGE OF THE GDP

AREAS WHERE BIG DATA COULD BE UTILIZED

POTENTIAL WORTH OF DATA

- R&D — Research and development; clinical trail design; personalized medicine: $108 BILLION
- CLINICAL — Transparency in clinical data and clinical decision support: $65 BILLION
- ACCOUNTING — Advanced fraud detection; performance-based drug pricing: $47 BILLION
- PUBLIC HEALTH — Public health surveillance and response systems: $9 BILLION
- NEW BUSINESS MODEL — Aggregation of patient records; online platforms and communities.: $5 BILLION

GOVERNMENT

Governments in many parts of the world need to increase their productivity through digital means. Examining the public sector of the European Union, we can where the utilization of big data can create value through efficiency.

€300 BILLION

UTILIZING BIG DATA, EUROPE'S PUBLIC SECTOR COULD REDUCE COSTS BY 20% OR 300 BILLION EUROS

AREAS WHERE BIG DATA COULD BE UTILIZED

POTENTIAL WORTH OF DATA

- OPERATIONAL EFFICIENCY SAVINGS: €200 BILLION
- REDUCTION IN FRAUD AND ERROR: €30 BILLION
- INCREASE IN TAX COLLECTION: €110 BILLION

RETAIL

The use of technology and digital data in the retail industry has allowed for a boost in profitability and productivity over the decade. The continued adoption of big data has the potential for further profitability.

60%

THE POTENTIAL INCREASE IN RETAILERS' OPERATING MARGINS FROM BIG DATA COULD BE 60%

MANUFACTURING

The manufacturing sector has adopted data in the use of information technology and automation. Continued adoption of big data could lead to increased production and decreased costs.

50%

BIG DATA HAS THE POTENTIAL TO CUT OPERATING COSTS BY NEARLY 50% ACROSS ALL SECTORS OF MANUFACTURING

PERSONAL LOCATION TECHNOLOGY

Personal location data volumes has increased rapidly with the adoption of mobile phones. The potential for this data is far greater than any other, because it is not confined to a single sector, but cuts across all industries.

$600 BILLION

THE POTENTIAL ANNUAL CONSUMER SURPLUS FROM GLOBAL PERSONAL LOCATION DATA IS $600 BILLION

1 Petabyte

AREAS WHERE BIG DATA COULD BE UTILIZED

POTENTIAL WORTH OF DATA

- GPS — Navigation including smart routing based on real-time traffic: $500 BILLION
- MARKETING — Geo-targeted mobile advertising (advertising platform providers): $100 BILLION
- SOCIAL — People tracking location sharing and entertainment: $57 BILLION

CAN WE CAPTURE THE FULL POTENTIAL OF BIG DATA?

There are several issues that must be dealt with before all industry sectors can access the full potential of this big data.

DATA POLICIES

Big data raises several legal issues due to the fact that data is fundamentally different from other assets. Since data can be so easily copied, intellectual property becomes an urgent consideration to policy makers. Also, there is an issue of culpability. Who is liable for an inaccurate piece of data when it leads to negative consequences?

TECHNOLOGY

Local systems and inferior standards/formats prevents the integration of big data in many sectors, especially the public sector. Making use of large datasets requires both adequate storage and compatible technology. These investment costs are sometimes far too large.

DATA ACCESS

In order to create the most broad data available, companies will need to increasingly rely on third-party data sources and integrate external information with their own. Currently, there are not completely efficient markets that allow for this transfer and sharing of data..

TALENT

In many instances, there is a lack of skilled personnel needed to mine big data, create the necessary structures, and make use of big data through informed decisions.

Sources:

- "Big Data: The Next Frontier for Innovation, Competition and Productivity." US Bureau of Labor Statistics | McKinsey Global Institute Analysis
- Column Five Infographic