

Data Mining: The Scientific and Industry Revolution

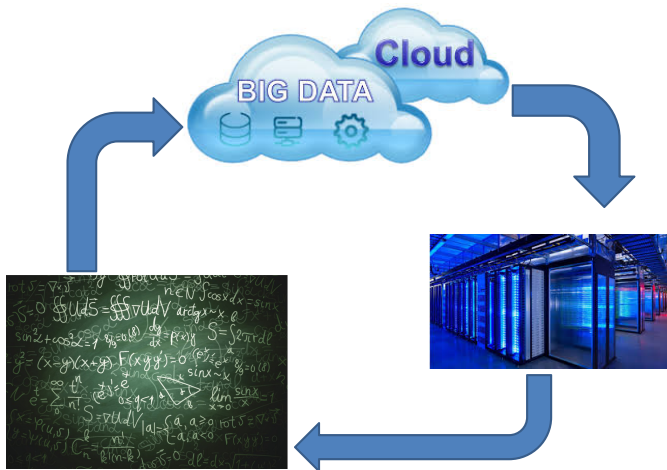
Alvaro J. Riascos Villegas
University of los Andes and Quantil

July 6 2018

Content

1 Introduction

The Scientific and Industry Revolution



Introduction: Data

- In the last 2 years more information has been produced and stored than that throughout all history of humanity.
- The projects Australia's radio telescopes, the Large Hadron Collider (CERN), etc. generate various petabytes ($10^6 GB = 10^3 TB$) of daily information.
- The daily consumption of Netflix is 100 million videos.
- 400 hours of video are uploaded to Youtube every minute.
- Cisco estimates that in 2020, 200 petabytes of video per second will be transferred around the world.

Introduction: Data

- In the last 2 years more information has been produced and stored than that throughout all history of humanity.
- The projects Australia's radio telescopes, the Large Hadron Collider (CERN), etc. generate various petabytes ($10^6 GB = 10^3 TB$) of daily information.
- The daily consumption of Netflix is 100 million videos.
- 400 hours of video are uploaded to Youtube every minute.
- Cisco estimates that in 2020, 200 petabytes of video per second will be transferred around the world.

Introduction: Data

- In the last 2 years more information has been produced and stored than that throughout all history of humanity.
- The projects Australia's radio telescopes, the Large Hadron Collider (CERN), etc. generate various petabytes ($10^6 GB = 10^3 TB$) of daily information.
- The daily consumption of Netflix is 100 million videos.
- 400 hours of video are uploaded to Youtube every minute.
- Cisco estimates that in 2020, 200 petabytes of video per second will be transferred around the world.

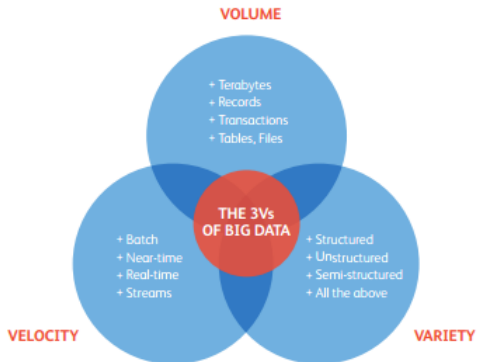
Introduction: Data

- In the last 2 years more information has been produced and stored than that throughout all history of humanity.
- The projects Australia's radio telescopes, the Large Hadron Collider (CERN), etc. generate various petabytes ($10^6 GB = 10^3 TB$) of daily information.
- The daily consumption of Netflix is 100 million videos.
- 400 hours of video are uploaded to Youtube every minute.
- Cisco estimates that in 2020, 200 petabytes of video per second will be transferred around the world.

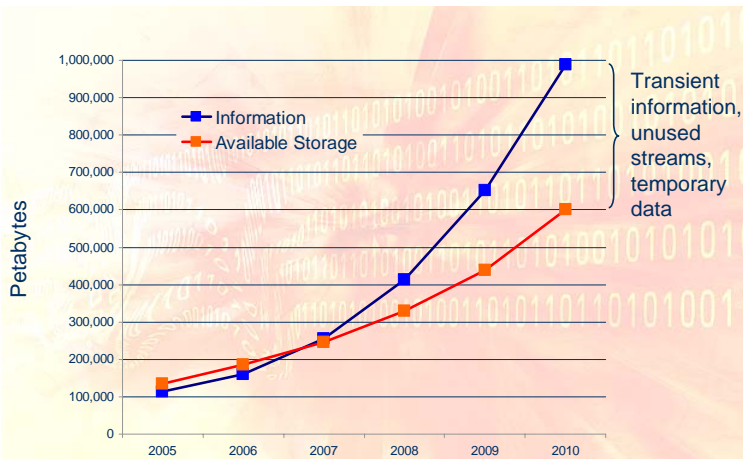
Introduction: Data

- In the last 2 years more information has been produced and stored than that throughout all history of humanity.
- The projects Australia's radio telescopes, the Large Hadron Collider (CERN), etc. generate various petabytes ($10^6 GB = 10^3 TB$) of daily information.
- The daily consumption of Netflix is 100 million videos.
- 400 hours of video are uploaded to Youtube every minute.
- Cisco estimates that in 2020, 200 petabytes of video per second will be transferred around the world.

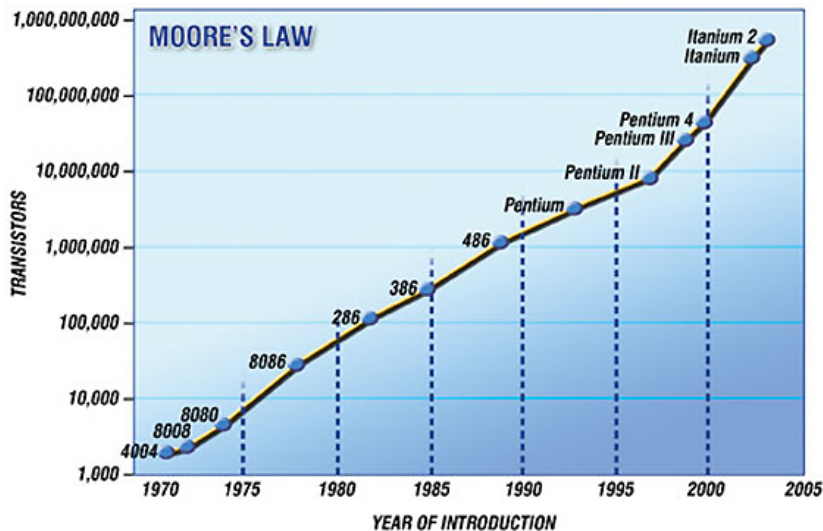
- What all these information has in common:



Introduction: Data

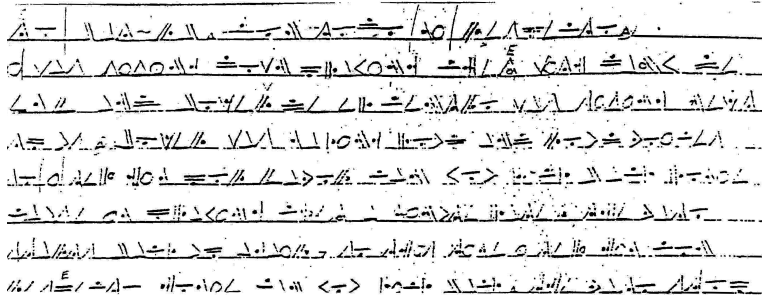


Introduction: Computation



Introduction: Mathematics

- A psychologist from a prison in California approached the Stanford Statistics Department with this:



The image shows a piece of lined paper with a dense, chaotic scribble of black ink marks. The marks are completely illegible and do not form any recognizable mathematical symbols or text. This is a well-known example of a 'proof' submitted by a prisoner to a statistics department, illustrating the concept of a 'proof by scribble'.

Introduction: Mathematics

ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS
NOBLER IN THE MIND TO SUFFER THE SLINGS AND ARROWS OF OUTRAGEOUS
FORTUNE OR TO TAKE ARMS AGAINST A SEA OF TROUBLES AND BY OPPOSING END

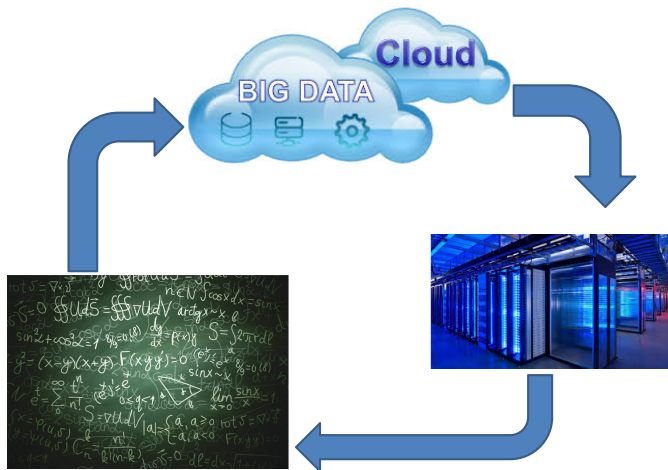
Introduction: Mathematics

100 ER ENOHLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDEL
300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOPHE HELO ITOSUVDEL
400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOS R O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODILO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFILO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHETHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHETHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBREL
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER

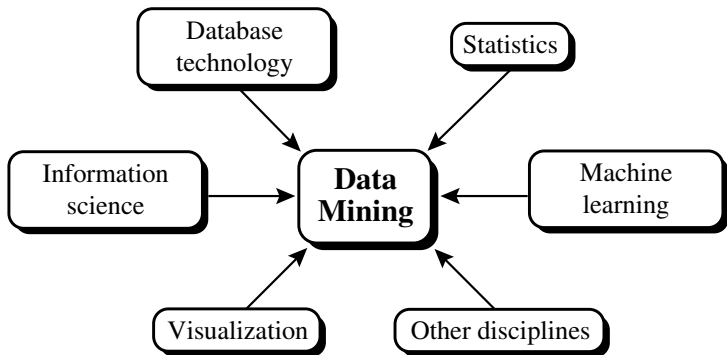
Introduction: Mathematics

to bat-rb. con todo mi respeto. i was sitting down playing chess with danny de emf and boxer de el centro was sitting next to us. boxer was making loud and loud voices so i tell him por favor can you kick back homie cause im playing chess a minute later the vato starts back up again so this time i tell him con respecto homie can you kick back. the vato stop for a minute and he starts up again so i tell him check this out shut the f**k up cause im tired of your voice and if you got a problem with it we can go to celda and handle it. i really felt disrespected thats why i told him. anyways after i tell him that the next thing I know that vato slashes me and leaves. dy the time i figure im hit i try to get away but the c.o. is walking in my direction and he gets me right dy a celda. so i go to the hole. when im in the hole my home boys hit doxer so now "b" is also in the hole. while im in the hole im getting schoold wrong and

The Scientific and Industry Revolution in a Nutshell



The new discipline



[http://www.alvaroriascos.com/mineriadatos/
The-age-of-analytics-Executive-summary.pdf](http://www.alvaroriascos.com/mineriadatos/The-age-of-analytics-Executive-summary.pdf)

Data mining process

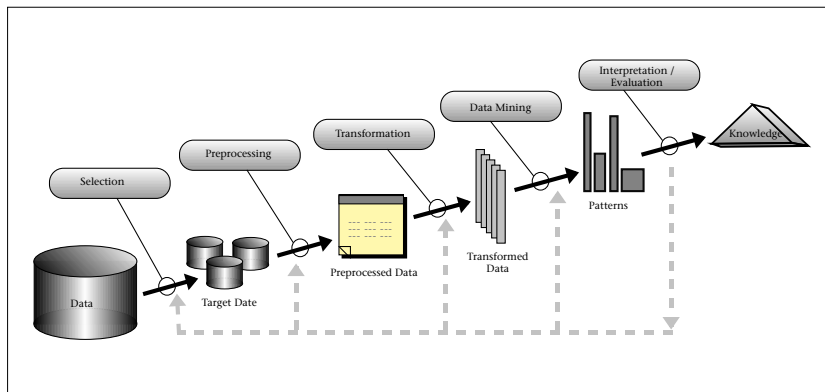


Figure 1. An Overview of the Steps That Compose the KDD Process.

- KDD: *Knowledge Discovery in Datasets*