Computational Communications and Data Analysis

Lecture 1: A Brief Introduction

Ting Wang

Outlines

1. Significance of Data Analysis
2. Definition of Data Analysis
3. History of Computer Data Analytics
4. Domains of New Media Data Analytics

上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

the significance of data analysis

Why Data Analysis

Why Data Analysis

Why Data Analysis

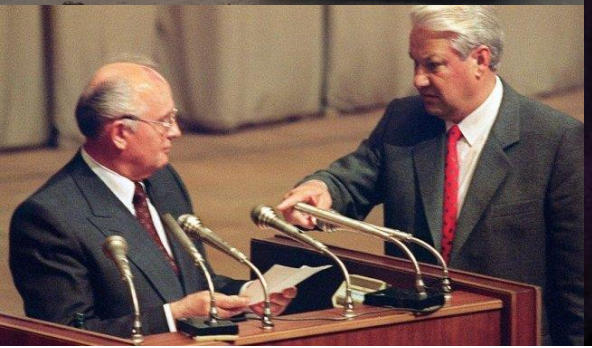
US Dollars,

The King of the World

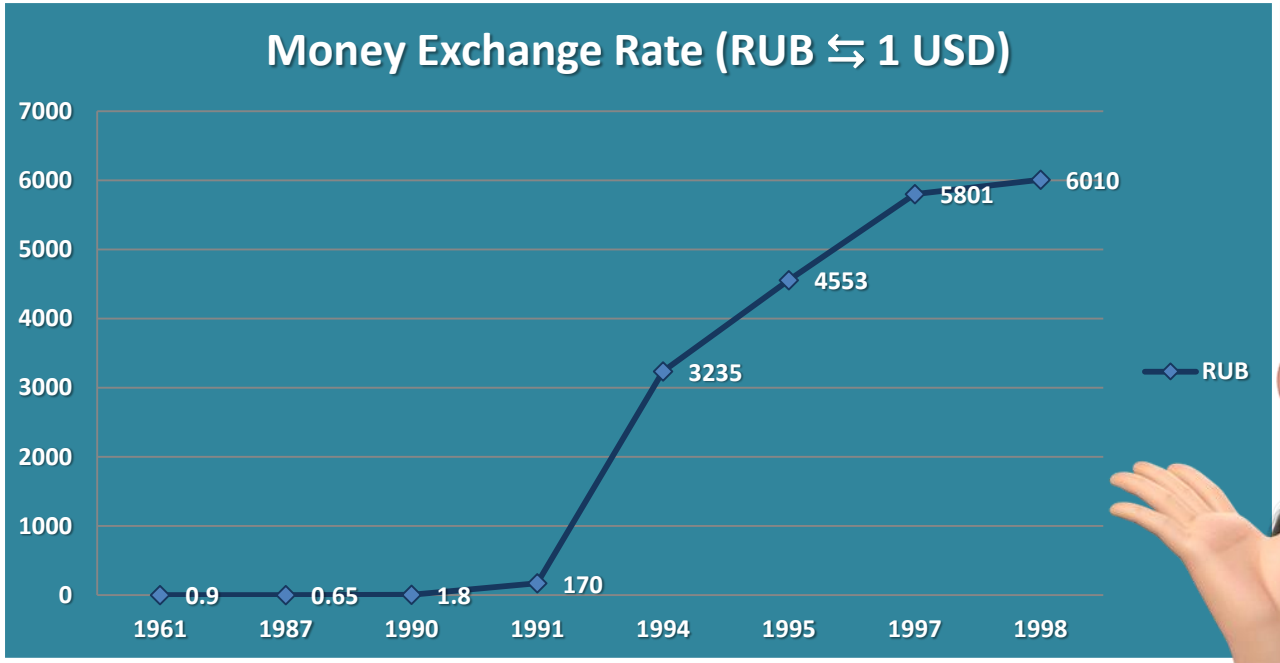
Bretton Woods, 1944

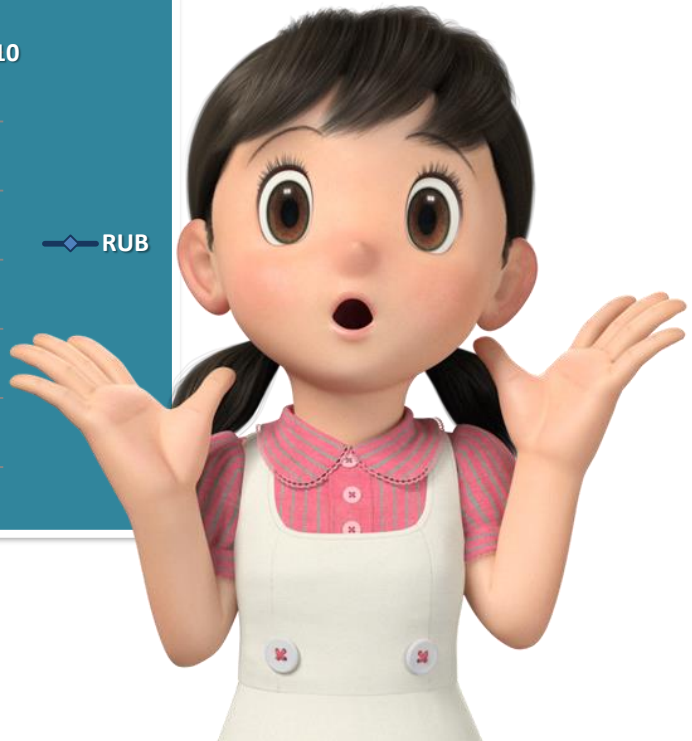
Why Data Analysis

Political events impact on economic trends



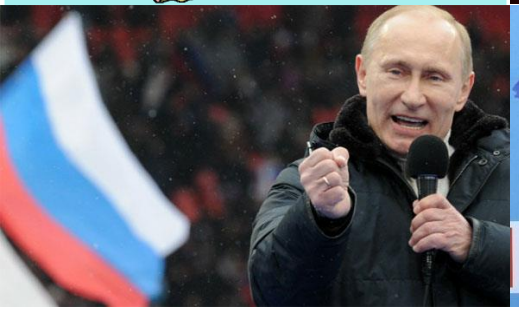
Why Data Analysis





Why Data Analysis

http://cn.investing.com/



卢布贬值引发代购潮，俄罗斯华人大淘便宜货

由于近期卢布的连续贬值，用
消售范围主要围绕化妆、日用

网易新闻 网易首页 > 新闻中心 > 滚动新

卢布大贬值 快去俄罗斯淘

2014-12-28 02:35:41 来源:华商报(西安)

又到岁末，人民币依然维持全年坚挺走势
是著名的金砖国家俄罗斯的卢布跌幅更是大。
少呢。

血拼乐淘：在莫斯科的高档商场Atrium
名牌商品，那里的迪奥、香奈儿、兰蔻等大牌
专柜的雅诗兰黛面霜小棕瓶50ml为例，在莫斯科卢布。
价格为950元；兰蔻一款在我国内地售价1460元的小
约880元，在莫斯科专柜价格不到600元。很多美妆和奢侈品的价格已经便宜过香港，部分品牌
至是全欧洲最便宜的。

腾讯教育 高考 考研 公务员 校园 留学 外语 中小学

卢布贬值震惊全球 中国留学生做起奢侈品

留学 浙江在线-钱江晚报[微博] 2014-12-18 09:35 我要分享

浙江在线12月18日讯：（钱江晚报记者 钟卉）在2014年的最后
场空前巨大的危机中。卢布在本周一暴跌超10%，创下1998年来最大
斯央行一口气加息650个基点，从10.5%提高至17%，仍拦不住卢布的
贬值25%。

受国际油价持续走低、西方对俄施加经济制裁、俄国际储备处
合作用，今年以来卢布贬值幅度已经超过50%。危机当前，一惯强
布“寒潮”中的俄罗斯民众已无法冷静。俄罗斯街头出现这样的场
数字不停地变化，银行提款机经常被提取一空，俄罗斯人冲进商场

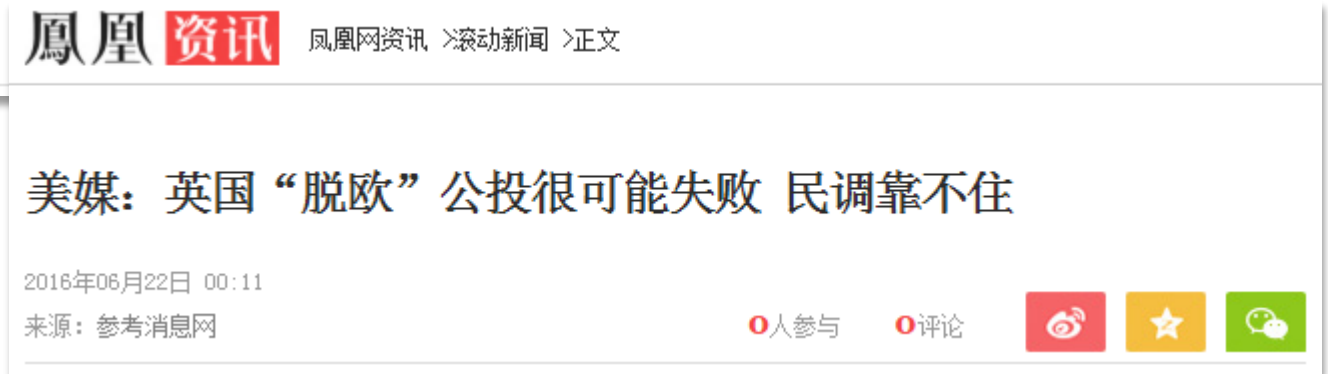
上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

Why Data Analysis

News and New Media are crucial to the investment.

Why Data Analysis

Why Data Analysis



Why Data Analysis

June 23, 2016

Why Data Analysis

Real time data analysis based on news and my EURO currency exchange

http://finance.qq.com/zt2016/gongtou/index.htm

Time	News	Rate
06-24 06:17	Gibraltar (IN)	1:7.52
06-24 06:31	New Castle (IN)	1:7.51
06-24 07:28	Sunderland (OUT)	1:7.47
06-24 09:23	Oxford (IN), North Ireland (IN)	1:7.43
06-24 10:01	49.79% (IN), 50.21% (OUT)	1:7.30
06-24 11:07	48.95% (IN), 51.05% (OUT)	1:7.25
06-24 11:40	Wales (OUT)	1:7.22
06-24 12:13	48.28% (IN), 51.72% (OUT), 339/382 Regions	1:7.24

Why Data Analysis

Final Currency Exchange Results:

(7.52-7.24)*2200=616 RMB

Why Data Analysis

Do you have a similar experience using data analysis based on news?

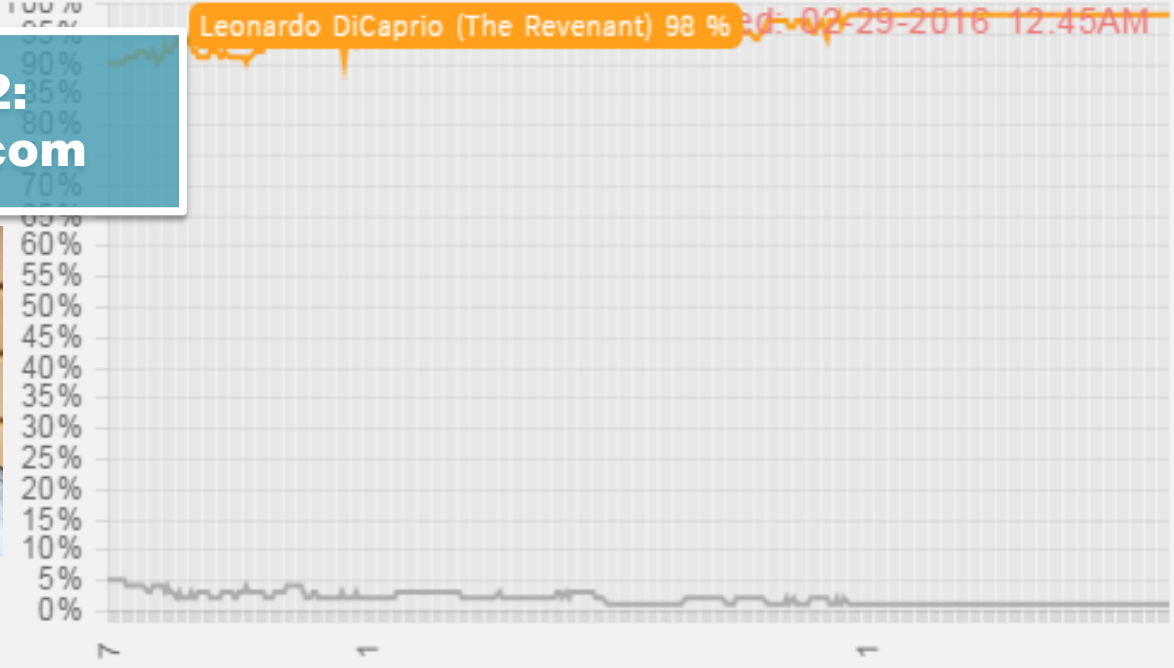
Why Data Analysis

Leading Actor

Leonardo DiCaprio (The Revenant)

Eddie Redmayne (The D...

EXAMPLE 2:
predictwise.com



Why Data Analysis

David Rothschild,

PhD of Wharton School of Business at the University of Pennsylvania

Microsoft researcher at Microsoft Research in New York City

He correctly predicted 50 of 51 Electoral College outcomes in February of 2012, average of 20 of 24 Oscars from 2013-5, and 15 of 15 knockout games in the 2014 World Cup.

- POLITICS
- SPORTS
- ENTERTAINMENT
- ECONOMIC/FINANCIAL

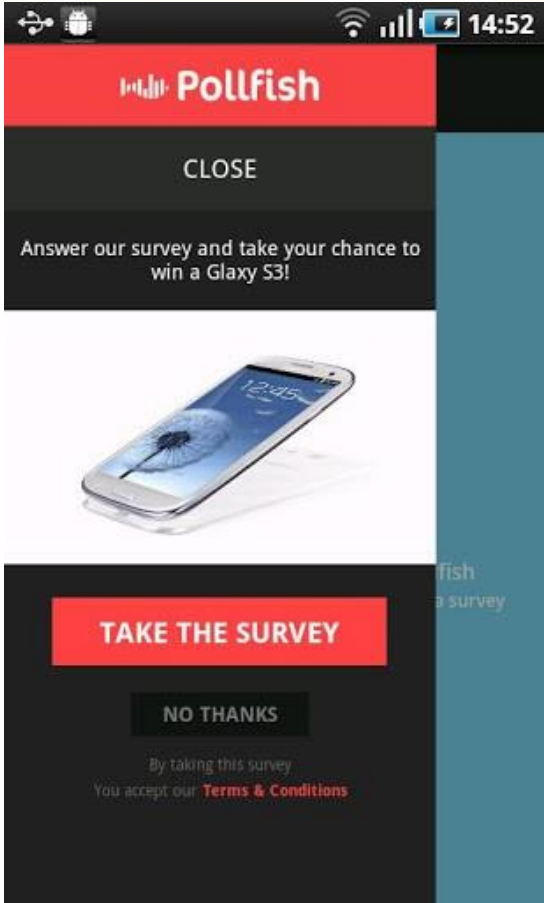
Why Data Analysis

Approaches

- Data Collection:
  Pollfish, MSN, Xbox

- Data Analysis:
  Statistical Analysis based on Historical Data

http://predictwise.com/

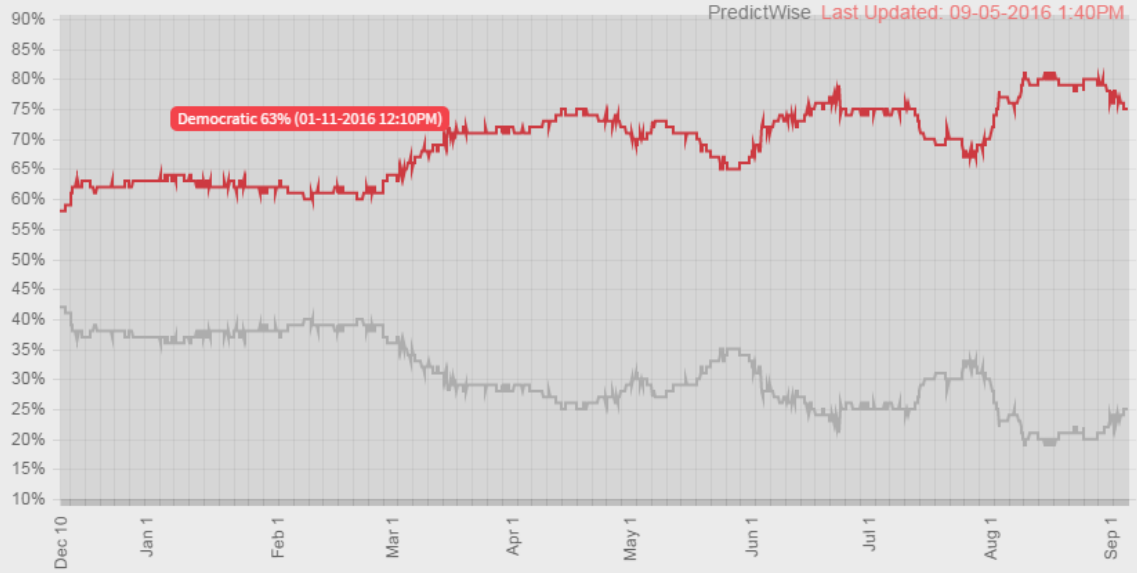


Why Data Analysis

Politics

2016 PRESIDENT - GENERAL ELECTION

Democratic	75 %
Republican	25 %

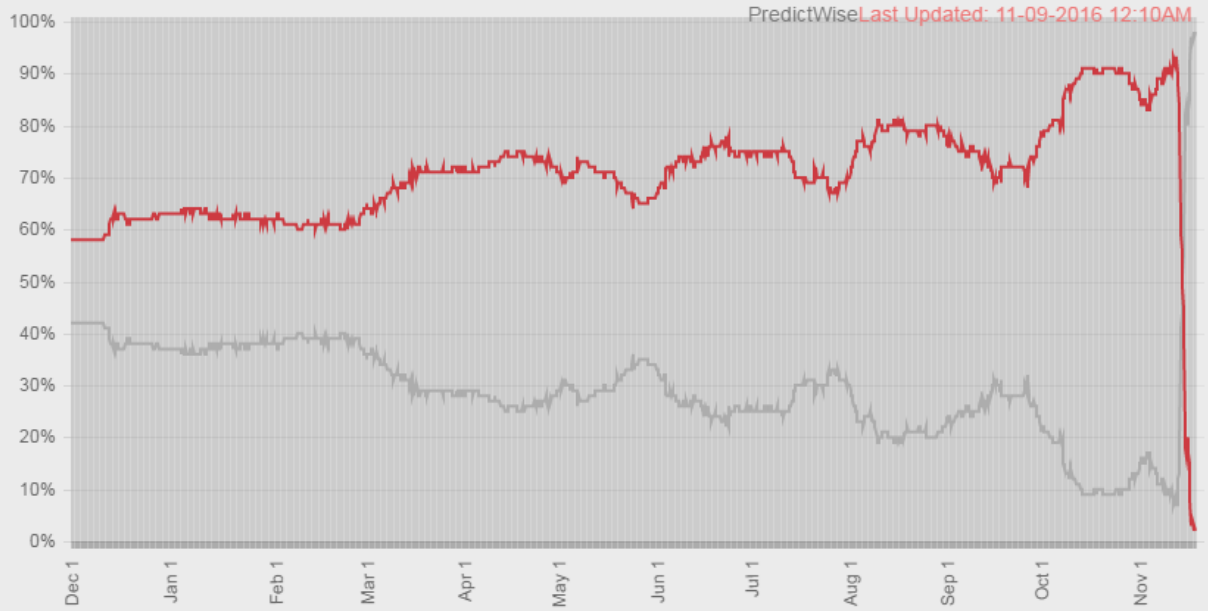


Why Data Analysis

Politics

2016 PRESIDENT - GENERAL ELECTION

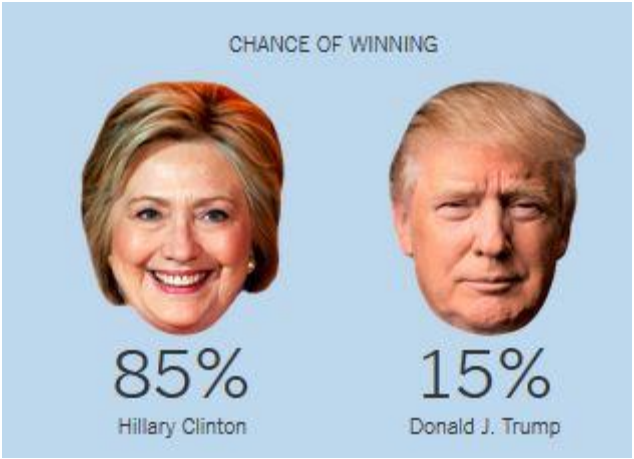
Republican	98 %
Democratic	2 %



Why Data Analysis

1. Examine data quality - in this election polls were not reaching all likely voters

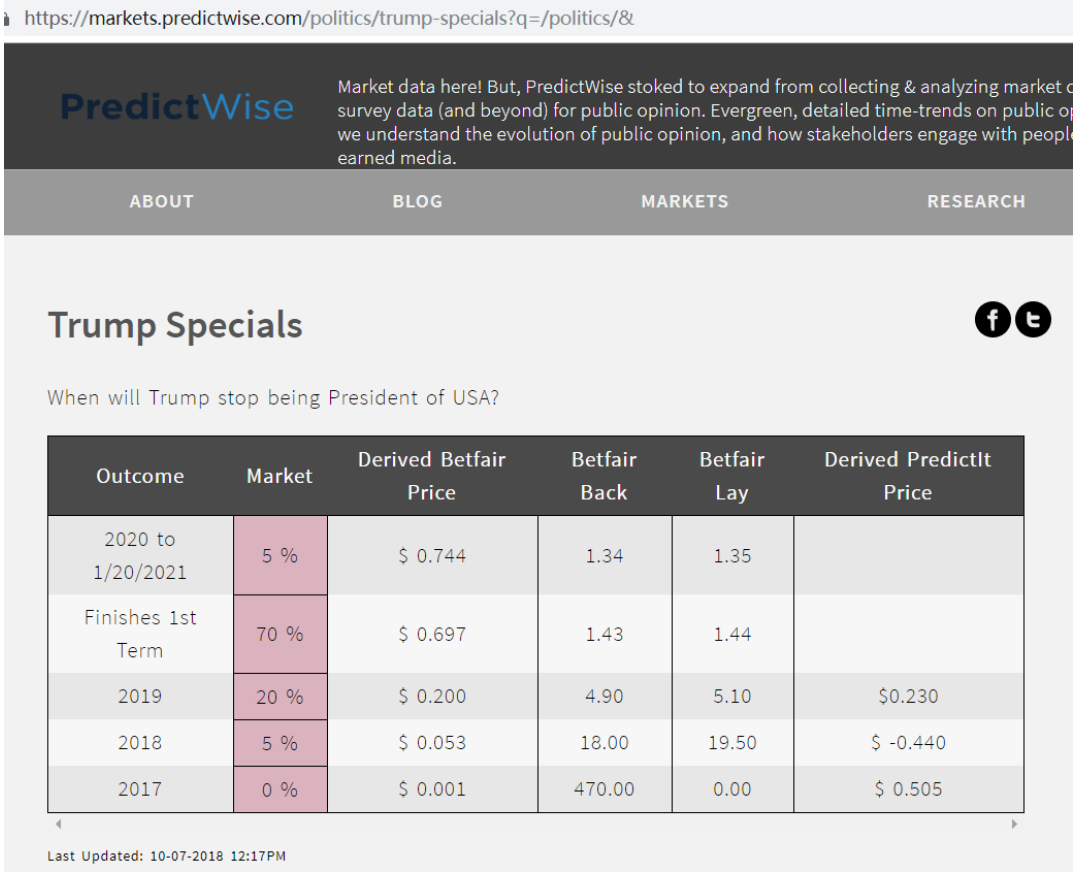
2. Beware of your own biases: many pollsters were likely Clinton supporters and did not want to question the results that favored their candidate. For example, Huffington Post had forecast 98% chance of Clinton Victory.



http://www.kdnuggets.com/2016/11/trump-shows-limits-prediction.html

Why Data Analysis

What is your opinion?



Why Data Analysis

PredictWise's Oscars 2019: Market Predictions

Picture		Director		Best Actor		Best Actress		Supporting Actor		Supporting Actress		Original Screenplay		Adapted Screenplay	
Roma	62%	Alfonso Cuarón, "Roma"	90%	Rami Malek, "Bohemian Rhapsody"	85%	Glenn Close, "The Wife"	84%	Mahershala Ali, "Green Book"	84%	Regina King, "If Beale Street Could Talk"	71%	The Favourite	69%	BlackKKlansman	74%
Green Book	26%	Spike Lee, "BlacKkKlansman"	6%	Christian Bale, "Vice"	12%	Olivia Colman, "The Favourite"	11%	Richard E. Grant, "Can You Ever Forgive Me?"	14%	Amy Adams, "Vice"	13%	Green Book	21%	If Beale Street Could Talk	12%
The Favourite	3%	Pawel Pawlikowski, "Cold War"	2%	Bradley Cooper, "A Star Is Born"	1%	Lady Gaga, "A Star Is Born"	4%	Sam Elliott, "A Star Is Born"	1%	Rachel Weisz, "The Favourite"	12%	Roma	3%	Can You Ever Forgive Me?	10%
BlackKKlansman	3%	Adam McKay, "Vice"	1%	Viggo Mortensen, "Green Book"	1%	Yalitza Aparicio, "Roma"	1%	Sam Rockwell, "Vice"	1%	Marina de Tavira, "Roma"	3%	First Reformed	4%	A Star Is Born	3%
Bohemian Rhapsody & Black Panther	2%	Yorgos Lanthimos, "The Favourite"	1%	Willem Dafoe, "At Eternity's Gate"	0%	Melissa McCarthy, "Can You Ever Forgive Me?"	0%	Adam Driver, "BlacKkKlansman"	0%	Emma Stone, "The Favourite"	2%	Vice	3%	The Ballad of Buster Scruggs	1%

Animated Feature		Animated Short		Foreign Language		Documentary Feature		Documentary Short		Live Action Short		Original Song		Original Score	
Spiderman-Into the SpiderVerse	90%	Bao	76%	Roma	85%	Free Solo	47%	Peroid. End of Sentence.	49%	Marguerite	50%	Shallow -A Star Is Born	82%	If Beale Street Could Talk	51%
Ralph Breaks the Internet	3%	Weekends	15%	Capernaum	3%	RBG	42%	Black Sheep	33%	Skin	25%	T.P.W.L.T.G-M Poppins Returns	4%	Black Panther	22%
Mirai	3%	Late Afternoon	4%	Never Look Away	3%	Minding the Gap	7%	End Game	10%	Detainment	9%	I'll Fight- RBG	4%	BlackKKlansman	12%
Incredibles 2	2%	Animal Behaviour	2%	Cold War	8%	Of Fathers and Sons	3%	Lifeboat	4%	Mother	8%	All the Stars - Black Panther	4%	Isle Of Dogs	8%
Isle Of Dogs	2%	One Small Step	2%	Shoplifters	2%	Hale County	2%	A Night at The Garden	4%	Fauve	8%	W.A.C.T.H.S.F.W-Buster Scruggs	4%	Mary Poppins Returns	7%

Cinematography		Costume Design		Film Editing		Makeup and Hairstyling		Production Design		Sound Editing		Sound Mixing		Visual Effects	
Roma	84%	The Favourite	50%	Vice	51%	Vice	82%	The Favourite	47%	First Man	42%	Bohemian Rhapsody	57%	Avengers: Infinity War	60%
Cold War	3%	Black Panther	32%	Bohemian Rhapsody	29%	Border	13%	First Man	9%	A Quiet Place	27%	A Star Is Born	35%	Solo: A Star Wars Story	10%
A Star Is Born	9%	The Ballad of Buster Scruggs	8%	The Favourite	14%	Mary Queen of Scots	5%	Mary Poppins Returns	9%	Bohemian Rhapsody	22%	First Man	5%	First Man	10%
The Favourite	3%	Mary Poppins Returns	8%	BlacKkKlansmen	6%			Black Panther	28%	Black Panther	5%	Black Panther	3%	Christopher Robin	10%
Never Look Away	2%	Mary Queen of Scots	2%	Green Book	0%			Roma	8%	Roma	4%	Roma	1%	Ready Player One	10%

Based on Betfair Market Prices at 8:00 PM ET on Feburary 23, 2019

the definition of data analysis for journalism

What is Data Analysis

What is Data Analysis

The significance of Data Analysis

1. *To obtain new information*
2. *To enlarge the benefits*
3. *To avoid the risks*

What is Data Analysis

INFORMATION DISCOVERY
CONCLUSION SUGGESTING
DECISION SUPPORT

are three objectives of data analysis

What is Data Analysis

Definition:

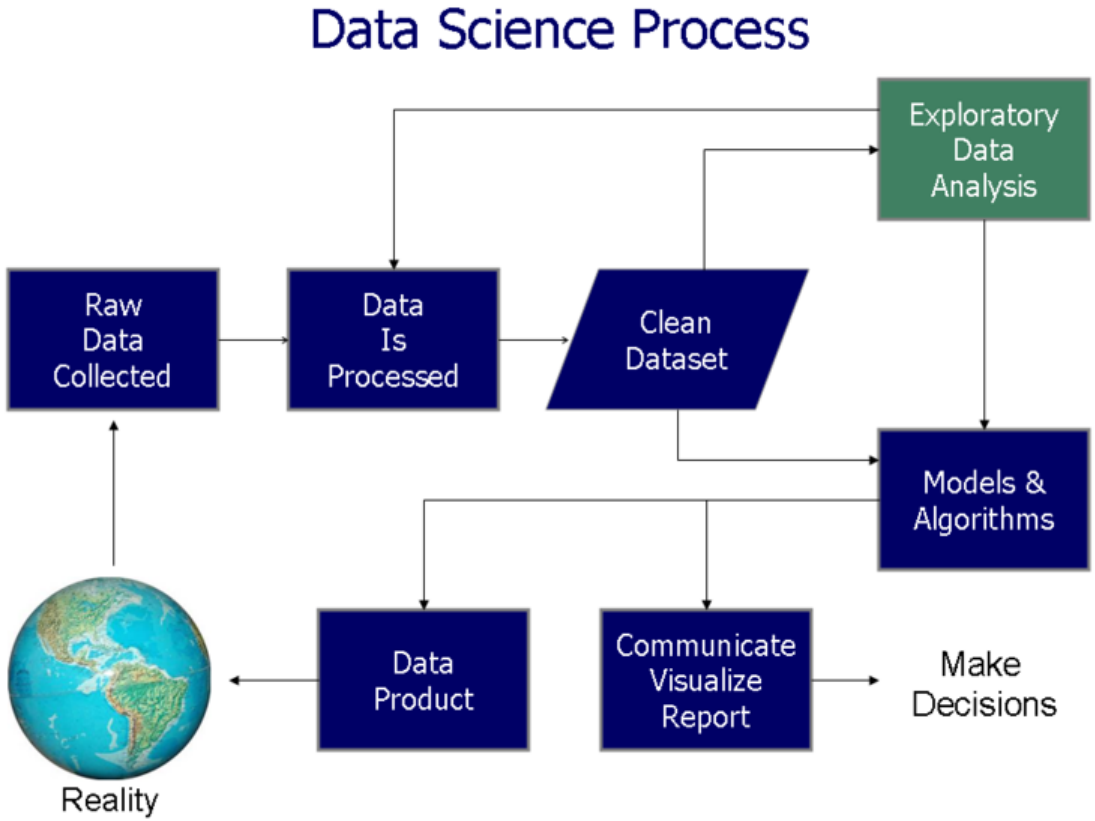
https://en.wikipedia.org/wiki/Data_analysis

Analysis of data *is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.*

What is Data Analysis

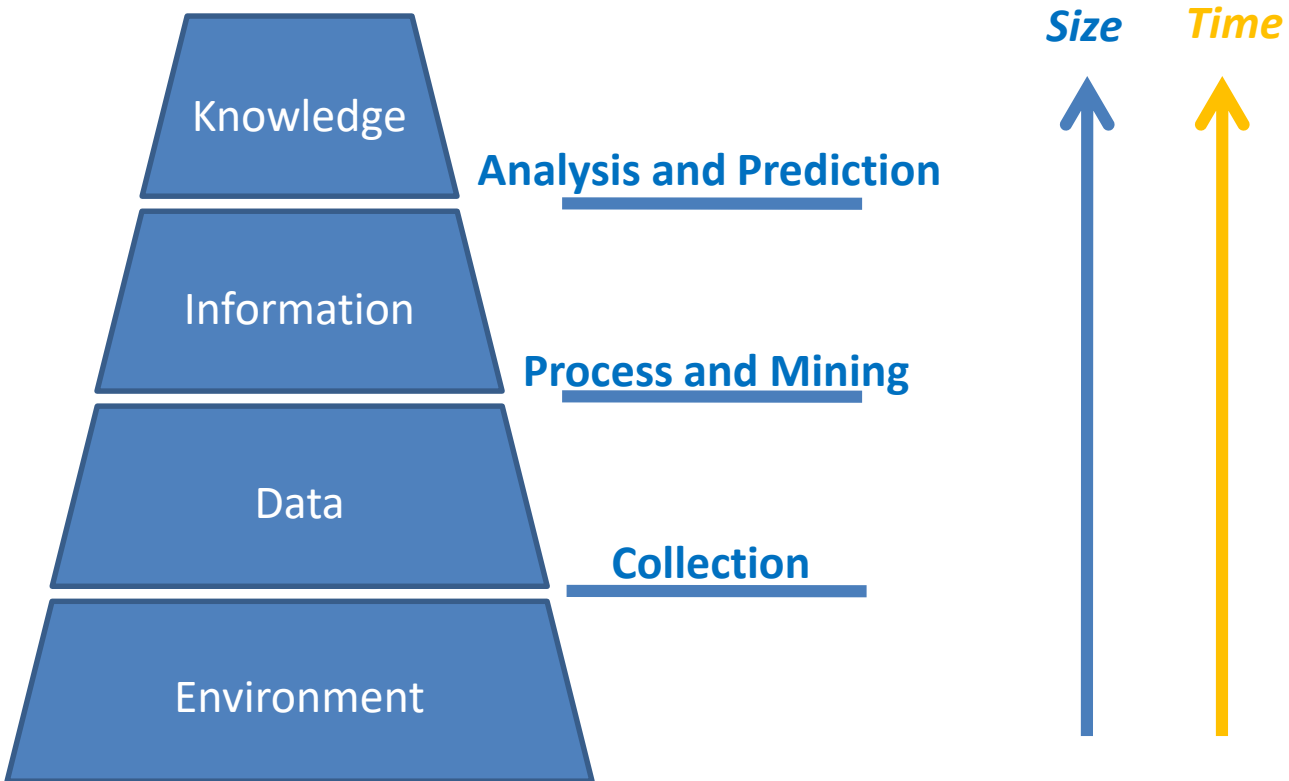
- Two types of decisions:
  - Quantitative Decision with a value
    - Prediction, Regression
  - Qualitative Decision with a label
    - Classification, Clustering

What is Data Analysis



What is Data Analysis

Relationship between data, information and knowledge



What is Data Analysis

Methodology

1 *Information Acquisition*

2 *Data Cleaning and Information Retrieval*

3 *Knowledge Fusion and Information Updating*

4 *Prediction or Classification*

Web Data → *Structured Web Information* → *Feature Analysis and Updating* → *Personalized Products or Services*

What is Data Analysis

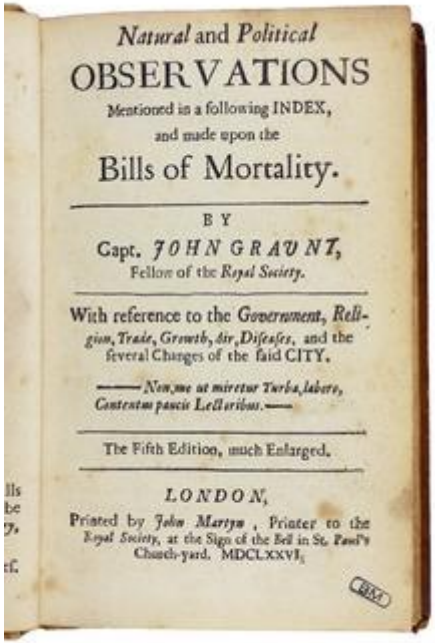
Do you have some methods to forecast the auto sale market in the next several months?

上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

data analysis will be everywhere in the future

The History of Computer Data Analytics

The History of Computer Data Analytics



Natural and *Political*

OBSERVATIONS

Mentioned in a following INDEX,
and made upon the

Bills of Mortality.

BY

Capt. JOHN GRAUNT,
Fellow of the *Royal Society*.

With reference to the *Government*, *Religion*, *Trade*, *Growth*, *Air*, *Diseases*, and the several Changes of the said CITY.

——*Non me ut miretur Turba, laboro,*
Contentus paucis Lectoribus.——

The Fifth Edition, much Enlarged.

LONDON,

Printed by *John Martyn*, Printer to the *Royal Society*, at the Sign of the Bell in St. *Paul's* Church-yard. MDCLXXVI.

BM

社会统计1662

John Graunt (24 April 1620 – 18 April 1674) used statistical analysis to predict the onset and spread of bubonic plague in London, which led him to the Royal Society.

The History of Computer Data Analytics

Thomas Bayes (1701-1761)

贝叶斯决策理论 1763

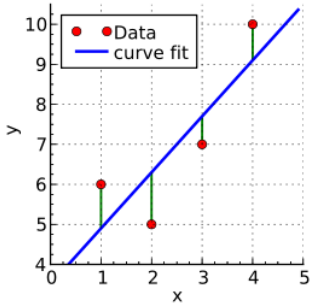
Bayes, Thomas; Price, Mr. (1763). "An Essay towards solving a Problem in the Doctrine of Chances.《机会问题的解法》". Philosophical Transactions of the Royal Society of London. 53 (0): 370–418. doi:10.1098/rstl.1763.0053

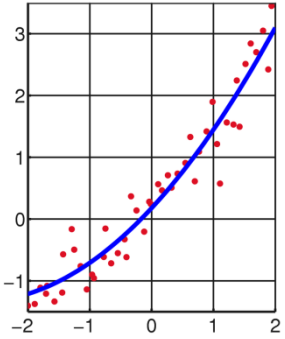
The History of Computer Data Analytics

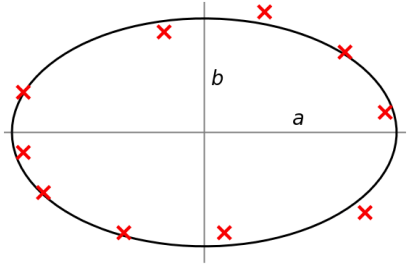
Carl Friedrich Gauss (1777–1855)

最小二乘法 1805

Least squares for data fitting and regression





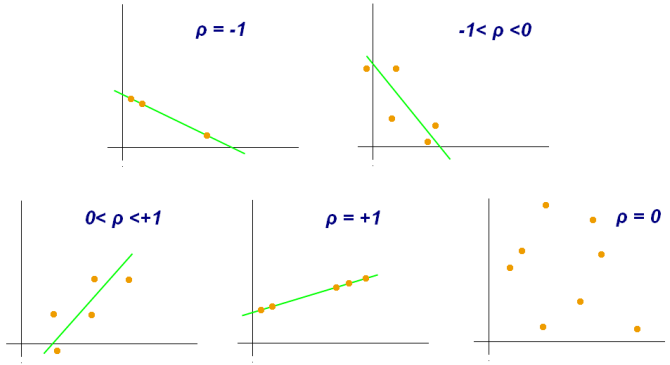


The History of Computer Data Analytics

Karl Pearson
(27 March 1857 – 27 April 1936)

相关系数 1880

Pearson Correlation Coefficient

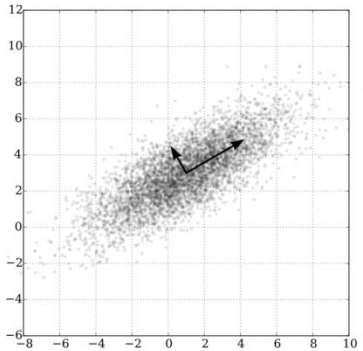


The History of Computer Data Analytics

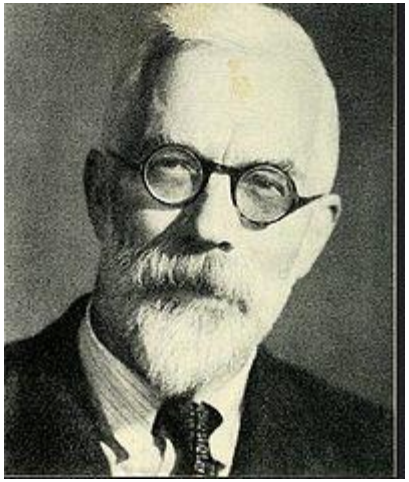
Karl Pearson
(27 March 1857 – 27 April 1936)

主成分分析1901

Principal Component Analysis



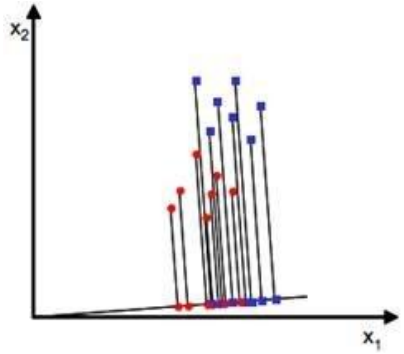
The History of Computer Data Analytics

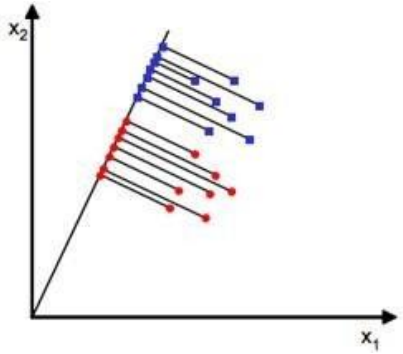


R. A. Fisher
(17 February 1890 – 29 July 1962)

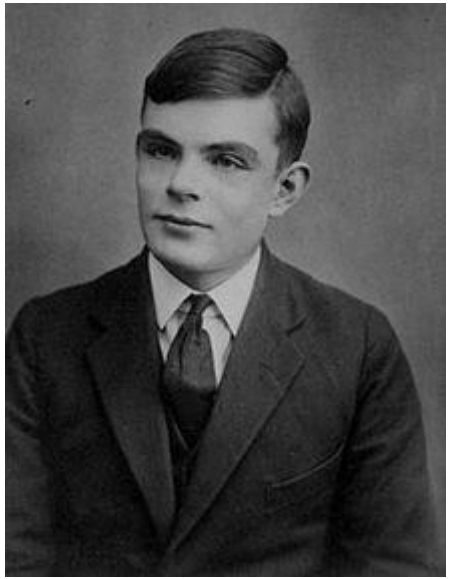
线性判别 1936

Linear Discriminant Aanalysis





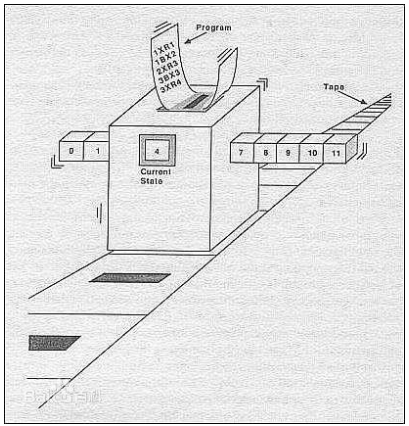
The History of Computer Data Analytics



Alan Turing
(23 June 1912 – 7 June 1954)

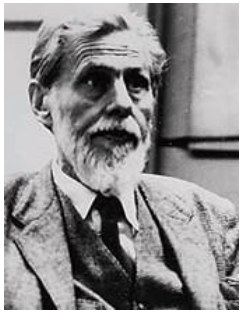
图灵机 1936

Turing Machine



The History of Computer Data Analytics

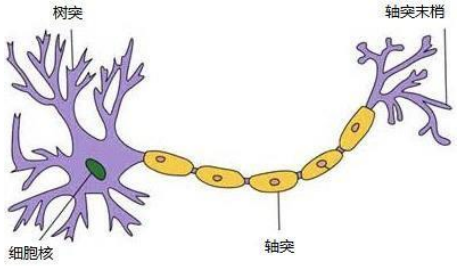
Warren McCulloch



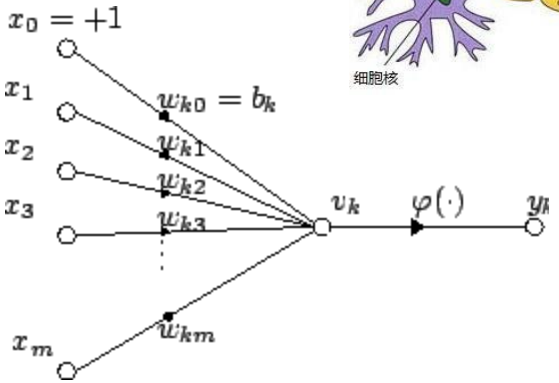
Walter Pitts

人工神经元 1943

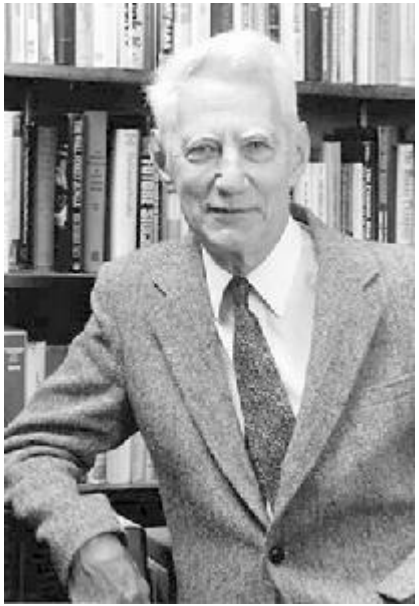
Artificial Neuron



$$y_k = \varphi\left(\sum_{j=0}^{m} w_{kj} x_j\right)$$



The History of Computer Data Analytics



Claude Shannon
(April 30, 1916 – February 24, 2001)

信息论 1948

Information theory

- –Entropy （信息熵）
- –Mutual Information （互信息）

The History of Computer Data Analytics

2006, 50 Years Anniversary

人工智能 1956

Summer Research Project on Artificial Intelligence, Dartmouth College

The History of Computer Data Analytics

Prof. Pedro Domingos
University of Washington

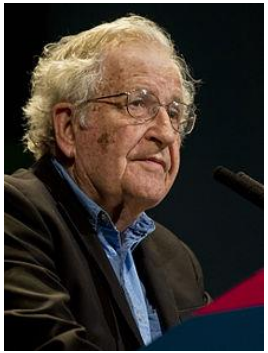
2015, ACM

1. 符号主义
2. 联结主义
3. 进化主义
4. 贝叶斯主义
5. 类推主义

Five Tribes in AI

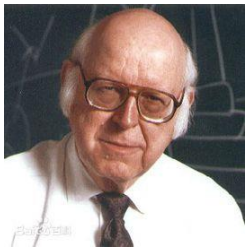
Tribe	Origins	Master Algorithm
Symbolists	Logic, philosophy	Inverse deduction
Connectionists	Neuroscience	Backpropagation
Evolutionaries	Evolutionary biology	Genetic programming
Bayesians	Statistics	Probabilistic inference
Analogizers	Psychology	Kernel machines

The History of Computer Data Analytics



Avram Noam Chomsky

Herbert Simon



Allen Newell

符号主义 1957
Symbolism

1. Plato is a man.
2. Man will die.
3. Plato will die.

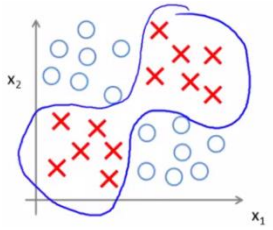
Expert System
Universal Grammar and Chomsky Hierarchy

The History of Computer Data Analytics

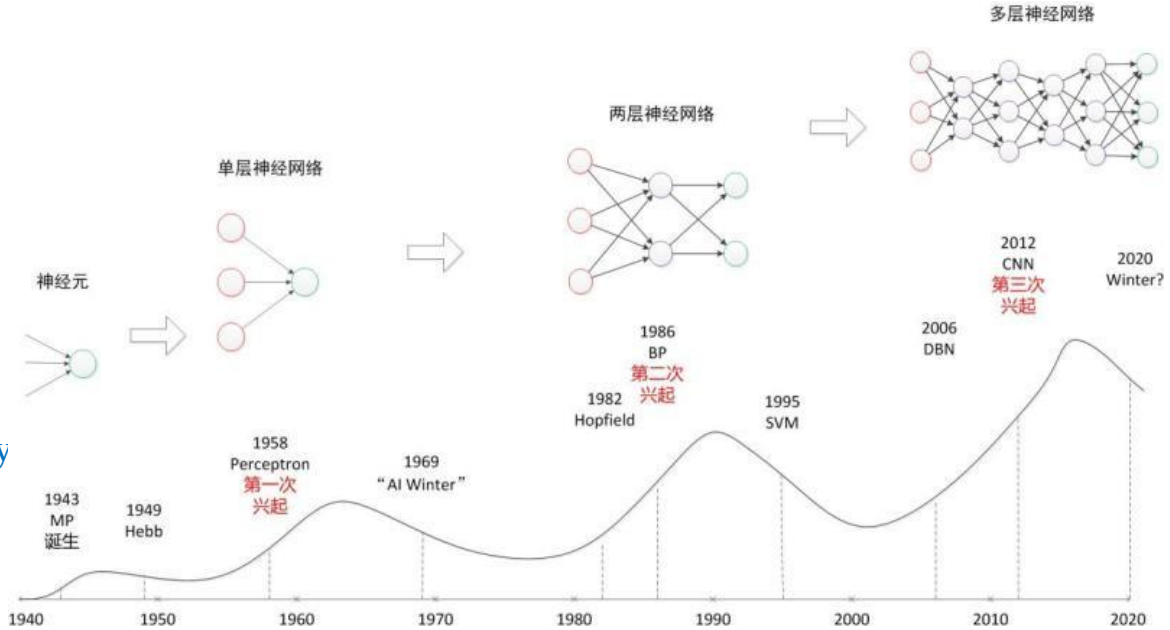
Donald Olding Hebb

Frank Rosenblatt

Marvin Lee Minsky



联结主义 1943 connectionism



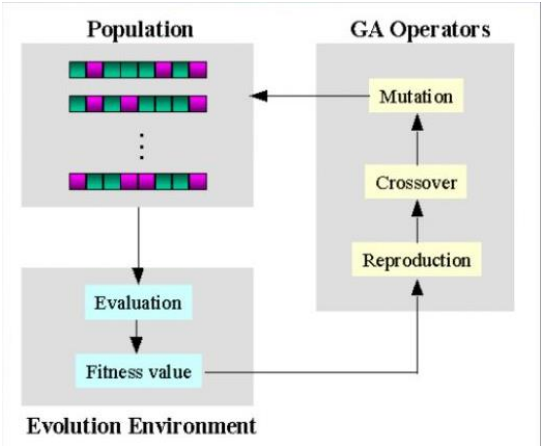
The History of Computer Data Analytics

John Henry Holland

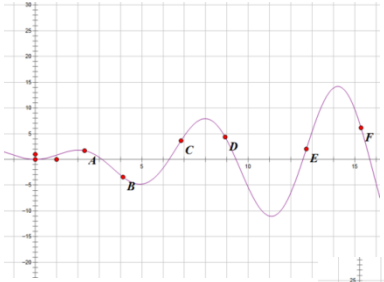
Yuhui Shi

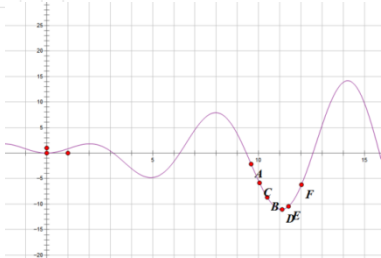
进化主义 1970's
Evolutionism

Genetic Algorithm



Particle Swarm Optimization

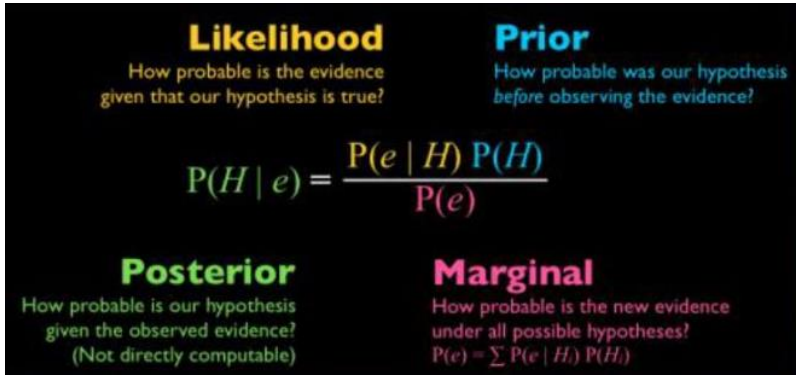




The History of Computer Data Analytics

Judea Pearl

贝叶斯主义 1763
Bayesianism



The History of Computer Data Analytics

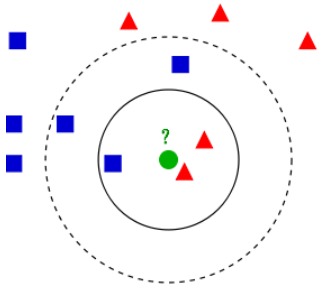
Vladimir Vapnik

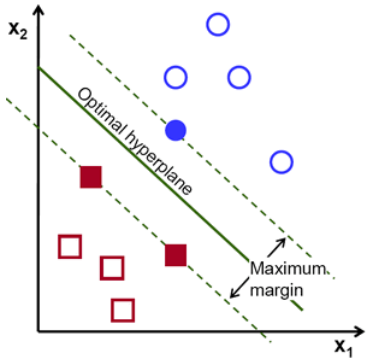
类推主义 1951

Analogism

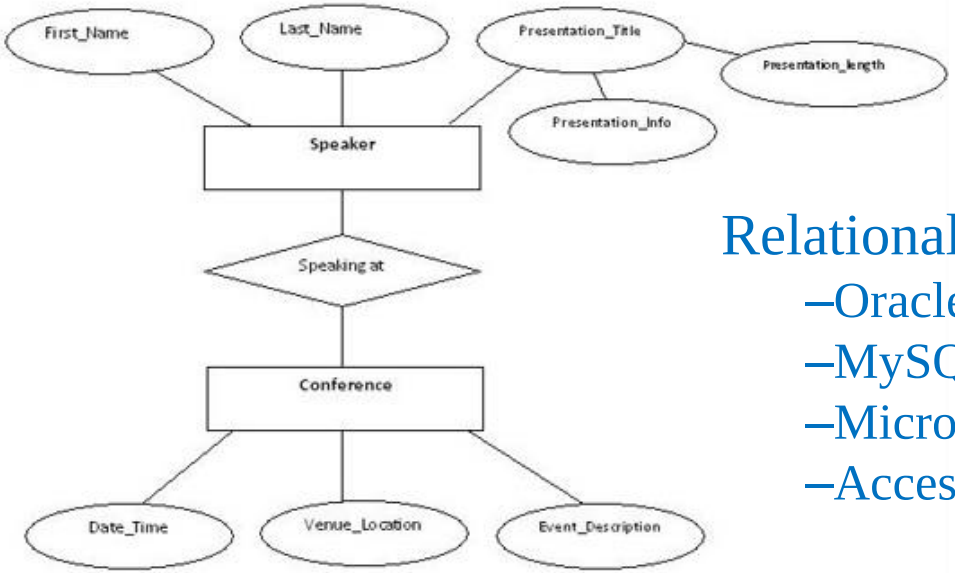
K-Nearest Neighbour

Support Vector Machine





The History of Computer Data Analytics



Entity Relationship Diagram

数据库 1951

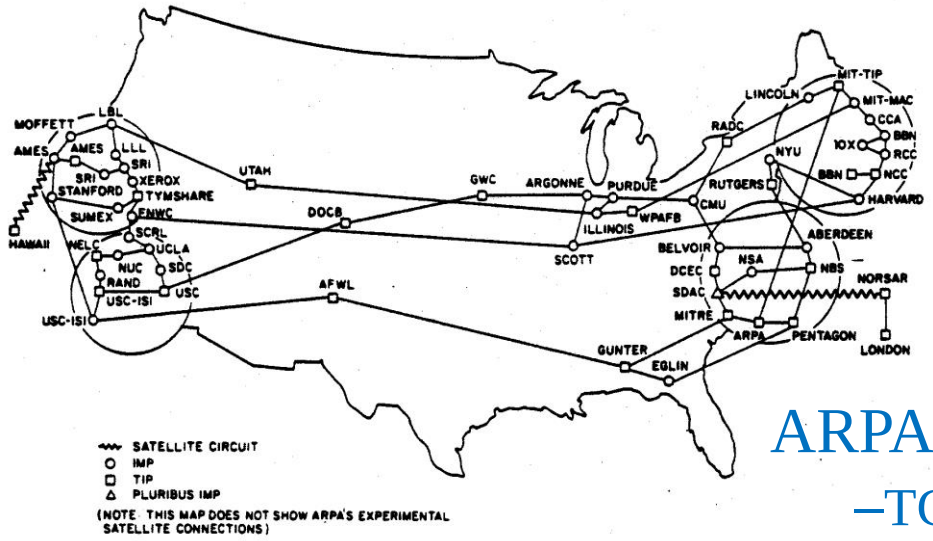
Relational Database:

- –Oracle
- –MySQL
- –Microsoft SQL Server
- –Access

NoSQL:

- – MongoDb

The History of Computer Data Analytics



互联网1969

ARPA Net:

–TCP/IP

The Internet:

–World Wide Web

–Email

The History of Computer Data Analytics

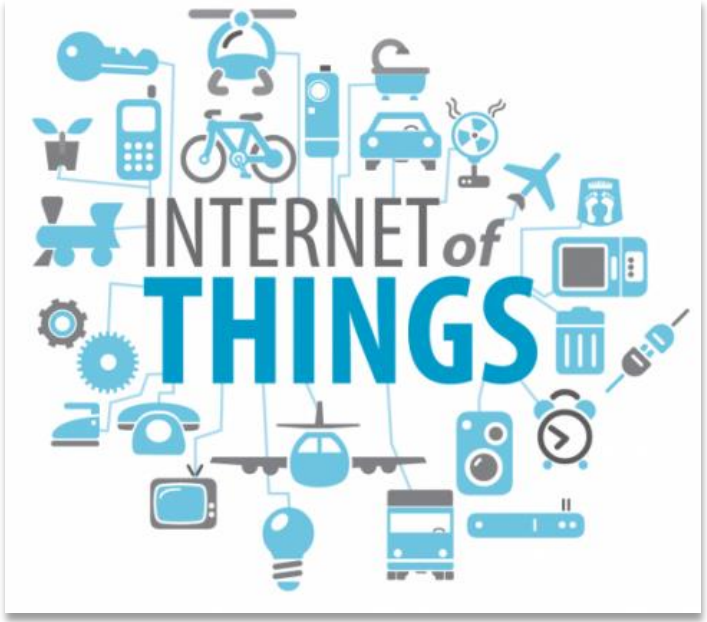
Web2.0 2004

Web 1.0 Contents made by Providers

Web 2.0 Contents made by Customers

Semantic Web: Web 3.0?

The History of Computer Data Analytics



物联网 2005

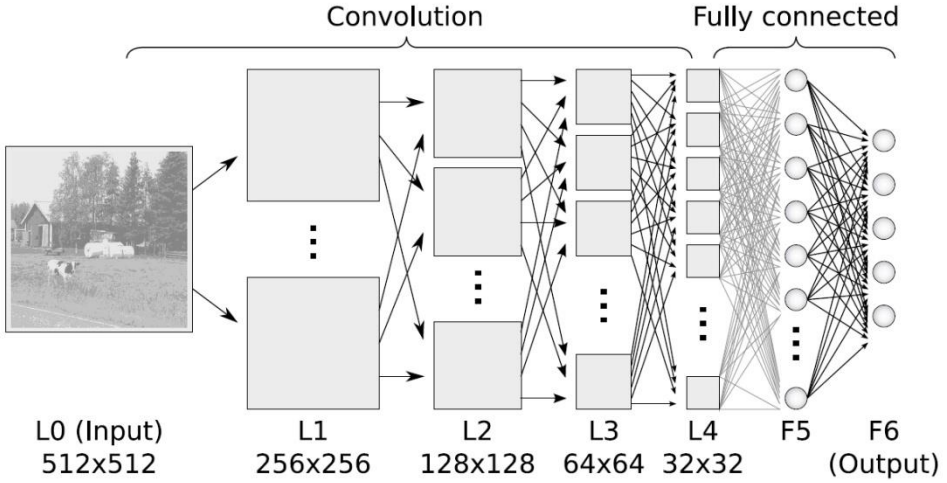
The Internet of Things

The History of Computer Data Analytics

Geoffery Hinton

深度学习 2006

Deep Learning



The History of Computer Data Analytics

云计算2006

Cloud Computing

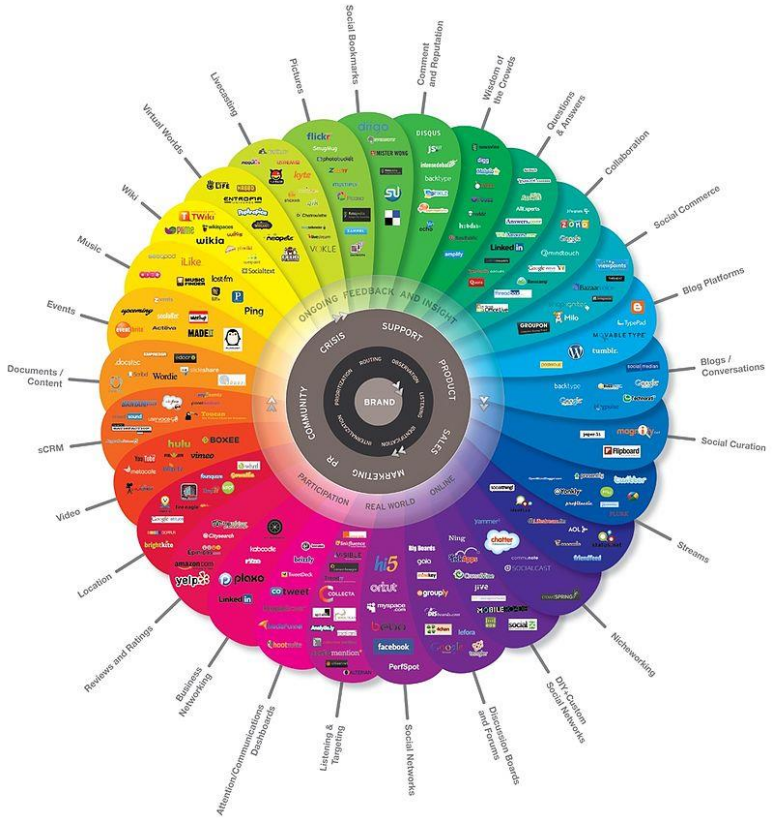
- –Grid Computing
- –Distributed Computing
- –Parallel Computing
- –Utility Computing
- –…

Google

The History of Computer Data Analytics

大数据 2008

The History of Computer Data Analytics

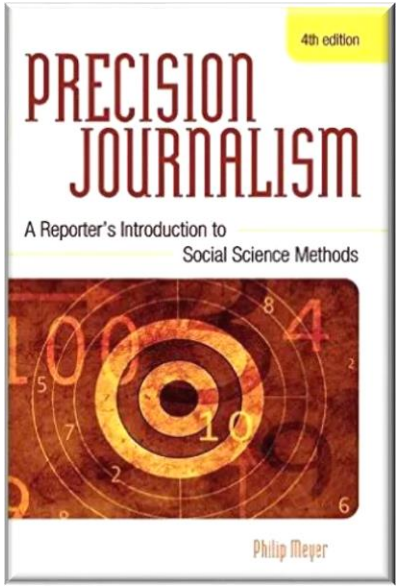


社交媒体的繁荣

2000's

Social Media

The History of Computer Data Analytics



数据新闻 1967

Data Analysis in Journalism

Philip Meye, Precision Journalism, 1970's

The National Institute for Computer-Assisted Reporting – NICAR, 1994

Computational Journalism, Georgia Institute of Technology, 2006

Frontiers of Computational Journalism, Columbia Journalism School, 2012

Masters in Computational Journalism, Syracuse University, 2015

Computational Journalism Lab, Stanford University, 2015

knowledge domains of computational journalism

Domains of New Media Data Analytics

Domains of New Media Data Analytics

Relevant Disciplines

- Journalism
- Computer Science
- Mathematics
- Psychology
- Economics
- Politics
- Linguistics
- ….

Domains of New Media Data Analytics

Corresponding Technologies

- Computer Science
- Artificial Intelligence
- Machine Learning
- Statistical Analysis
- Natural Language Processing
- Pattern Recognition
- …

Domains of New Media Data Analytics

Related Fields

1. Database journalism
2. Computer-assisted reporting
3. Data-driven journalism

Home Work

Home Work

1. Identify at least three major side effects of information sharing on social media.

2. Rumors spread rapidly on social media. Can you think of some methods to block the spread of rumors on social media?

The End of Lecture 1

Thank You

http://www.wangting.ac.cn