Adaptive Business Intelligence: Using Prediction and Optimization for Decision Support

Paulo Cortez

(ALGORITMI Centre, Dep. Information Systems, University of Minho; Machine Learning EPMQ Senior Researcher, Computer Graphics Centre - CCG)

http://www3.dsi.uminho.pt/pcortez

University of Minho

CENTROALGORITMI

Centro de Computação Gráfica

Investigação e Desenvolvimento Tecnológico

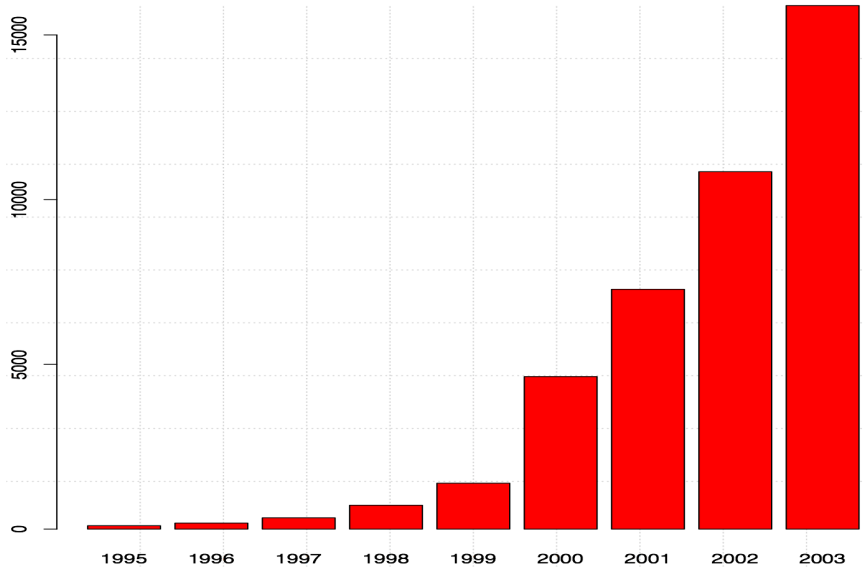
Research and Technological Development

Part1
Adaptive Business Intelligence: Introduction

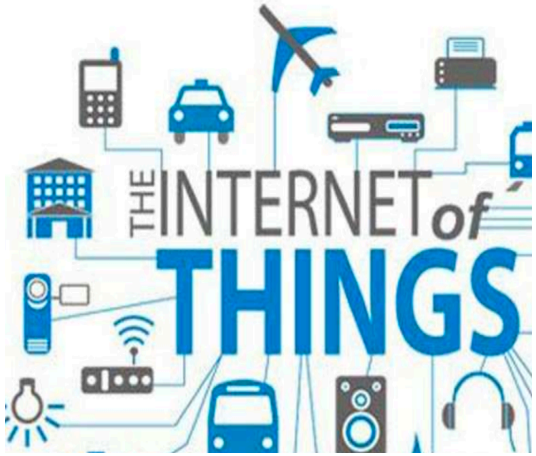
The Rise of **Data**

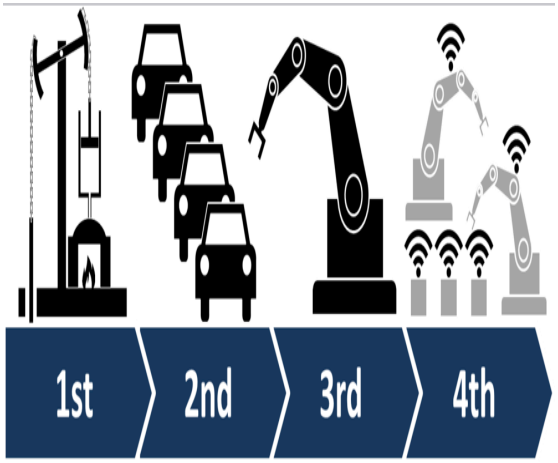
Many sources of data, including

- Web & Social Networks.
- Internet of Things (IoT).
- Industry 4.0.
- Smart Cities.

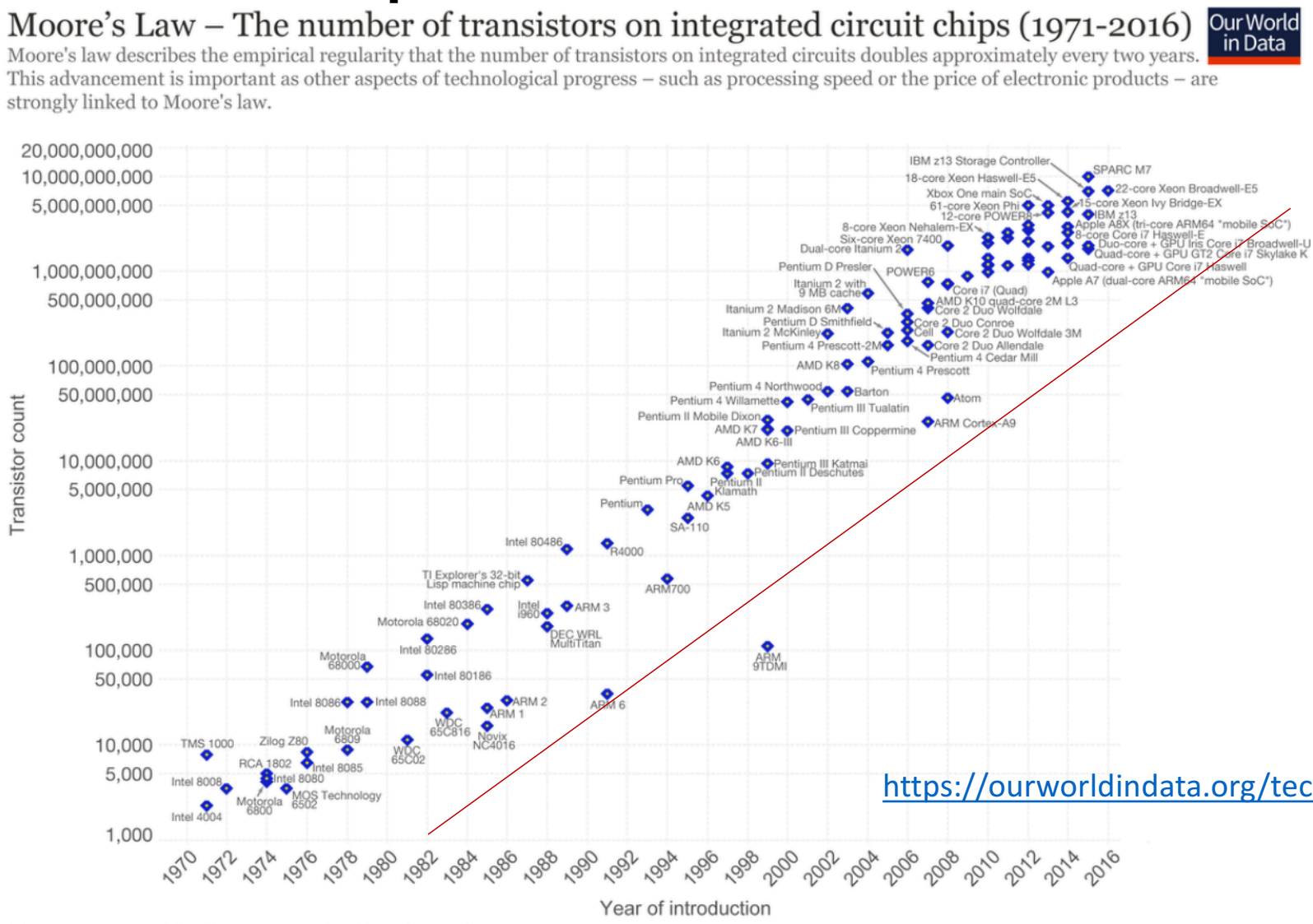


http://www2.sims.berkeley.edu/research/projects/how-much-info-2003





The Rise of **Computational Power**

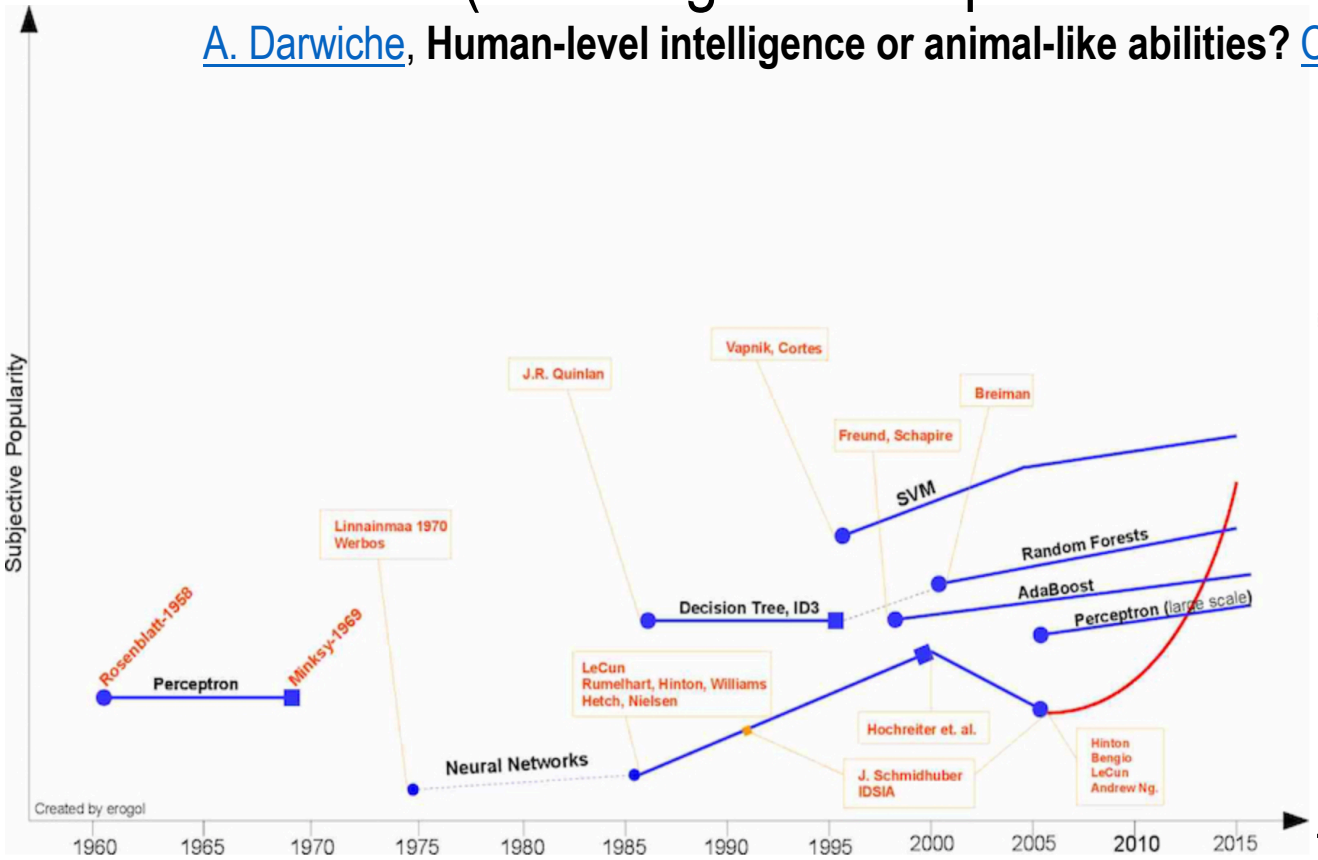


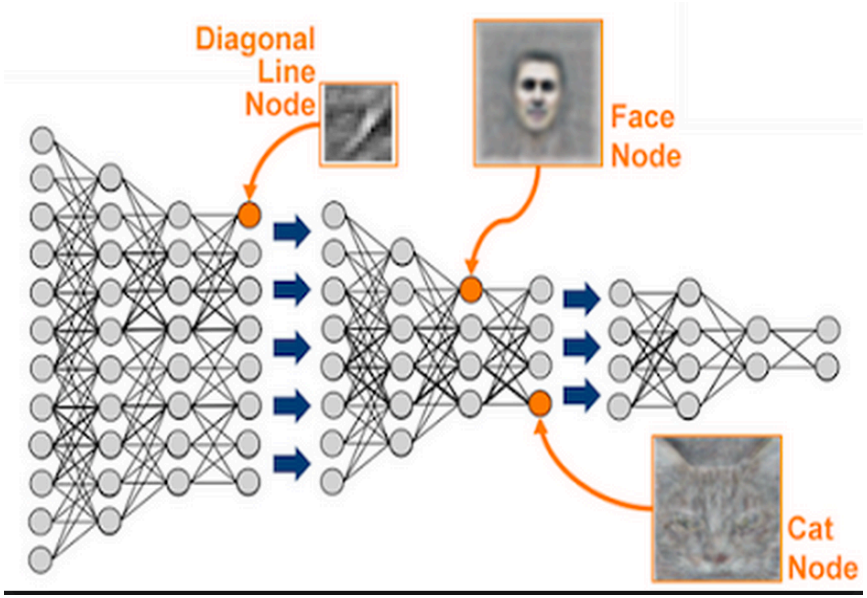
https://ourworldindata.org/technological-progress

The Rise of **Algorithms** (including Machine Learning)

”More sophisticated statistical and optimization techniques for fitting functions (including new/deeper network structures)”

A. Darwiche, **Human-level intelligence or animal-like abilities?** Commun. ACM 61(10): 56-67 (2018)





= The Rise of **Artificial Intelligence (AI)**

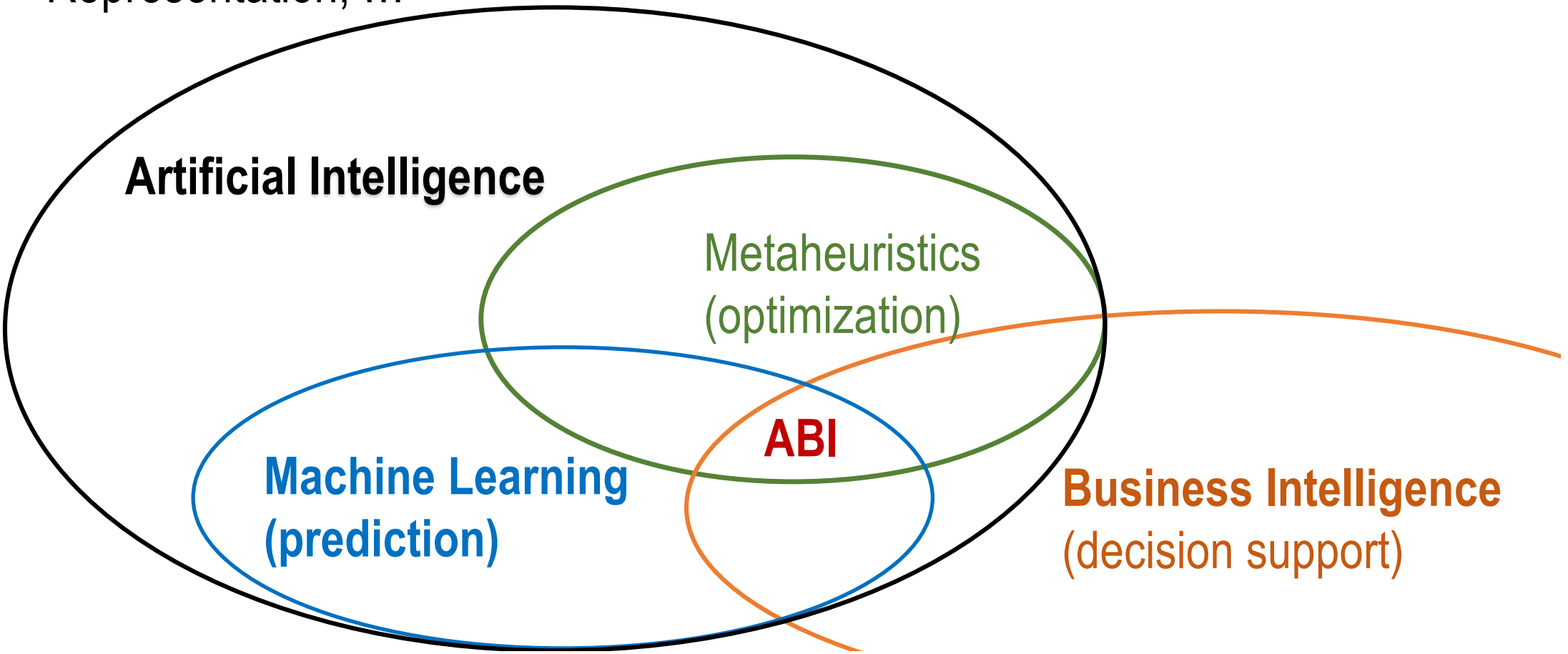
https://ai100.stanford.edu

AI has achieved many remarkable milestones, including:

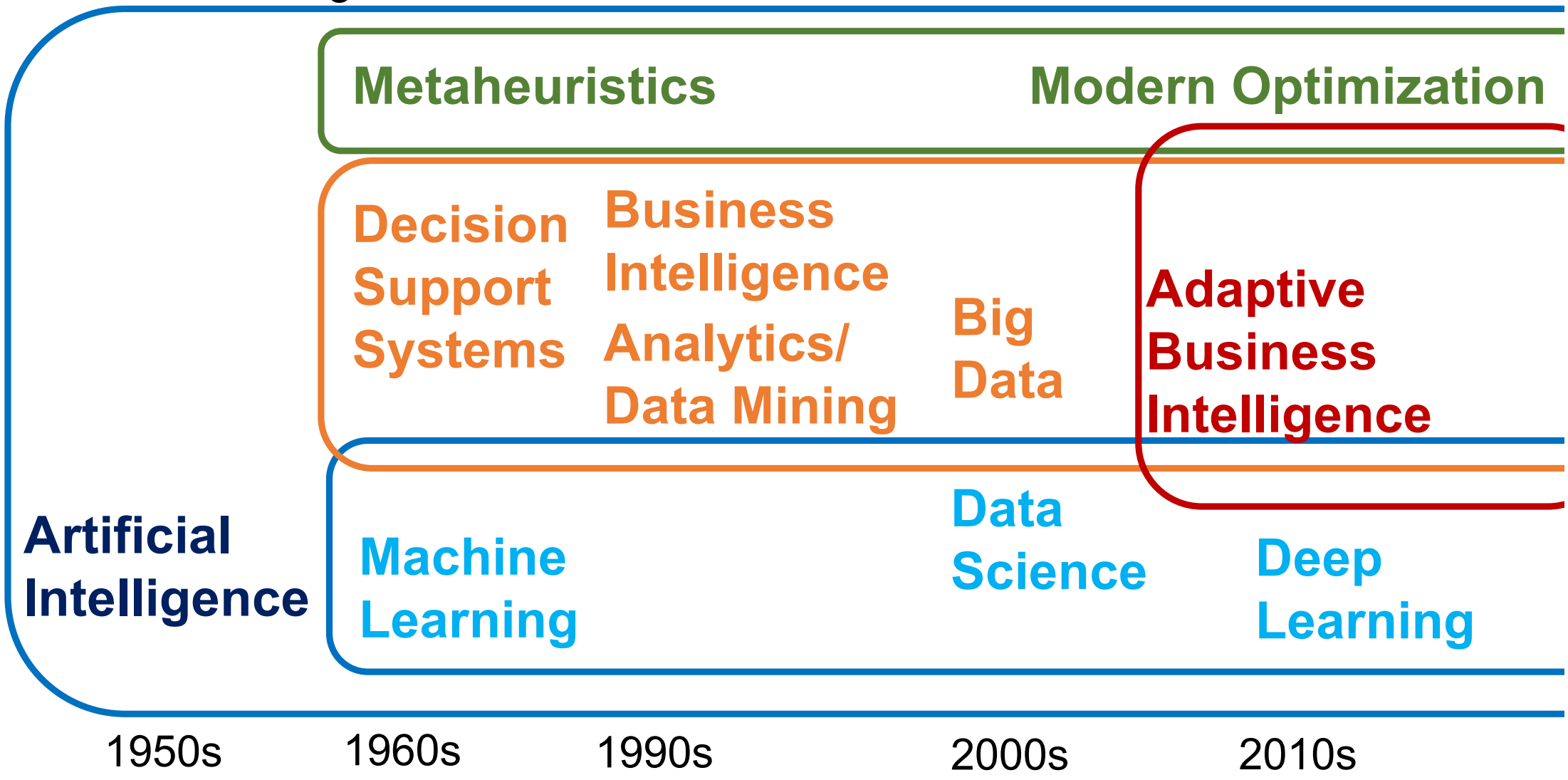
- 1997**:** IBM's Deep Blue beats Garry Kasparov in Chess.
- 2011: IBM's Watson beats two best human players on Jeopardy.
- 2016: Google's AlphaGo wins Korea's Lee Sedol Go player.

Artificial Intelligence (AI) = Data + Algorithms

Includes several subfields: Machine Learning, Metaheuristics, Knowledge Representation, …



Artificial Intelligence and Data Related Terms

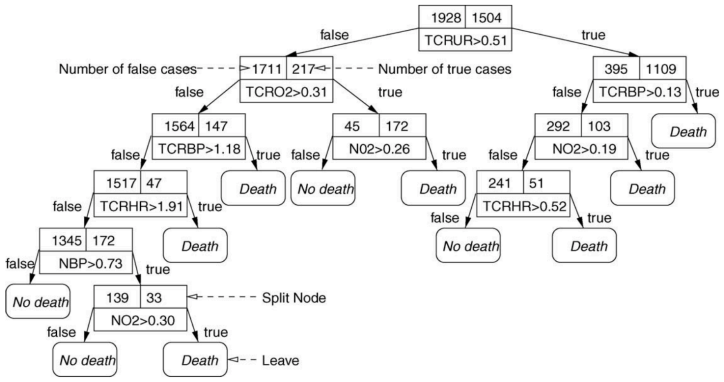


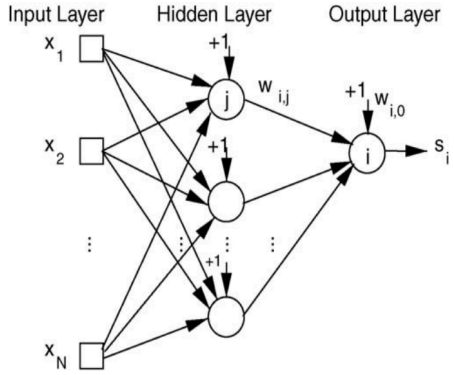
Data-driven **Prediction** (predictive analytics)

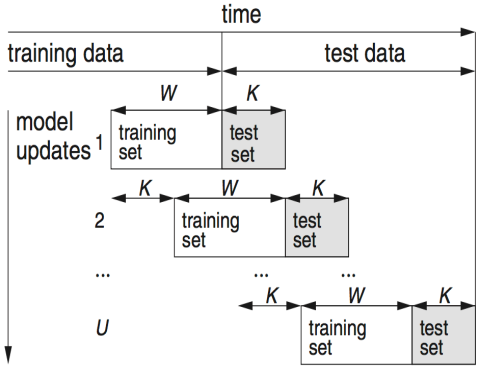
"The ultimate goal of data mining is **prediction** - and **predictive data mining** is the most common type of data mining and one that has the most direct business applications."

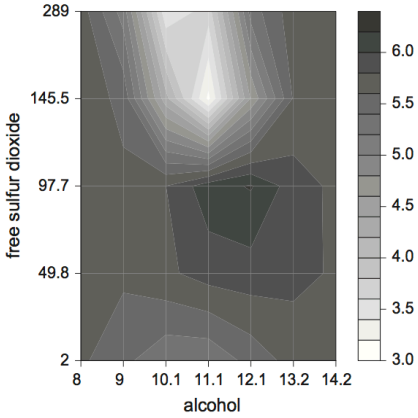
http://www.statsoft.com/textbook/data-mining-techniques

Machine Learning: decision trees, neural networks, ensembles, random forests, support vector machines, deep learning, …







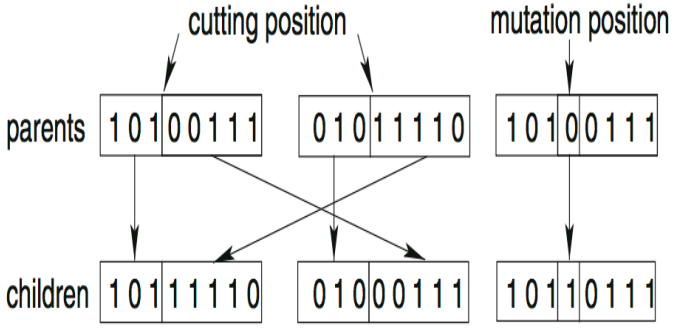


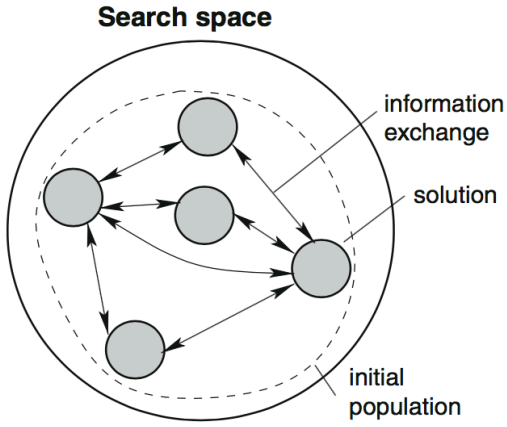
Modern Optimization (prescriptive analytics)

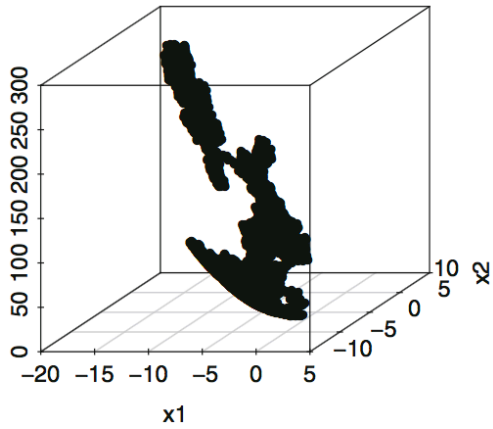
Also known as Metaheuristics

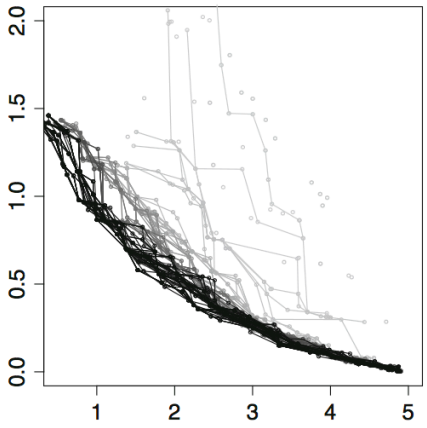
"Related with general purpose solvers that use few domain knowledge, iteratively improving a solution (or population of solutions) to minimize or maximize a goal." http://www.springer.com/gp/book/9783319082622

Metaheuristics: simulated annealing, tabu search, genetic algorithms, genetic programming, multi-objective optimization (e.g., NSGAII), particle swarm optimization, …



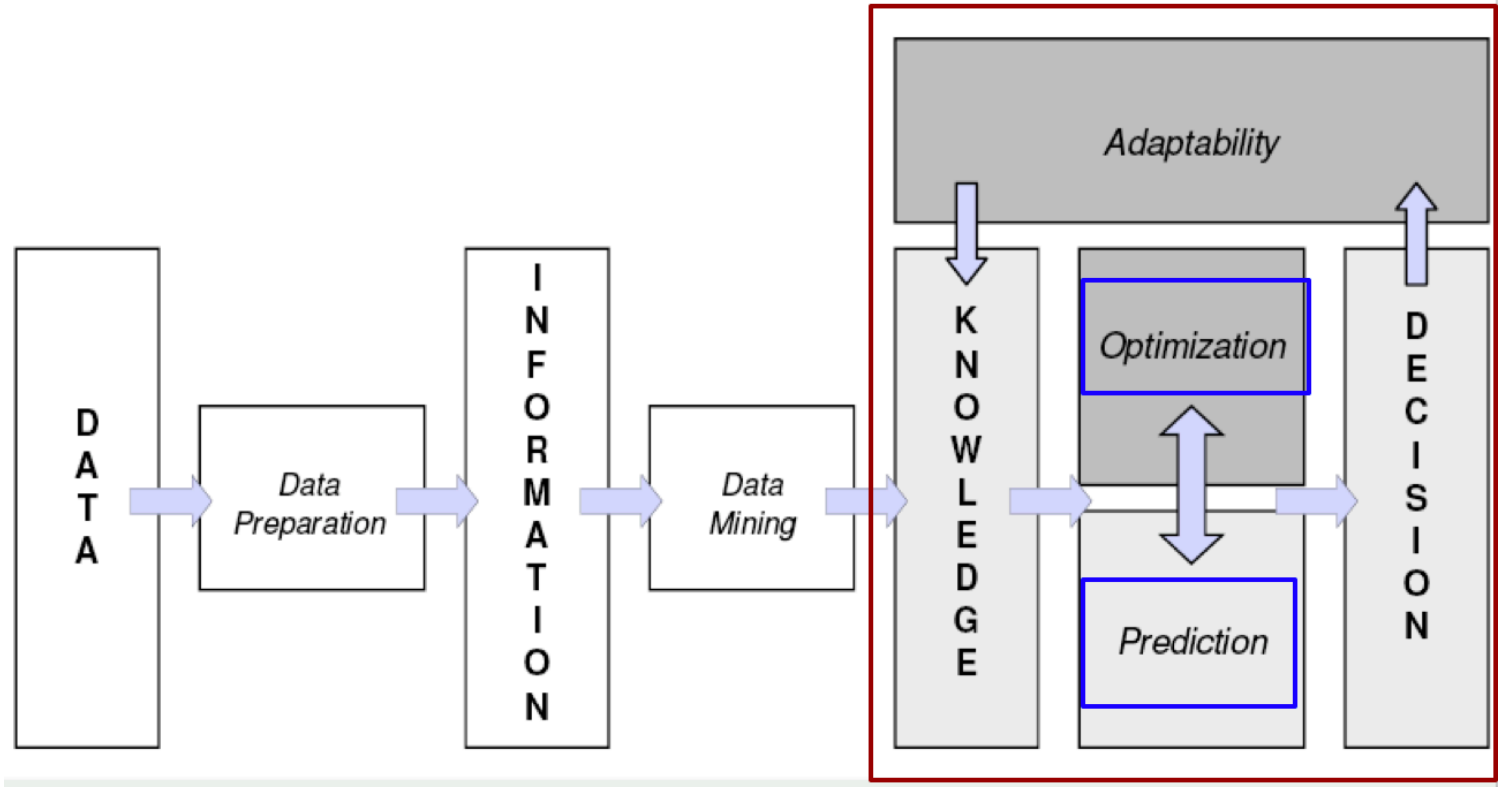


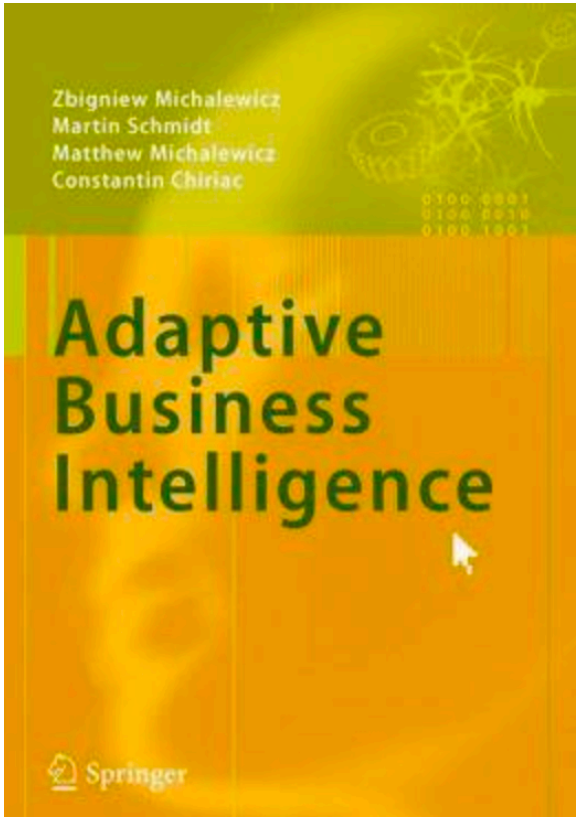




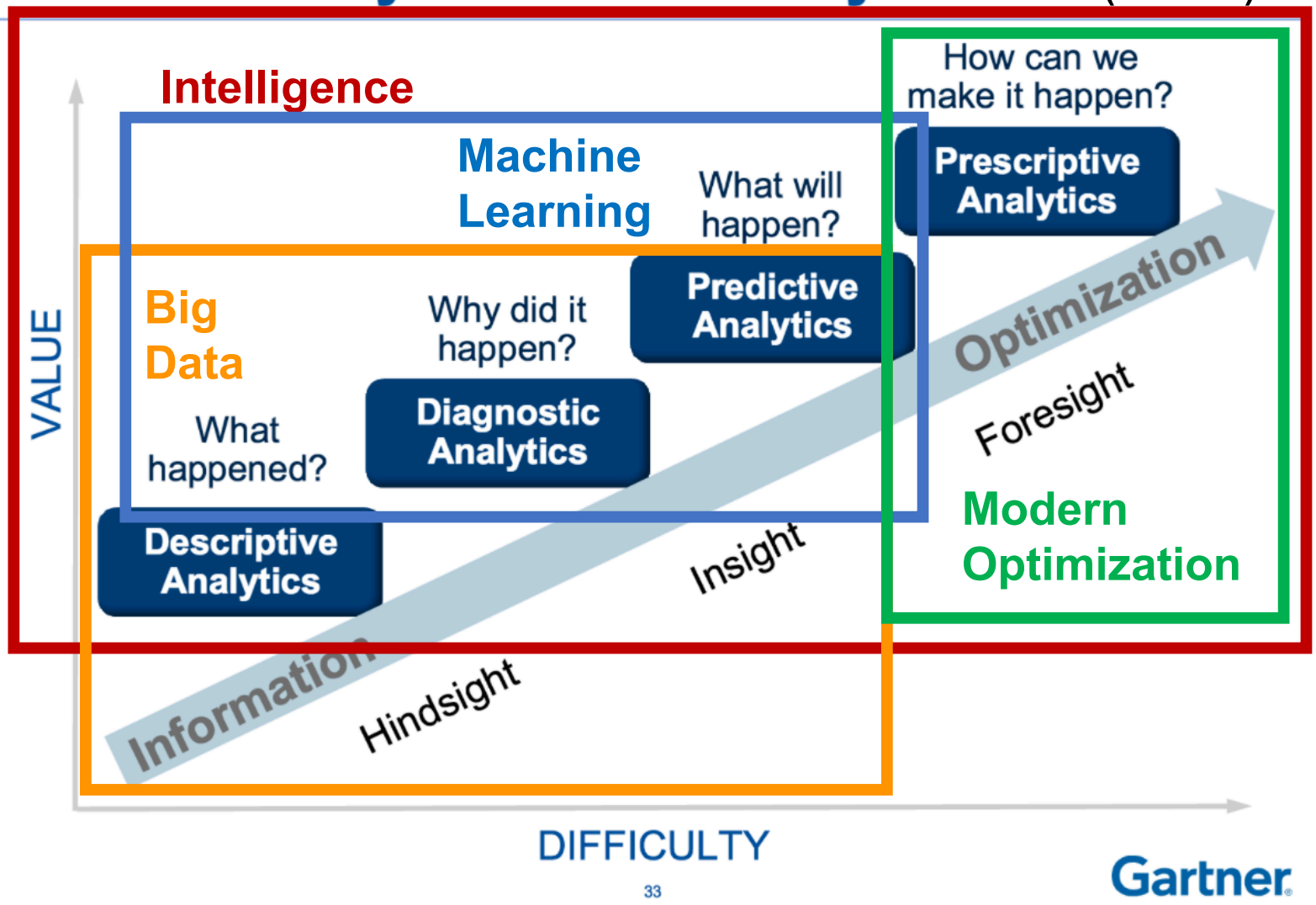
Adaptive Business Intelligence (ABI) (Michalewicz et al. **2006**)

Adds intelligent adaptive modules to standard BI systems: Data-driven **Prediction** and **Modern Optimization**.





Gartner Analytic Ascendancy Model (2013):



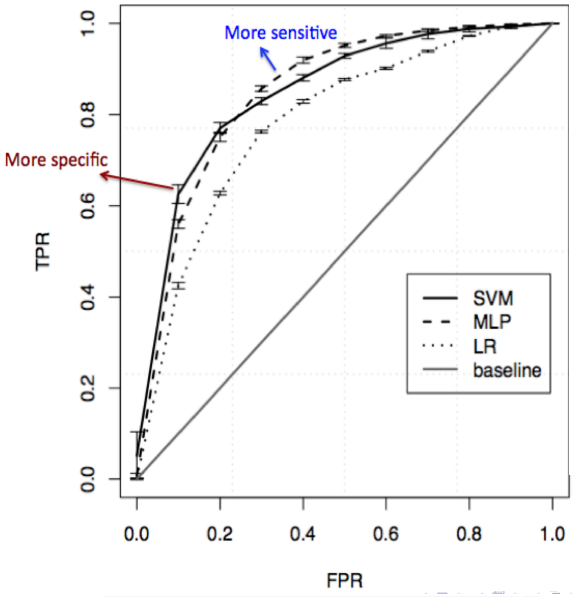
Gartner

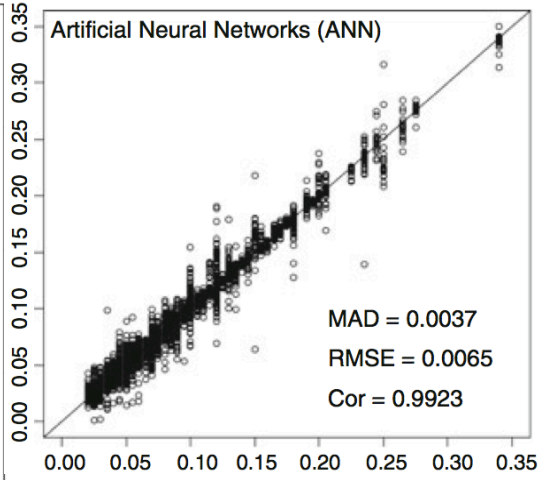
Open Source Tools: R and Python

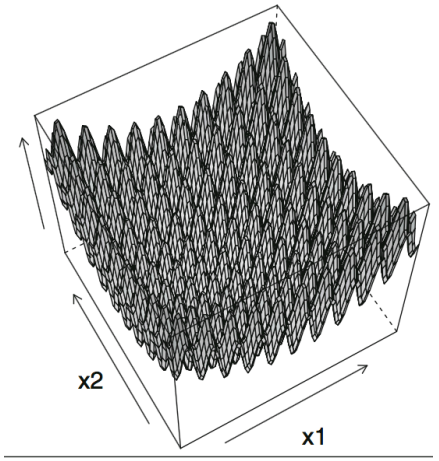
IEEE 2018 computer language ranking:

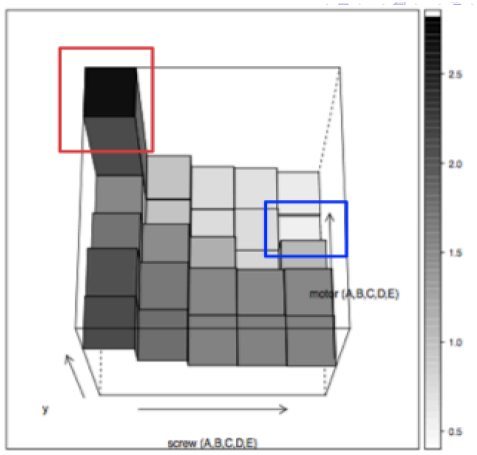
Language Rank	Types	Spectrum Ranking
1. Python		100.0
2. C++		99.7
3. Java		97.5
4. C		96.7
5. C#		89.4
6. PHP		84.9
7. R		82.9
8. JavaScript		82.6
9. Go		76.4
10. Assembly		74.1

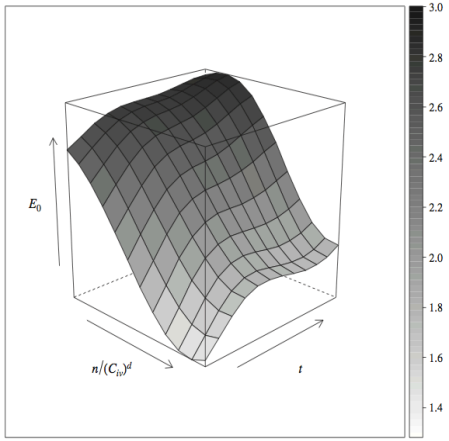
R http://www.r-project.org, **rminer** package, R Springer book

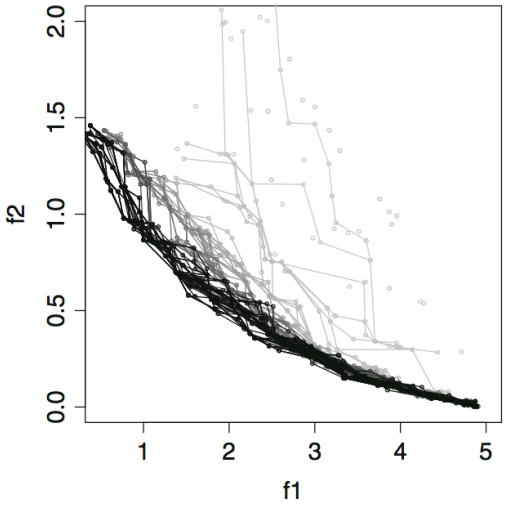


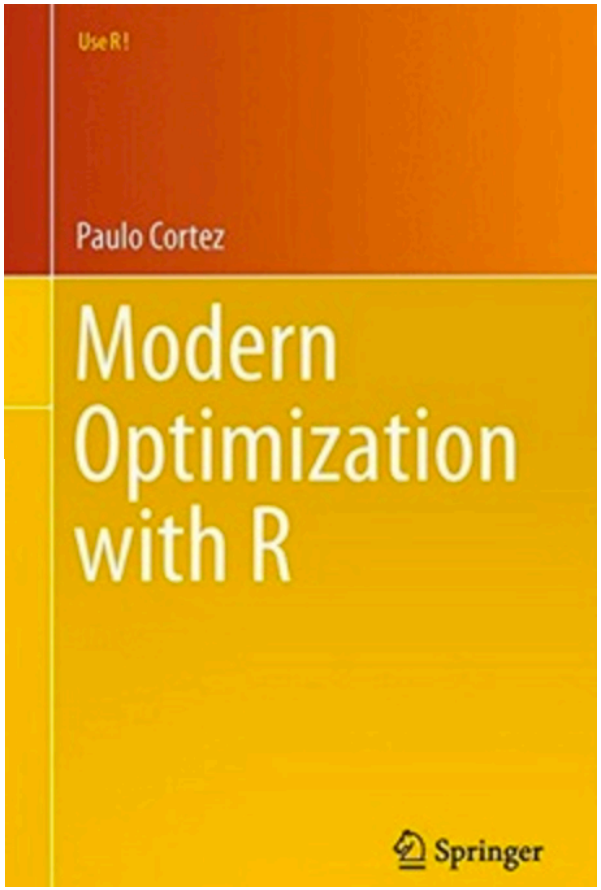












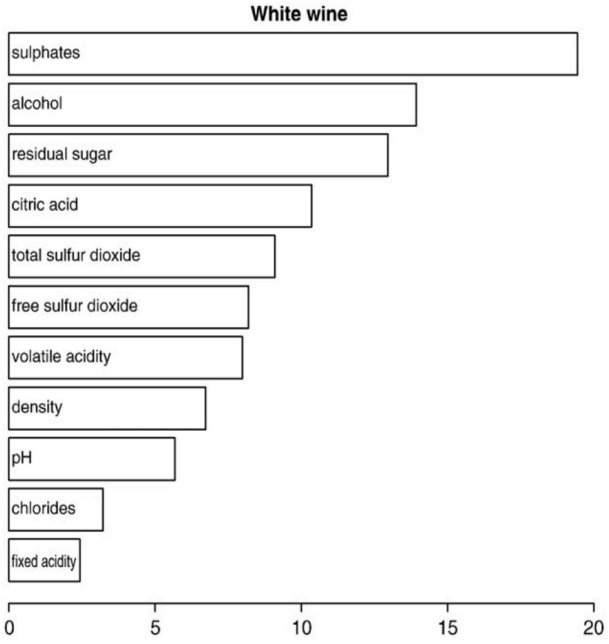
Part 2
Application Examples

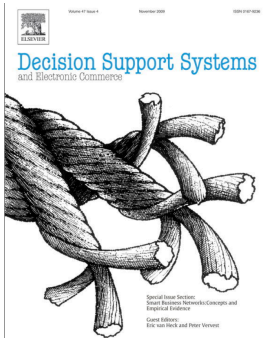
Predictive Analytics

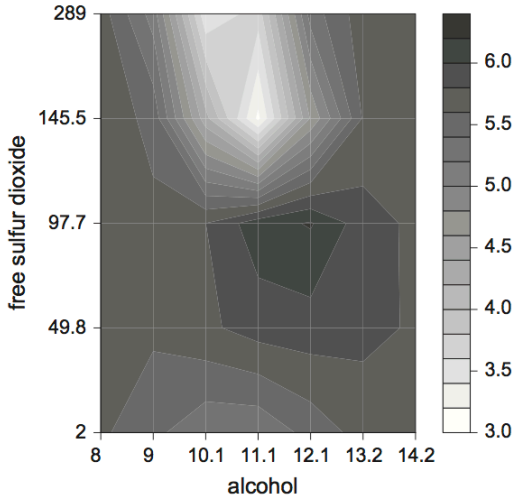
Wine Quality

Prediction: Regression, Support Vector Machines

"Modeling wine preferences by data mining from physicochemical properties", DSS 2009.







Wine Quality Data Set

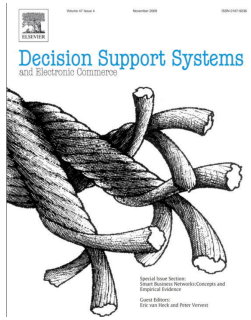
Download: Data Folder, Data Set Description

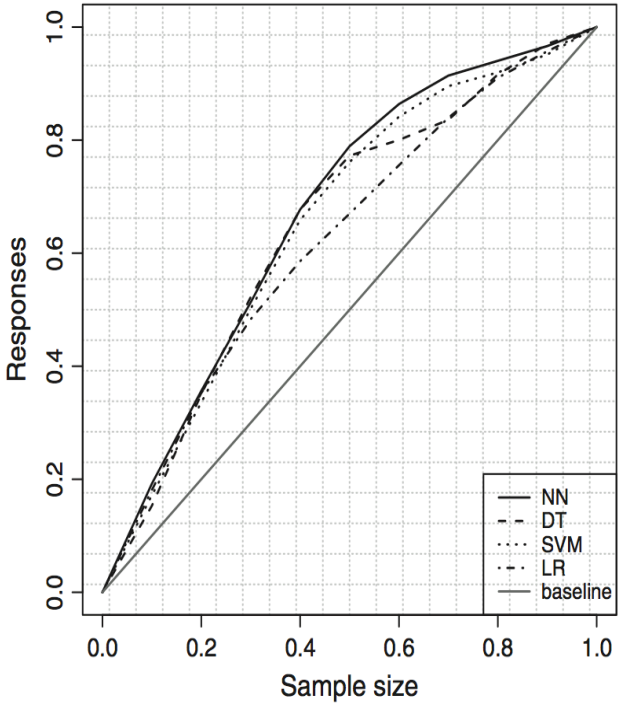
Abstract: Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], [Web Link]).

Bank Telemarketing

Prediction: Feature Engineering, Neural Networks

"A data-driven approach to predict the success of bank telemarketing", DSS 2014.





Examples of DT, SVM and NN response values:

Sample size	DT	SVM	NN
5%	8.0%	9.0%	10.4%
10%	15.3%	17.2%	19.2%
20%	35.1%	33.6%	35.6%
30%	52.2%	50.2%	51.3%
40%	67.7%	65.9%	67.7%
50%	77.2%	76.1%	78.9%
60%	80.0%	84.1%	86.4%
70%	83.6%	89.6%	91.4%

Bank Marketing Data Set

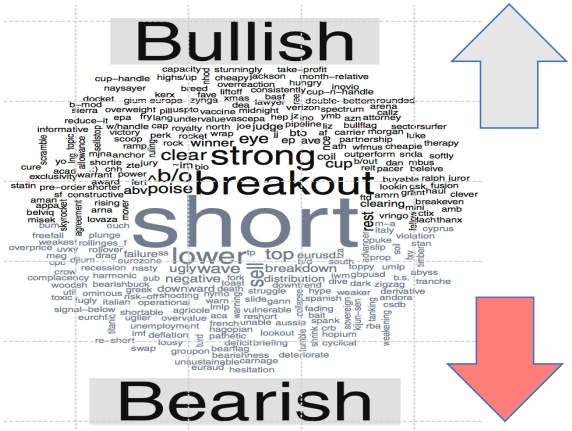
Download: Data Folder, Data Set Description

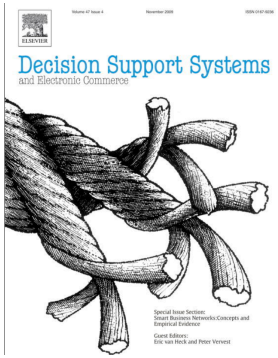
Abstract: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Stock Market

Prediction: Text mining, Sentiment Analysis

"Stock market sentiment lexicon acquisition using microblogging data and statistical measures", DSS 2016.





New microblog financial lexicon:

https://github.com/nunomroliveira/stock_market_lexicon

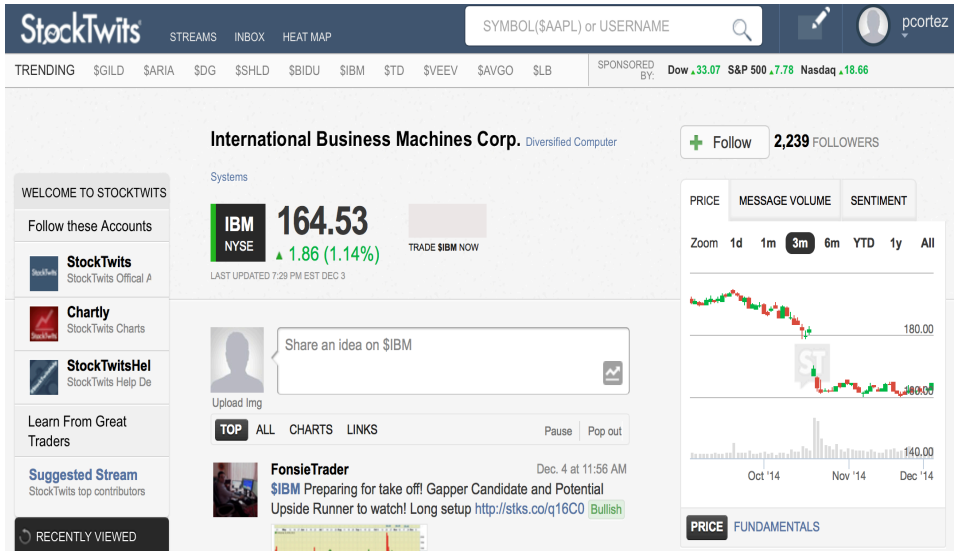


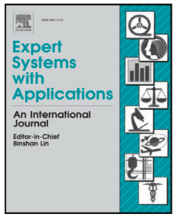
Table 26.: Short sample (10 messages) of StockTwits messages, author supplied label and SA classification

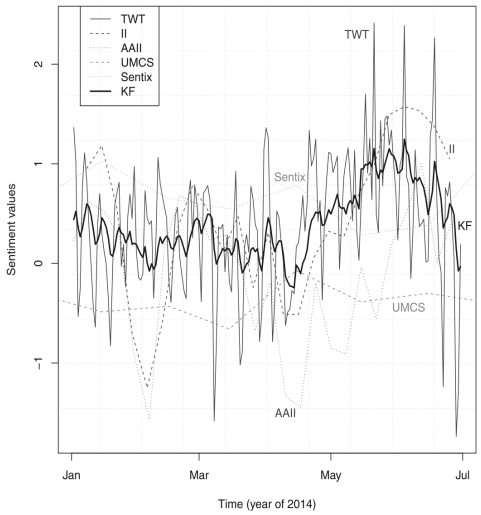
StockTwits Message	Author Label	SA Label
$FIVE settin **up well for a big breakout in the coming weeks.**	Bullish	Bullish
Thats why it rallied kinda late. You might want to exit now before it retraces 10 percent tomorrow.	Bearish	Bearish
$SHLD $LNKD $CMG calls on our **morning watchlist before the open. Buy when its cold sell while is** steaming **hot.**	Bullish	Bearish
Let the fat cats **buy the 3 dip on** ocz! **lol.**	Bullish	Bullish
$AAPL **will continue strong into close.**	Bullish	Bullish

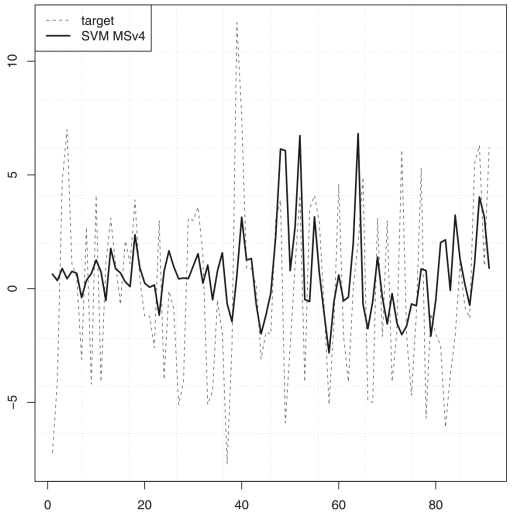
Stock Market

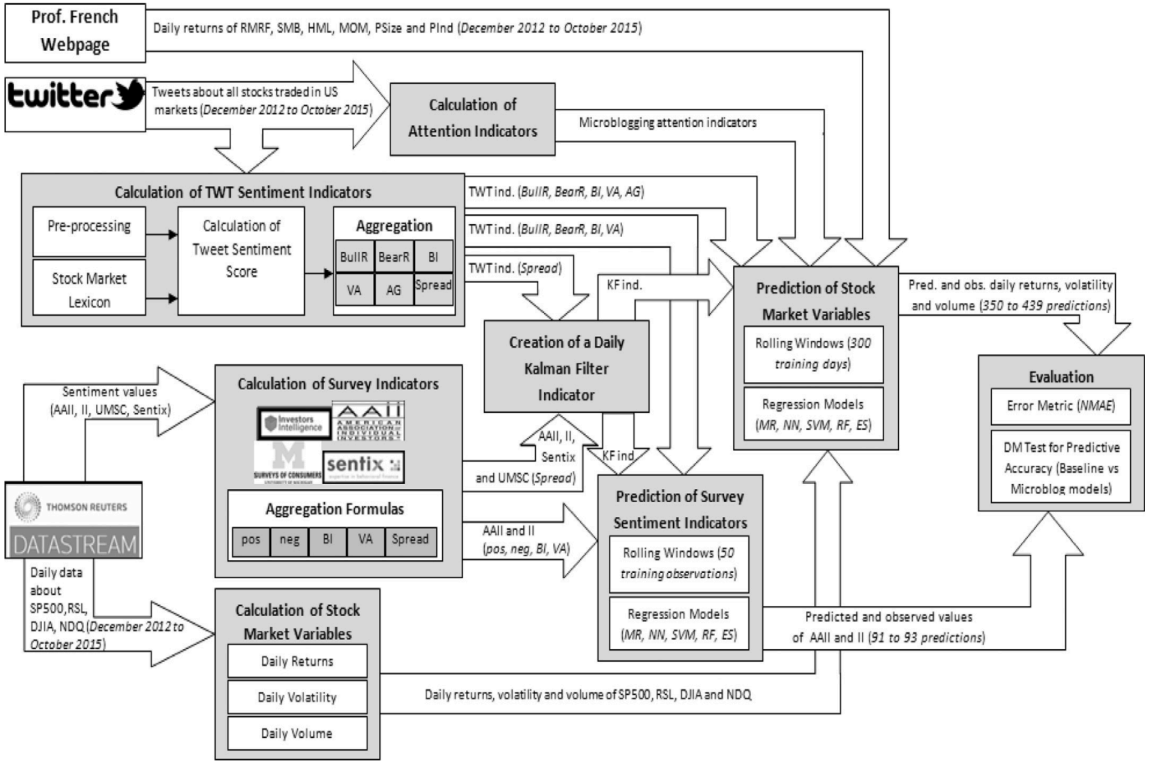
Prediction: Kalman Filter, Regression

"The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices", ESWA 2017.







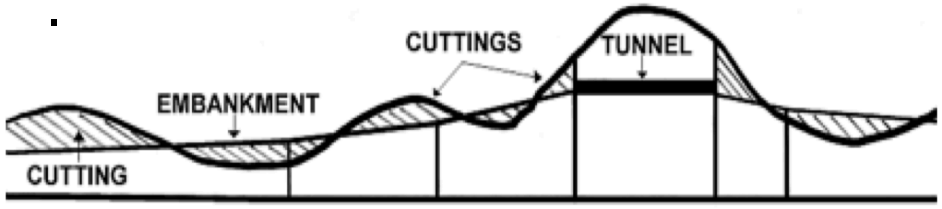


Adaptive Business Intelligence (ABI)

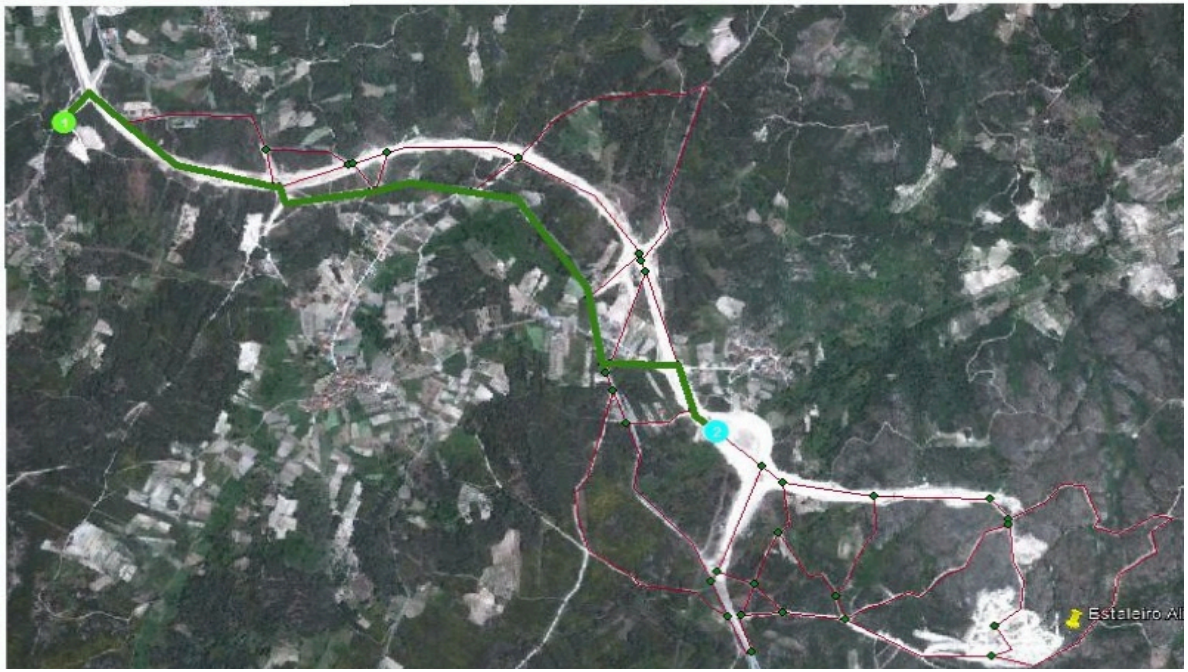
Design of Earthworks

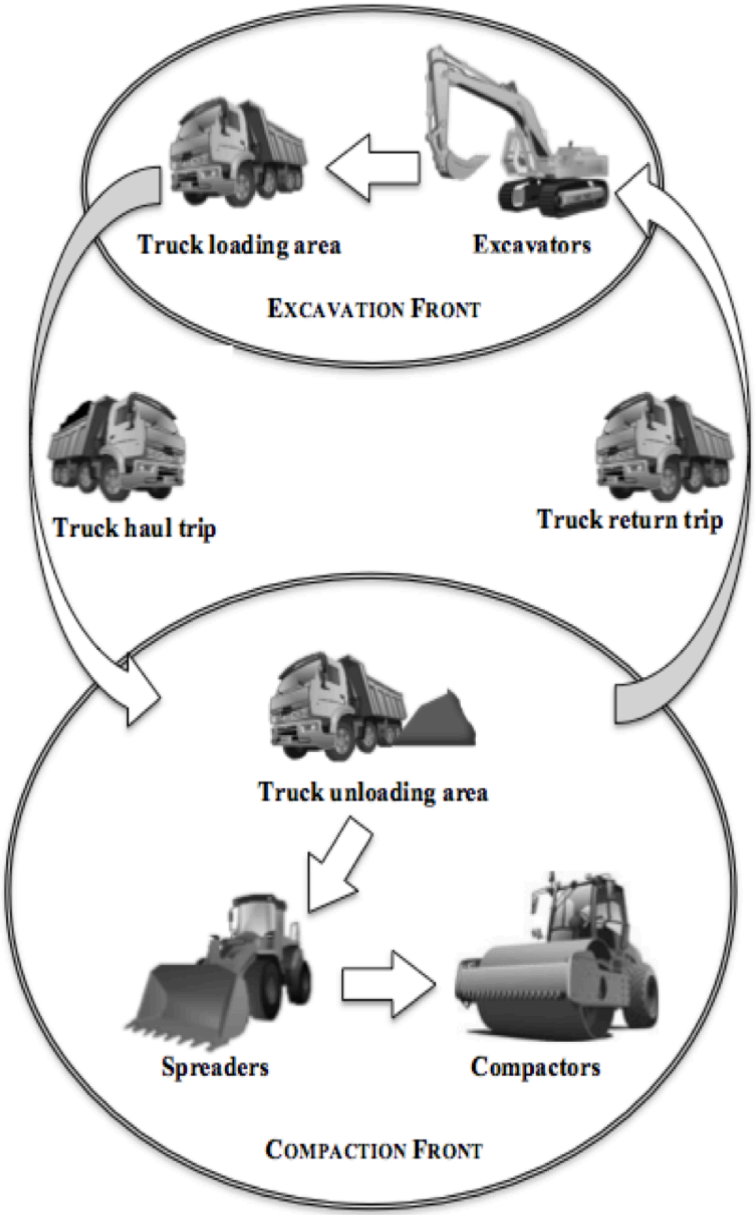
Leveling or shaping of a target area: need to manage construction equipment

.



Earthwork tasks schema

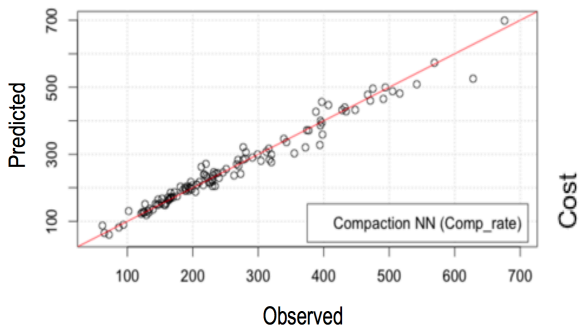




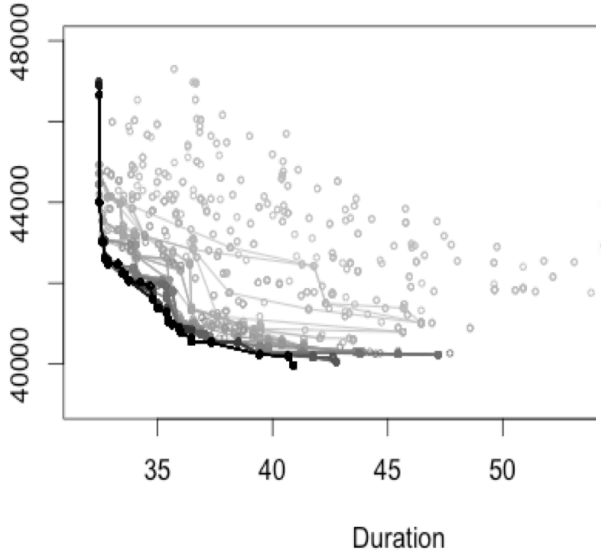
ABI approach:

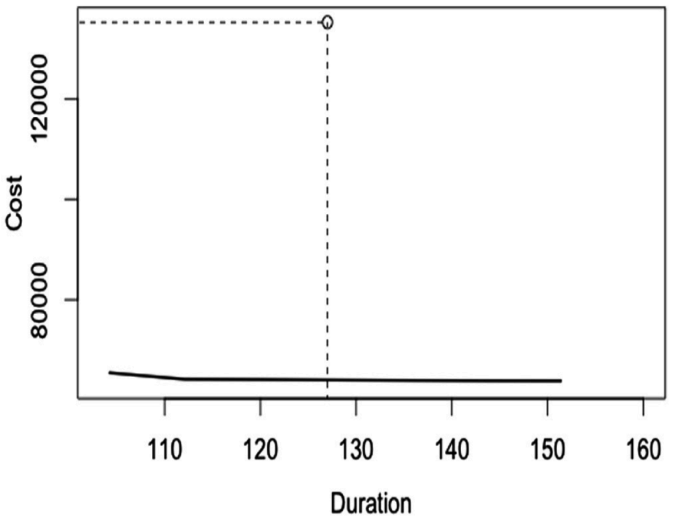
Prediction of compaction equipment conditions for a particular soil type (regression, Neural Network)

Optimization: Assigning compactors to embankment areas (multi-objective, NSGAII algorithm)



c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
2	1	2	1	1	3	3	0	3
Phase 1			Phase 2			Phase 3		





M. Parente, P. Cortez, A.G. Correia, An evolutionary multi-objective optimization system for earthworks, Expert System with Applications, 42:6674-6685 2015.

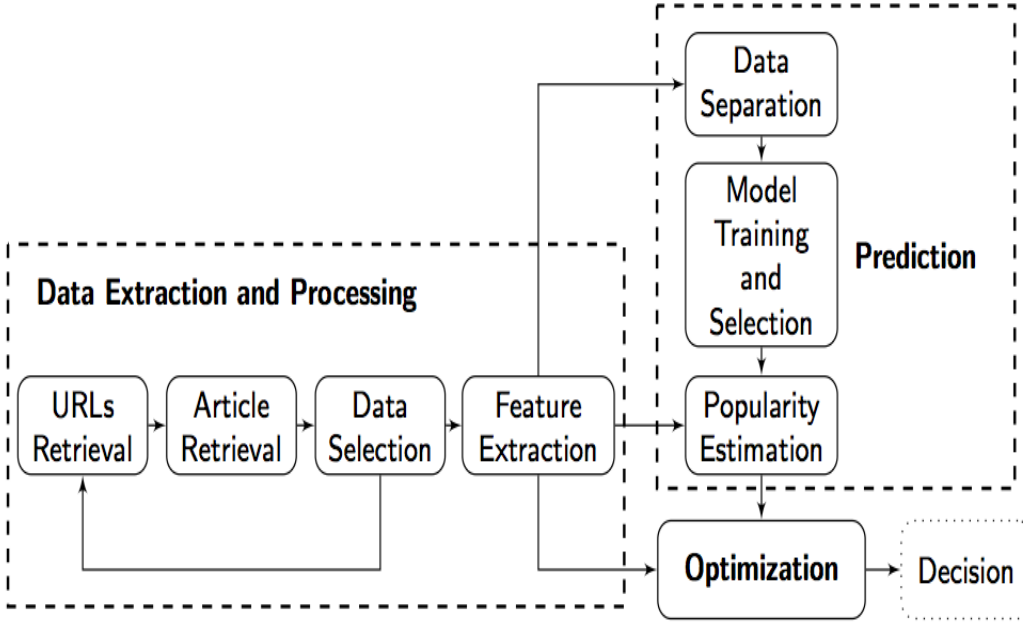
Design of Online News

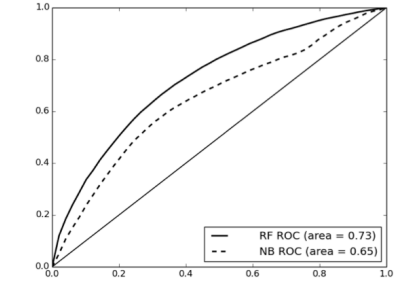
Assisting users in the writing of online news: how to improve popularity?

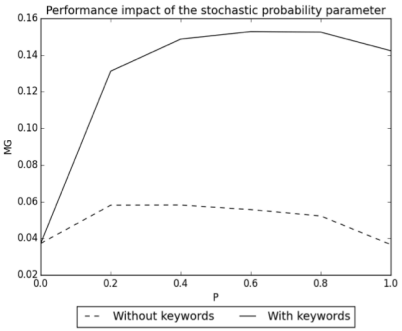
ABI approach:

Prediction of news shares based on text and multimedia features (classification, Random Forest)

Optimization: Searching for more popular news features (stochastic hill-climbing)







Online News Popularity Data Set

Download: Data Folder, Data Set Description

Abstract: This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity).

K. Fernandes, P. Vinagre and P. Cortez, A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News, EPIA 2015.

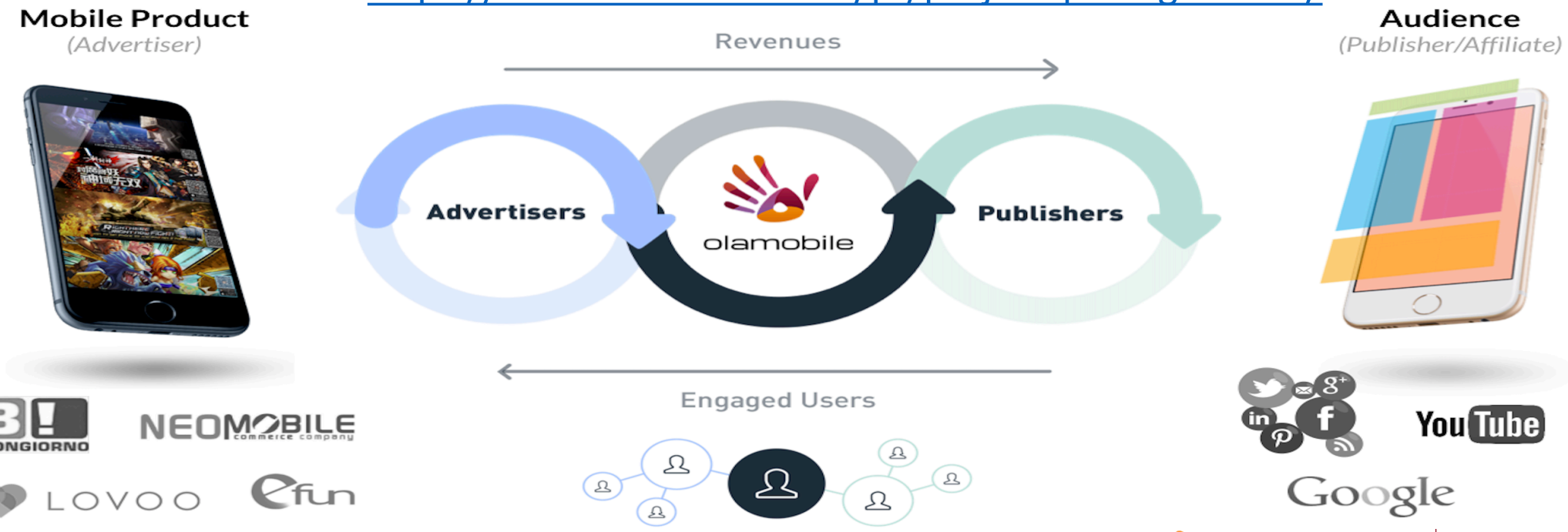
EPIA 2015 – COIMBRA 17TH PORTUGUESE CONFERENCE ON ARTIFICIAL INTELLIGENCE

Mobile Marketing: PROMOS – Prediction and optimization of advertising campaigns for mobile devices

Prediction and Optimization: Big Data, Classification, Modern Optimization

http://promos.dsi.uminho.pt/

https://www.olamobile.com/pt/projeto-portugal-2020/



PORTUGAL 2020
NORTE2020
NORTH-01-0247-FEDER-017497
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional
olamobile
University of Minho

Textile Industry:

http://www.ccg.pt/my-product/texboost/

Centro de Computação Gráfica
Investigação e Desenvolvimento Tecnológico
Research and Technological Development

TexBoost
less Commodities
more Specialities

(3-year project, funded by Portugal 2020/Adi, total of 9.2 MEUR)

ABI approach (for PPS1 textile digitalization and dematerialization):

Prediction of textile properties based on design features

Optimization: searching for best textile design features

DTx Digital Transformation Colab

"aims to contribute to make Portugal a reference in the exploitation of digital transformation"
Cyber-Physical Systems (CPSs) and ABI
http://www.dtx-colab.pt/

University of Minho

CENTROALGORITMI

Centro de Computação Gráfica
Investigação e Desenvolvimento Tecnológico
Research and Technological Development

Associate Professor
(Habilitation, PhD)

My contacts:
email: pcortez@dsi.uminho.pt
URL: http://www3.dsi.uminho.pt/pcortez
Linkedin: https://pt.linkedin.com/in/paulocortez
twitter: @PauloCortez4

Thank you!