



Inspire...Educate...Transform.

PGP in Big Data Analytics and Optimization

Orientation Session

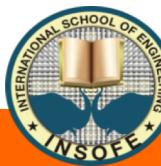
Kalpa K S
Data Scientist, INSOFE

Class Structure

- 4 hours lecture...focus on explaining concepts.
- 4 hours lab...focus on hands-on.
- Hands-on activities using R & Python. Do not underestimate its importance.

Will have focus in exams.

- Welcome to come to office during weekdays and clarify questions. But please make an appointment as Data Scientists are also involved in CPR activities.



Course Material and Discussion - Piazza



Screenshot of the Piazza platform interface for CSE 7202c Batch 10. The top navigation bar includes Q & A, Resources, Statistics, Manage Class, and a user profile for Sridhar Pappu. The main content area shows a question titled "Few basic questions...." posted by Sridhar Pappu on May 5, 2015. The question asks about multiple predictors, outliers, and various regression metrics. Below the question is a response from the instructor, Sridhar Pappu, dated April 29, 2015, discussing the importance of understanding assumptions and making decisions based on multiple metrics. The interface also shows other posts and discussions related to linear and logistic regression, step-wise regression, and data splitting.

- Ask a lot of questions
- Post questions, assignments, etc. in the correct module; else, they will be lost.



Video Recordings - Impartus

- Do not miss classes.

Video ≠ Classroom

- Will be available for 3 months after the program.

- Contact if you need extension.

The screenshot shows the Impartus Lecture Capture interface. On the left, there's a sidebar with icons for home, user profile, search, and help. Below it, a list of classroom recordings is displayed:

- Class Room 1, Batch 17, Session 2016-17, 139 Lectures. One recording is shown: "Lecture 51: Decision Trees" (20th Jul 2016, 11:00 am).
- Class Room 2, Batch 18, Session 2016-17, 174 Lectures. One recording is shown: "Lecture 174: No Topic Entered" (20th Nov 2016, 11:00 am).
- Class Room 1, Batch 23, Session 2016-17, 96 Lectures. One recording is shown: "Lecture 78: No Topic Entered" (26th Mar 2017, 12:00 pm).
- Class Room 2, Batch 24, Session 2016-17, 128 Lectures. One recording is shown: "Lecture 98: The Art And Science..." (15th Apr 2017, 12:00 pm).
- Classroom 3, Batch 26, Session 2016-17, 96 Lectures. One recording is shown: "Lecture 38: Rules And Rule Ind..." (19th Mar 2017, 12:00 pm).

In the center, three video thumbnails are shown:

- Lecture 1 - Fundamenta... (237 Views, 4 stars)
- Lecture 11 - Fundamenta... (184 Views, 5 stars)
- Lecture 34 - Statistics an.. (182 Views, 4 stars)

Below these, a message says "and 4 Others also watched this lecture" followed by a user profile for Samrat Putcha. A large video player at the bottom shows a live lecture with a male professor standing in front of a whiteboard. The video player has a play button and a circular progress bar.

On the right side, there are several panels:

- A news feed with articles like "Why Tata Nano Did Not Take Off Despite Being The Cheapest Car? Detailed Analysis Would Rea..." and "A Fire Fighter Has To Get To A Burning Building As Quickly As He Can. There Are Three Path...".
- A "My Topics" section with a search bar and a list of topics.
- A "Help" section showing 50413 followers.
- A "Trending Topics" section with "Business", "Management", and "Financial" categories.
- A "Recently Added Lectures" section with two entries: "No Topic Entered test" (17th May 2017, 02:50 pm) and "No Topic Entered BIZ Calls" (16th May 2017, 01:30 pm).

Assessments

- 25+ ROTes, 5 CUTes, Hackathons
- Weightage
 - Recall Output Test (**ROTe**): 15%
 - Conceptual Understanding Test (**CUTE**): 25%
 - Mid-term Hackathon (**MiTH**): 15%
 - Project Hackathon and Defense (**PHD**): 35%
 - Feedback: 5%
 - Attendance (85%+): 5%
- Non-zero cumulative score in each a must

Report Date: February 11, 2015

The below report summarizes your performance in the program of your study. Absence of a bar in the chart(s) indicates that you did not participate in the corresponding activity. If you find any discrepancy in the report, please contact the academic team at INSOFE.

Cumulative Scores* Required for:

Certificate of Engineering Excellence – 50% with a non-zero score in each assessment activity
INSOFE Career Services – 60%



* Weightage of each assessment activity in the Cumulative Score is:

Feedback - Zoho

- 5% grade as incentive.
- One per module.
- Provide as soon as you receive the link.

Instructors: Dr. Sridhar Pappu

PLEASE LIMIT YOUR FEEDBACK TO THIS MODULE, ITS FACULTY AND DATA SCIENTISTS WHO HANDLED LABS FOR THIS MODULE.

1. Please tell us about yourself. We will ask your permission in the last question on using this information in our collateral. We will NOT use this information without your explicit permission.

Your Name:*

Your Designation:*

Your Company:*

2. Content*

	Poor	Average	Good	Excellent	Not Applicable
How smooth was the flow of the course material?	<input type="radio"/>				
Were the topics covered in sufficient detail?	<input type="radio"/>				
Did it provide real world experience?	<input type="radio"/>				
How useful is the instructional and reference material provided?	<input type="radio"/>				



Certification (or not)

- No certificate if attendance is below 70% and cumulative grade < 50%.
- PGP certificate if cumulative grade > 50%
- **Certificate of Participation**, if attendance is above 70% and cumulative grade < 50%.
- You have to work really hard to be (in)eligible for this.

The board of
International School of Engineering

on the recommendation of its faculty,

hereby confer upon

the award of

**Postgraduate Certificate in
Big Data Analytics and Optimization**

on successful completion of all the requirements of the 336-hour program conducted between _____ and _____ followed by a project hackathon and defense.

This program is certified for quality of content, assessment and pedagogy by the Language Technologies Institute (LTI) of Carnegie Mellon University (CMU).

The program curriculum has been developed in collaboration with LTI.



Dated this _____ day of _____, _____

Dr. Dakshinamurthy V Kolluru
President

Dr. Sridhar Pappu
Executive VP-Academics



ROTe, CUTe, MiTH and PHD Postponement Policy



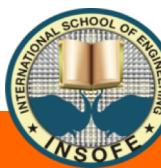
Recall Output Test (ROTe)

- Only 2 ROTe postponements will be allowed throughout the program. Students must inform in advance.
- If someone takes a ROTe from outside the classroom, they will be marked 0 for that ROTe.
- While 2 postponement policy stays, if there is an extended business trip, On sharing the details, INSOFE can make exceptions. INSOFE may request candidates for proofs like tickets, etc. as well.



Conceptual Understanding Test (CUTe)

- No postponement of CUTe allowed during the program.
- If any student misses a CUTe and INSOFE determine at the end of the program that student met the performance criteria only for a Certificate of Participation and not the PGP Certificate, INSOFE will check if scoring up to 80% in the missed CUTe would have enabled student to get the PGP Certificate. If so, INSOFE will conduct a separate CUTe for student at that time. If scoring even 80% would not impact Student's certification in that case INSOFE will not conduct CUTe.



Conceptual Understanding Test (CUTe)

- In case student miss more than one CUTe. INSOFE will determine which module will enable him / her to cross the line for PGP Certificate among the missed modules. Please note only ONE module can be selected.
- Student must submit **valid documented proofs** of why (s)he missed the CUTe in the first place. Without such a proof, the CUTe will not be conducted even if it allows student to cross the finish line.



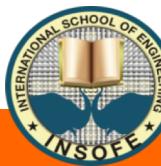
Mid-term Hackathon (MiTH)

- No postponement of MiTH allowed during the program.
- MiTH will be treated as CUTe for postponement purposes.
- MiTH can be conducted again only along with a future batch as it is a competition. It cannot be taken at your convenience.
- If you miss the immediate next MiTH opportunity as well, you will not be provided further option of taking it again.
- Student will also not be eligible for the scholarship for Phase 2 that is based on the MiTH.



Project Hackathon and Defense (PHD):

- No postponement of PHD allowed during the program.
- PHD can be conducted again only along with a future batch as it is a competition. It cannot be taken at your convenience.
- If you miss the immediate next PHD opportunity as well, you will not be provided further option of taking it again.
- You will also not be eligible for the scholarship for Phase 3 that is based on the PHD



Get a RAISE – Research Assistantship, Internship and Scholarship Events

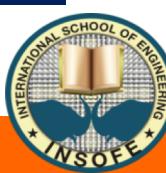


Up to 10% of your batch revenue is allocated as the total scholarship amount.

Number of scholarships and the amount disbursed will depend on the performance at each phase. The academic team will conduct the evaluations and it is up to their discretion to decide the amount and number of scholarships.

A student can win back part of or even full fee amount in this process.

Internship and Research Assistantship announcements as required by INSOFE.



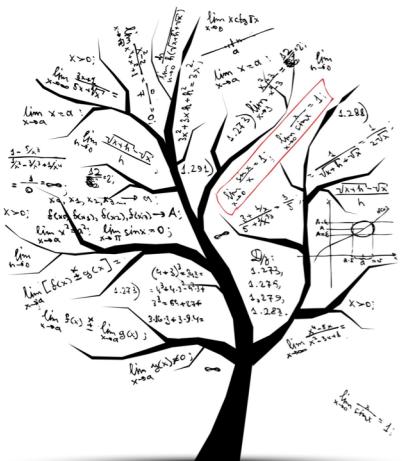
Internship

- Guaranteed Research Internship
 - for non-working/working professionals
 - 6 months duration
 - minimum of ₹ 5000 for all good internship projects.
 - Top 3 internships will be awarded with a prize money every quarter.
- Internship – consulting
- Research assistantship



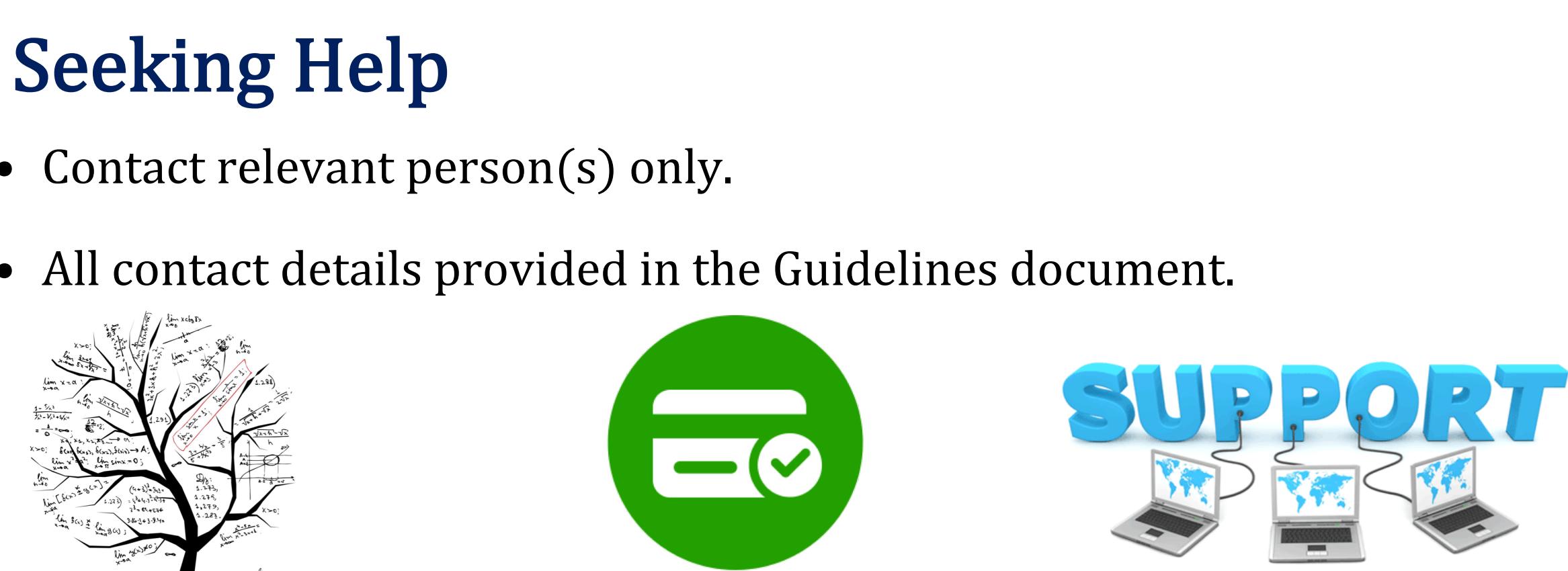
Seeking Help

- Contact relevant person(s) only.
- All contact details provided in the Guidelines document.



For any subject-matter help

Seema | Chaitanya| Mahidhar |
Tarun



All payments and loan related

Ravindra Karanam



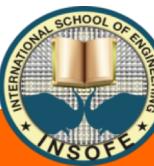
For any in-class issues

Karthik Katamreddy
Reddaiah Kavuturu



Some Good Practices

- Classes start at 9 AM. Please respect others' time. ROTEs conducted at 9 AM.
- Please switch off phones or keep them in silent mode. Please do not distract others.
- Please do not browse internet unless instructed to do so. We want you to learn.
- We take attendance in the last hour of the day. We do want you to learn.
- Please do not ask questions on topics not yet covered. No ransom will be paid for hijacking class discussions.
- Refer to the Guidelines document in your welcome kit.
- Write all your exams.

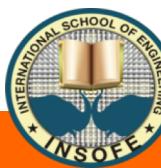




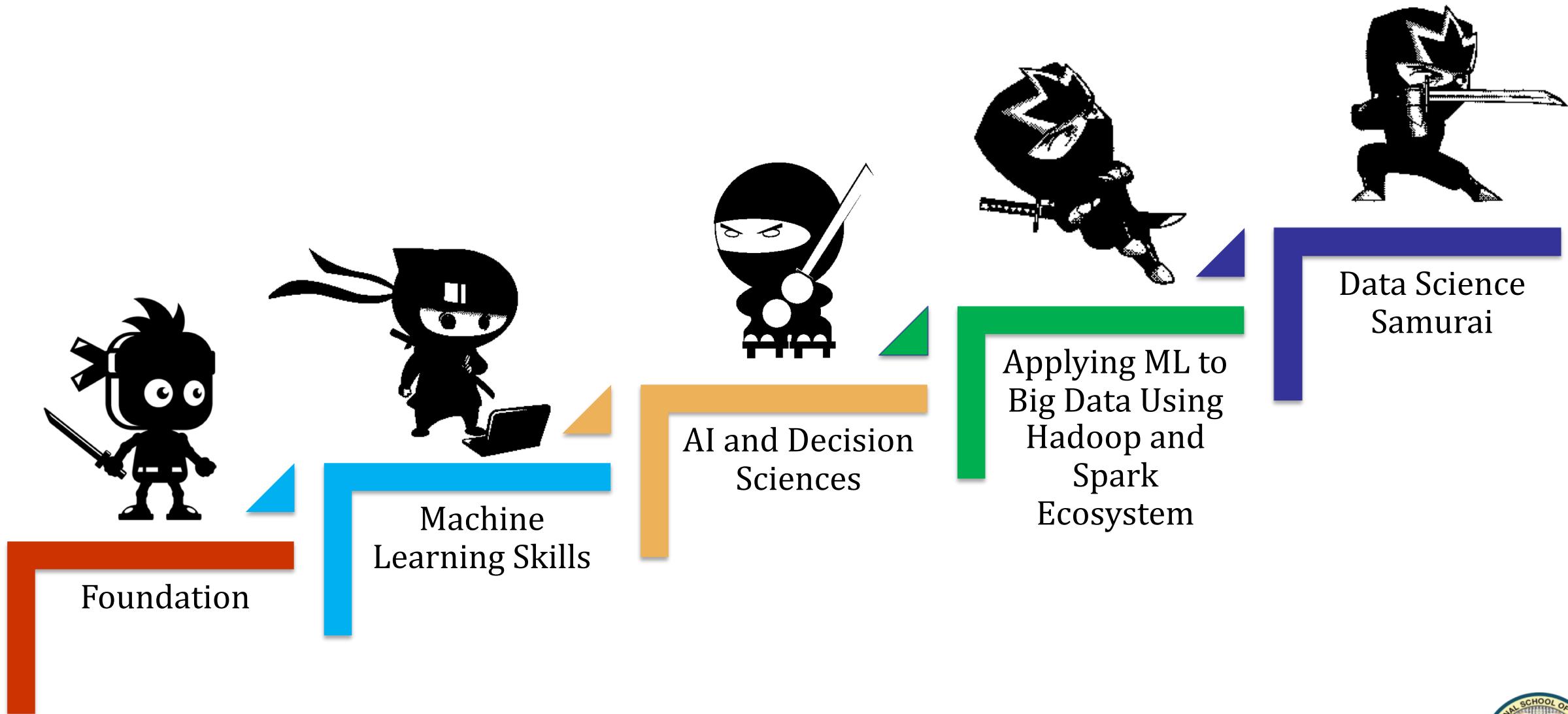




“Data is the sword of the 21st century. Those who wield it well, the Samurai.” – Jonathan Rosenberg, Advisor to Google Alphabet CEO, Larry Page

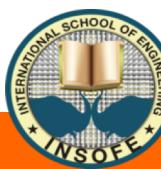


Curriculum structure – how are we heading?

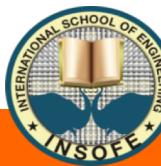


Module Names

- Essential Engineering Skills in Big Data Analytics Using R and Python
- Foundations of Probability and Statistics for Data Science
- Statistics and Probability in Decision Modeling
- Methods and Algorithms in Machine Learning
- The Art and Science of Storytelling with Data Visualizations
- Foundations of Text Mining and Search
- AI and Decision Sciences
- Applying ML to Big Data Using Hadoop and Spark Ecosystem



BREAK



20,000-FOOT VIEW OF DATA SCIENCE



Why Data Science?



STUDENTS CAN OPT FOR NEW TECH: EXPERTS

DC CORRESPONDENT
VISAKHAPATNAM, MAY 17

Data analytics, cyber security, cloud computing, virtual reality, artificial intelligence, digital marketing could be emerging sectors in the IT sector, say experts.

"Students equipped with the right skills will have plenty of job opportunities in the future. However, virtual and augmented reality, data analytics, robotics, computer numerical control laboratory, etc could play a key role in the future job market," said G.S.N. Raju, vice-chancellor of Centurion University.

He predicted an impending shift in the job market from traditional job roles to innovative roles.

According to experts, the WannaCry ransomware attack is an example of how the malware threats can create lakhs of cyber security jobs in the future.

They suggested engineering students to learn new competencies and skills to capture future job market.

According to reports, India will have a shortage of one million cyber security experts by 2020 and about 2 lakh data analysts by 2018.

Beyond Hype and Hoopla

Data analytics, cyber security, cloud computing, virtual reality, artificial intelligence, digital marketing sectors could be emerging sectors in the IT, says experts.

Students equipped with right skills will have plenty of job opportunities in the future...

"There's no denying that 'data scientist' is a hot job title to have right now, and for good reason. It's a tremendously fun and challenging field to be in, and despite all of the often undeserved hoopla that surrounds it, data scientists are doing some pretty amazing things. So it's no surprise that many people are clamoring to find out how to become data scientists." http://treycausey.com/getting_started.html

Read more: <https://www.kdnuggets.com/2018/02/2018-predictions-analytics-data-science-hiring-market.html>





Colombia will win the 1994 World Cup in the USA.

Spain are favourites for the 1998 World Cup in France.

China will qualify for the next round from their 2002 World Cup group.

Argentina & France will both reach the final in WC 2002.

Brazil won't even get past the group stages in WC 2002.

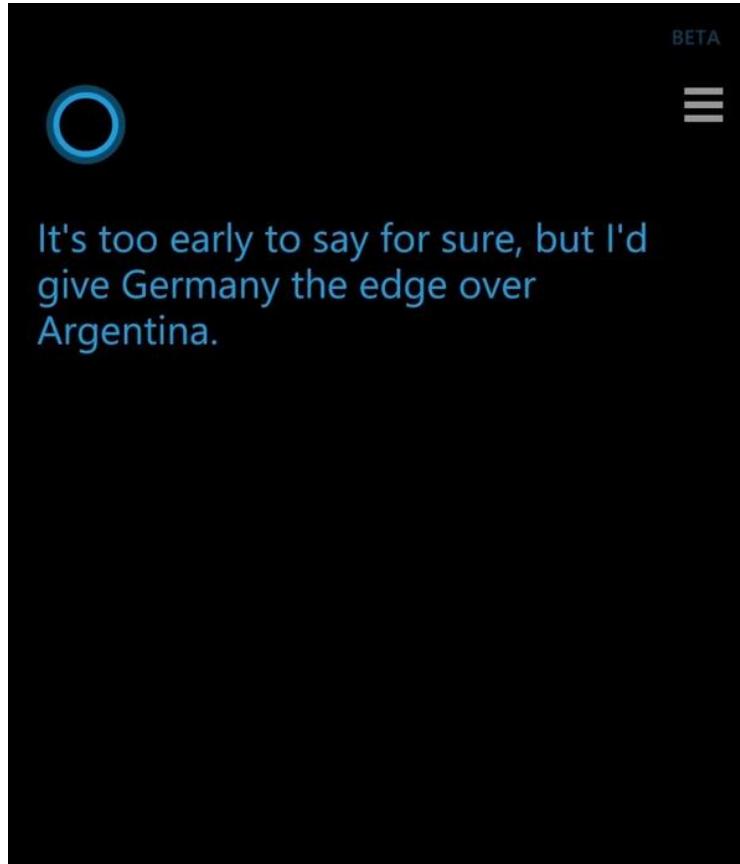
Pele, about to predict a Tom Cleverley hat-trick will hand England World Cup glory

(©GettyImages)

Source: <http://www.givemesport.com/474982-world-cup-five-worst-predictions-by-pele>

Last accessed: July 8, 2014





Google's Cloud Platform: 14 predictions correct out of 16
Baidu Trends: 15 predictions correct out of 16
Microsoft Cortana: 15 predictions correct out of 16

All 3 got Brazil-vs-Netherlands third-place result wrong.

Image Source: <https://twitter.com/EdemXII/status/487211669479641088/photo/1>
Last accessed: July 22, 2014



Power of models

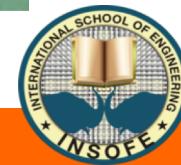


Robyn M. Dawes, The robust beauty of improper linear models in decision making, *American Psychologist*, Vol 34(7), 1979, 571-582.

Simple models do better than experts

Image Source: <http://www.psychologicalscience.org/index.php/publications/observer/2005/august-05/convention-snapshots.html>
Last accessed: July 22, 2014





rtel



DELHI
DAREDEVILS



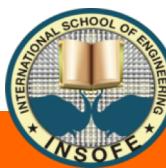


Clarifying the Jargon



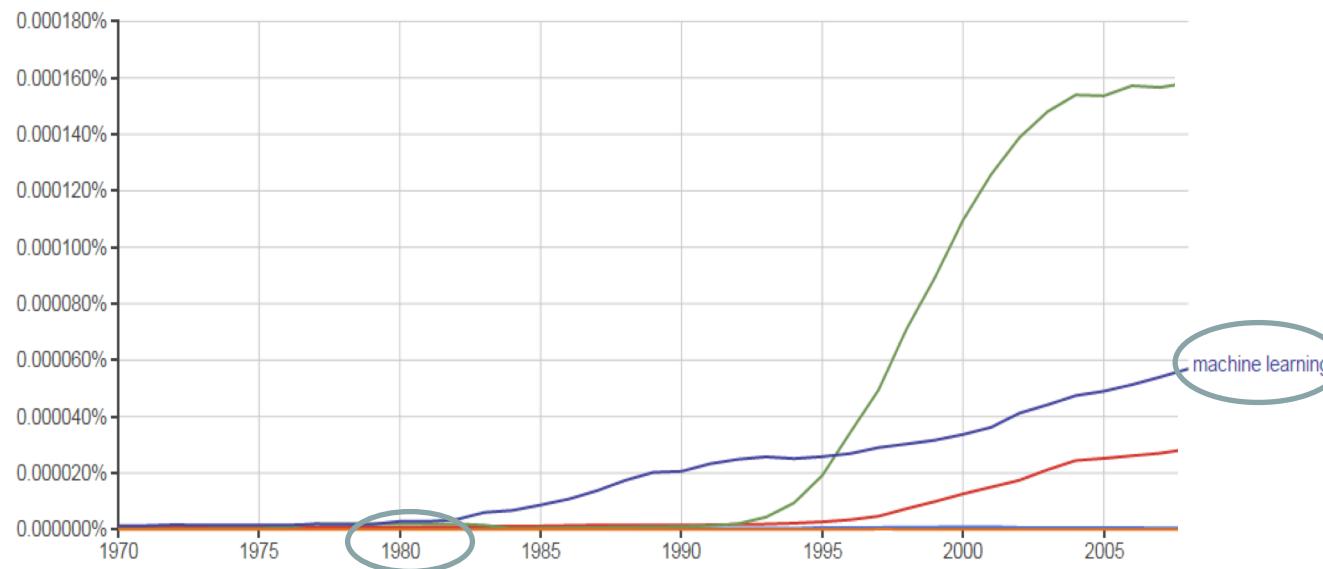
- AI, Machine Learning
- Predictive Analytics/Data Mining
- Big Data Analytics
- Data Science

ALL THESE TERMS ARE PART OF THE SAME FIELD USED AT DIFFERENT TIMES BY
PEOPLE WITH DIFFERENT BACKGROUNDS



Clarifying jargon: Chronology

- Machine learning – 1980s
 - Computing Departments named it
 - Focus was on algorithm and the amount of data was limited

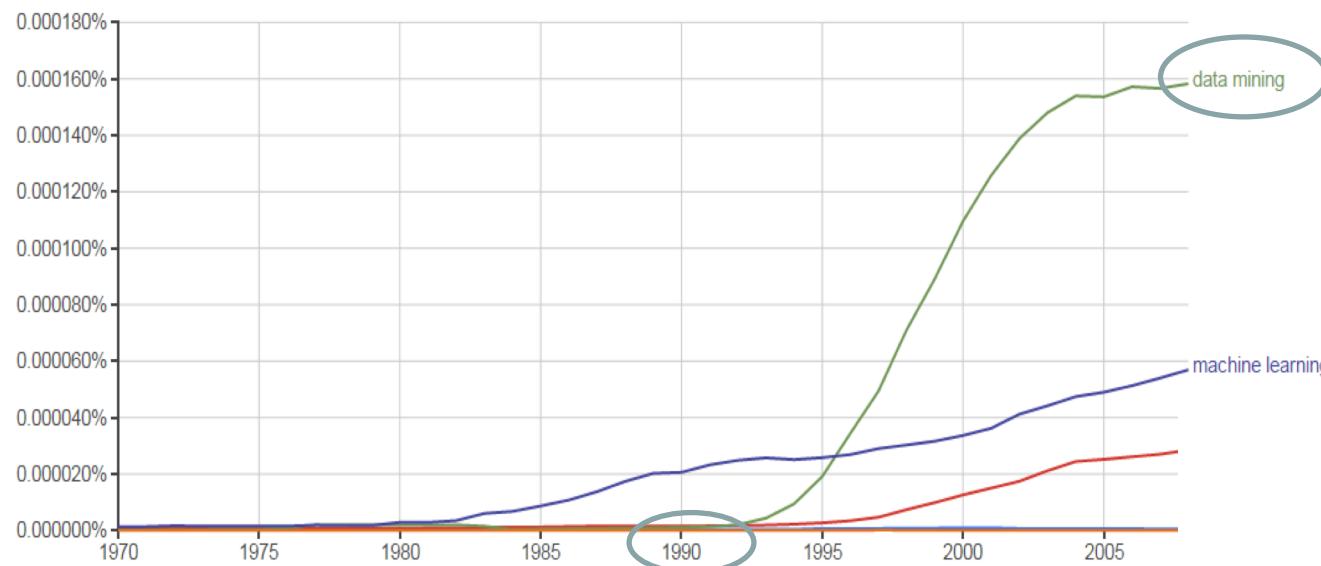


Source: Google Books chart comparing the frequency of occurrence of “big data”, “business intelligence”, “data mining”, “data science” y “machine learning” in the historical records of this service.

www.mikelnino.com

Clarifying jargon: Chronology

- Predictive analytics | Data Mining – 1990s
 - Business world started adopting Data Analytics
 - Used algorithms that are developed and applied on large amount of data



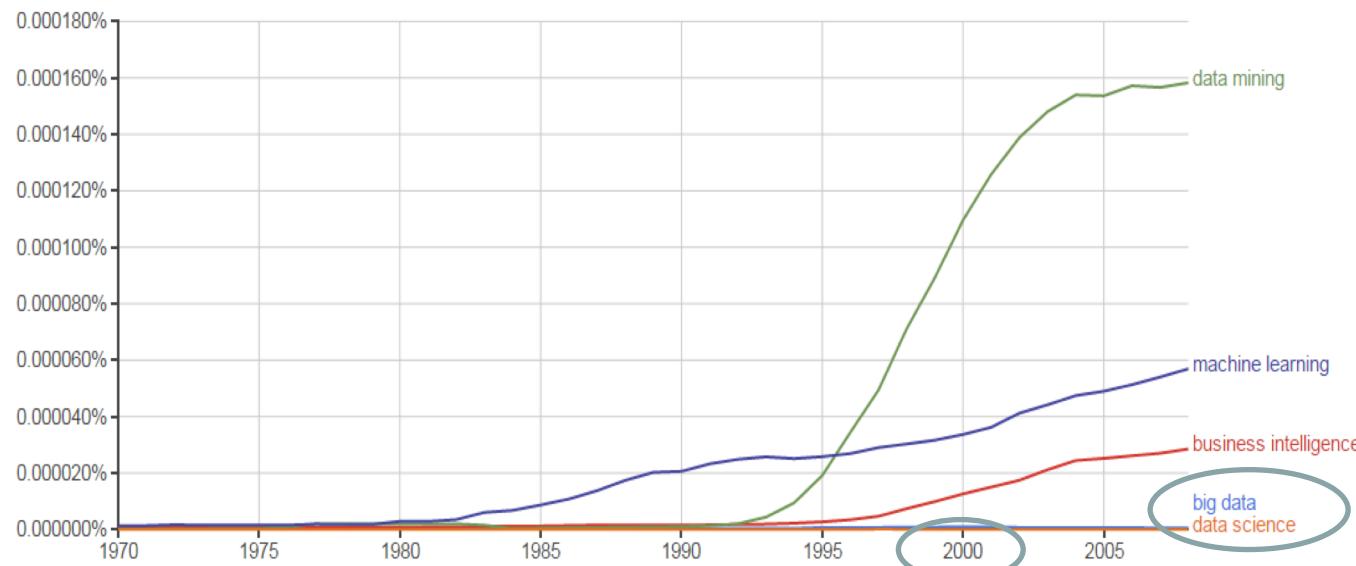
Source: Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.

www.mikelnino.com



Clarifying jargon: Chronology

- Big Data Analytics – 2000s
 - Focus was on computing on big volume of data
 - Google, Yahoo, Facebook, Twitter, etc. applied algorithms on much larger amounts of data

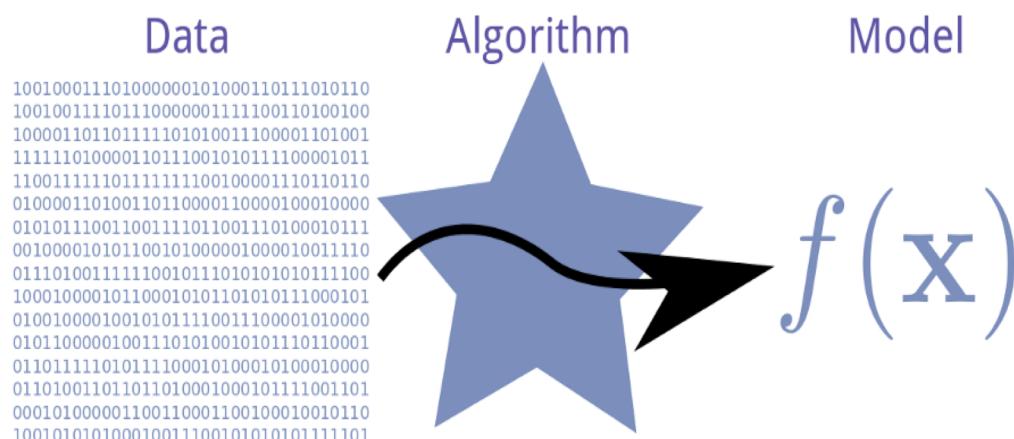


Source: Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.

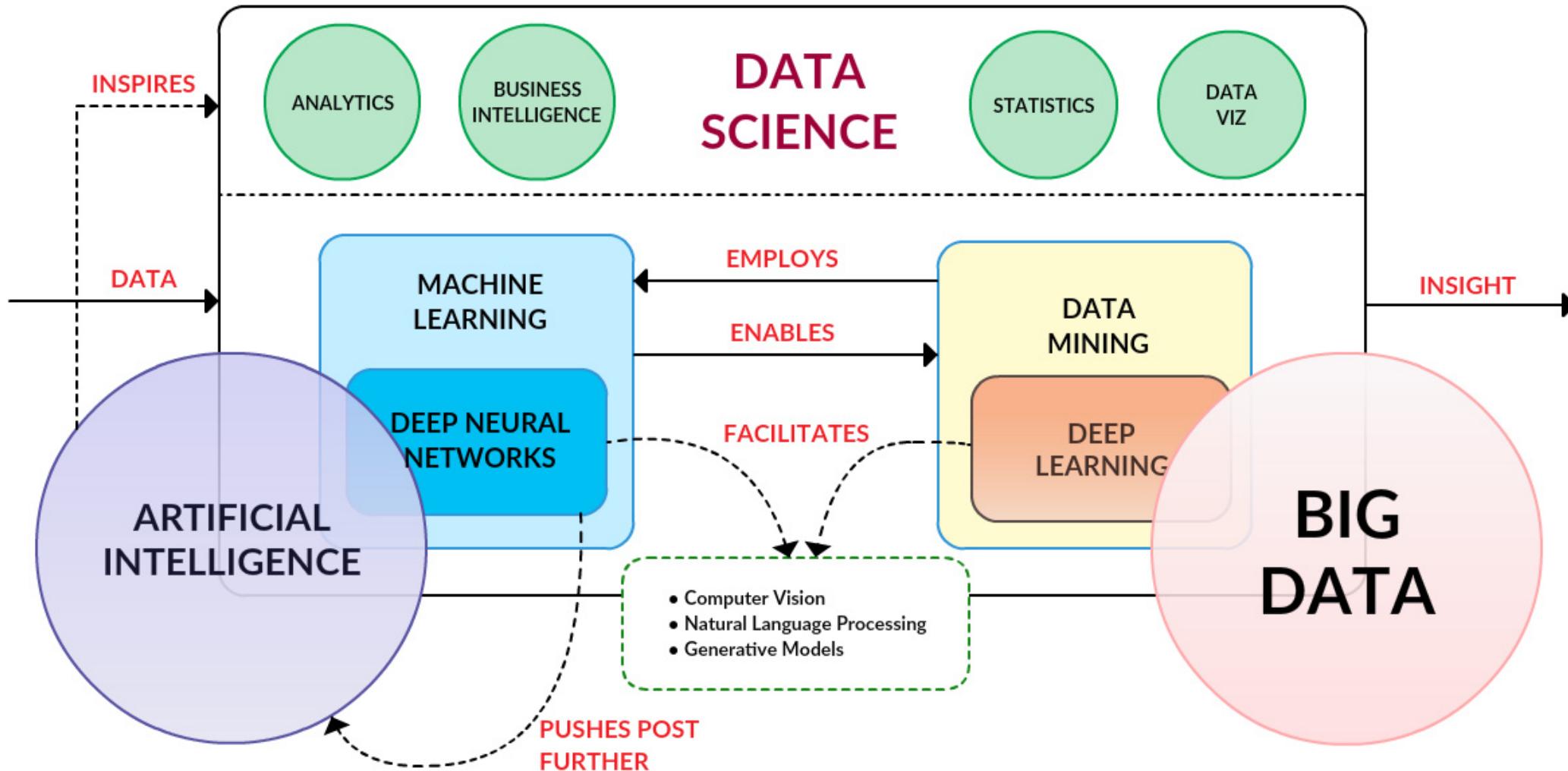
www.mikelnino.com

Clarifying jargon: Chronology

- Data Science – over the last decade
 - Term is used to indicate a field where complex algorithms work on large volumes of data to solve important business problems
 - A lot of emphasis on Visualization and storytelling to non-practitioners



Source: Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service. www.mikelnino.com



Business Intelligence Vs Data Analytics



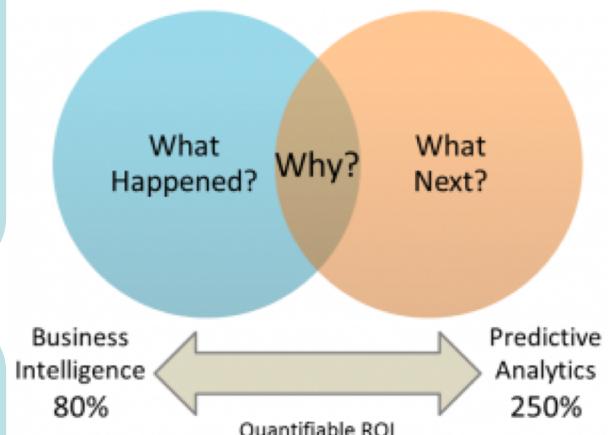
Business Intelligence (Descriptive)

Backward Looking
Limited to your data
Focus on the past



Data Analytics (Predictive & Prescriptive)

Forward Looking
Can take advantage & build upon intelligence existing elsewhere
Focus on the future



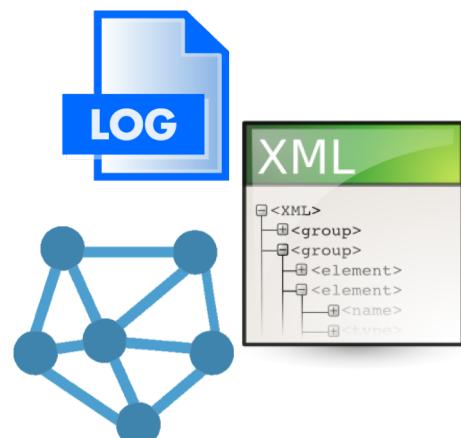
What are various Types of data?

Examples

Structured Data



Semi-structured Data



Unstructured Data



Data – Numeric and Categorical



18 kg



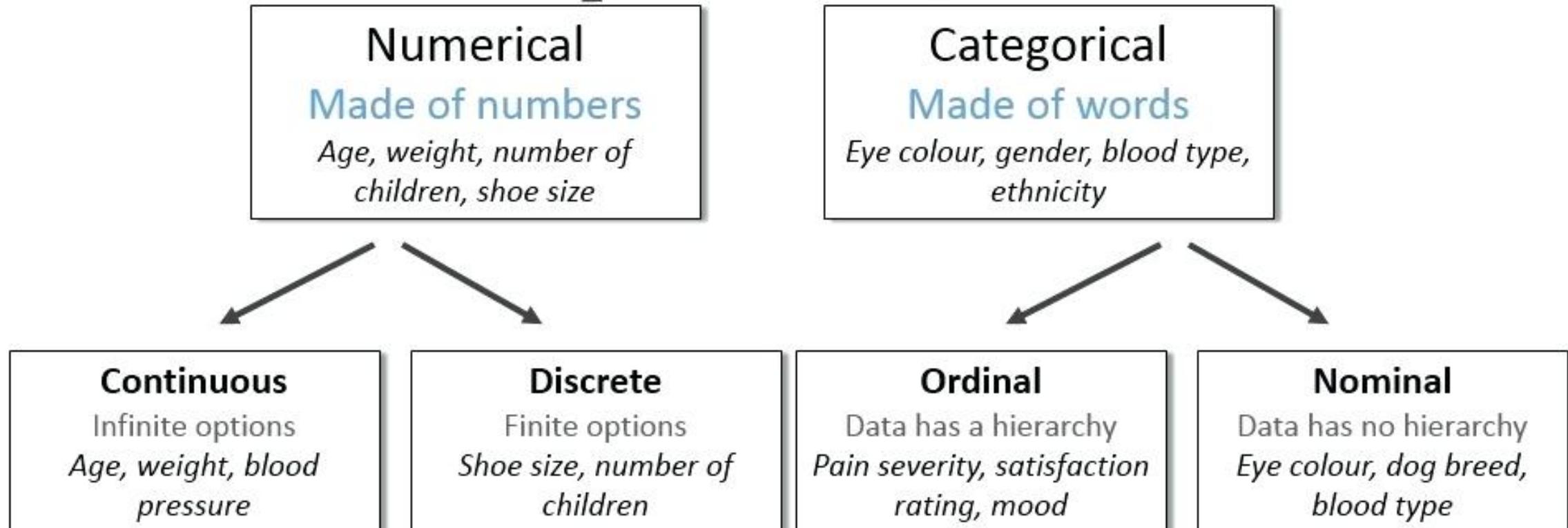
27 kg



Sources: <http://banglanews24.com/en/files/2013August/SM/Gold-sm20130830024804.jpg>, <http://myoor.com/wp-content/uploads/2014/01/gold.jpg> and <http://im.rediff.com/cricket/2014/feb/01india1.jpg>

Last accessed: November 22, 2014

Data



Variables and Data

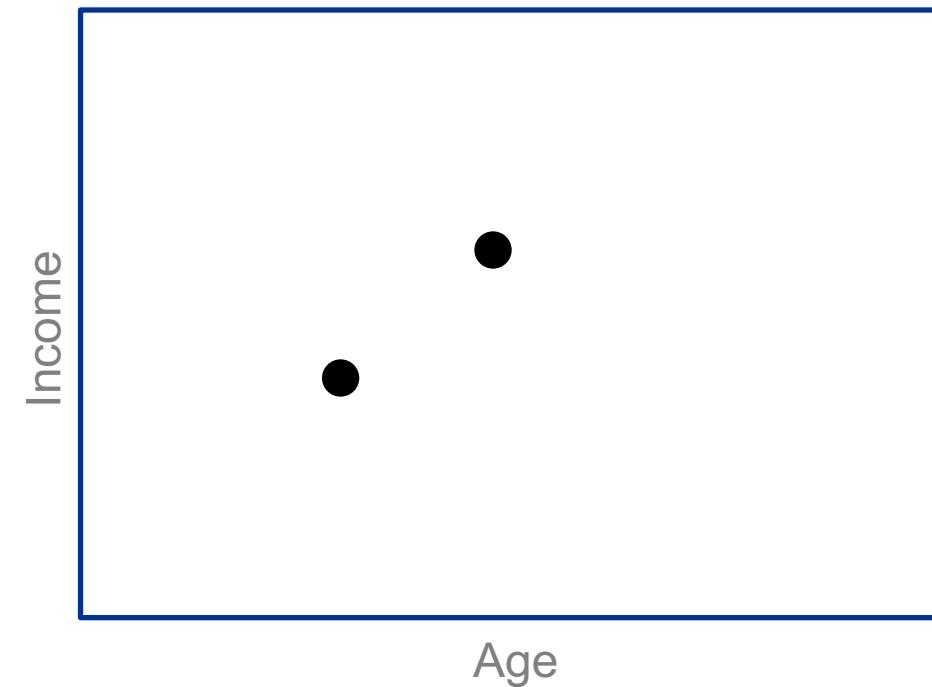
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	0	unknown	no
58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	0	unknown	no
43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	0	unknown	no
41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	0	unknown	no
29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	0	unknown	no
53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	0	unknown	no
58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	0	unknown	no
57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	0	unknown	no
51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	0	unknown	no
45	admin.	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	0	unknown	no
57	blue-collar	married	primary	no	52	yes	no	unknown	5	may	38	1	-1	0	unknown	no
60	retired	married	primary	no	60	yes	no	unknown	5	may	219	1	-1	0	unknown	no
33	services	married	secondary	no	0	yes	no	unknown	5	may	54	1	-1	0	unknown	no
28	blue-collar	married	secondary	no	723	yes	yes	unknown	5	may	262	1	-1	0	unknown	no
56	management	married	tertiary	no	779	yes	no	unknown	5	may	164	1	-1	0	unknown	no
32	blue-collar	single	primary	no	23	yes	yes	unknown	5	may	160	1	-1	0	unknown	no
25	services	married	secondary	no	50	yes	no	unknown	5	may	342	1	-1	0	unknown	no
40	retired	married	primary	no	0	yes	yes	unknown	5	may	181	1	-1	0	unknown	no



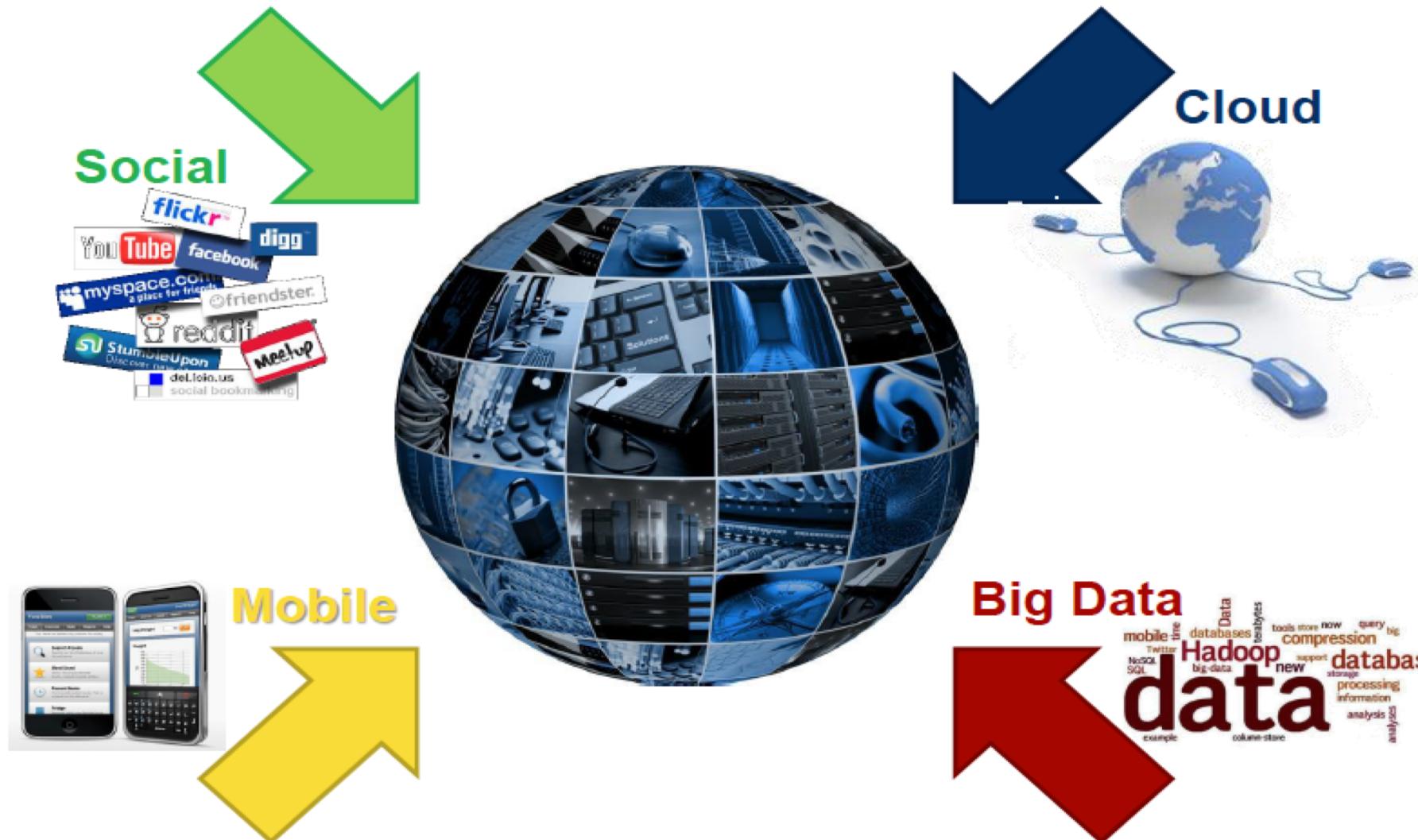
Structured Data Representation

- DB & Excel like format
- Each column represents one aspect of data and each row is one record
- The points represent a 2-dimensional vectors.
- Each point talks about a customer with age 'x' and income 'y'

	Age	Income
1	25	30
2	35	50



Data sources



What happens in an INTERNET MINUTE?



Volume

Variety

Velocity

Modeling philosophy

- In any business, there are some easy-to-measure metrics
 - Age; Gender; Income; Education level; etc. or in a sequence data recorded until then and sometimes a difficult-to-measure metric
 - Amount of loan to give; Will she buy or not; How many days will he stay in the hospital; etc.
- Unsupervised learning is about learning patterns in the easy to measure metrics
- Supervised learning is about computing the latter using the former

Supervised and Unsupervised learning

Supervised Learning



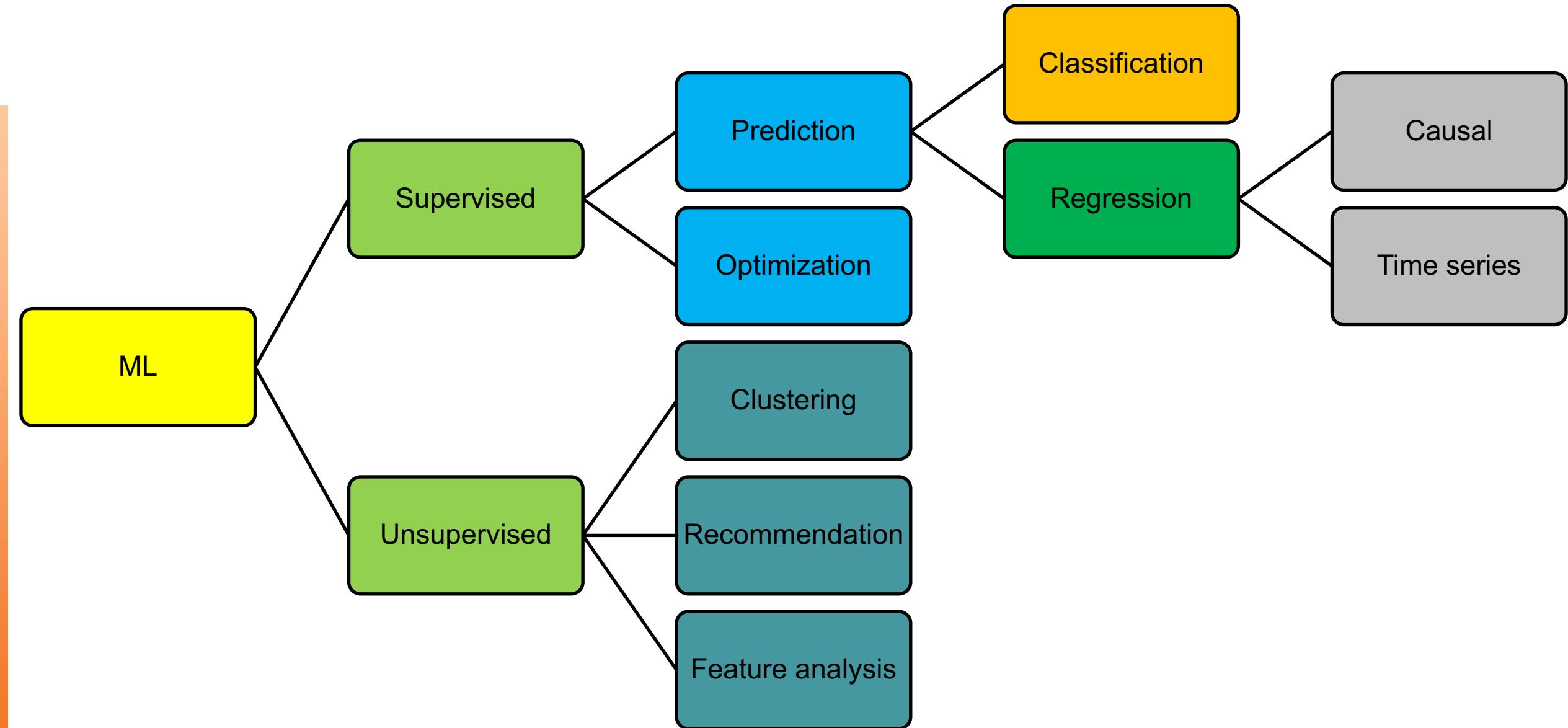
- Known number of classes/value
- Based on training samples
- Use to predict future value/class

Unsupervised Learning



- Unknown number of classes
- No prior knowledge
- Used to understand or explore data for patterns

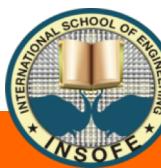




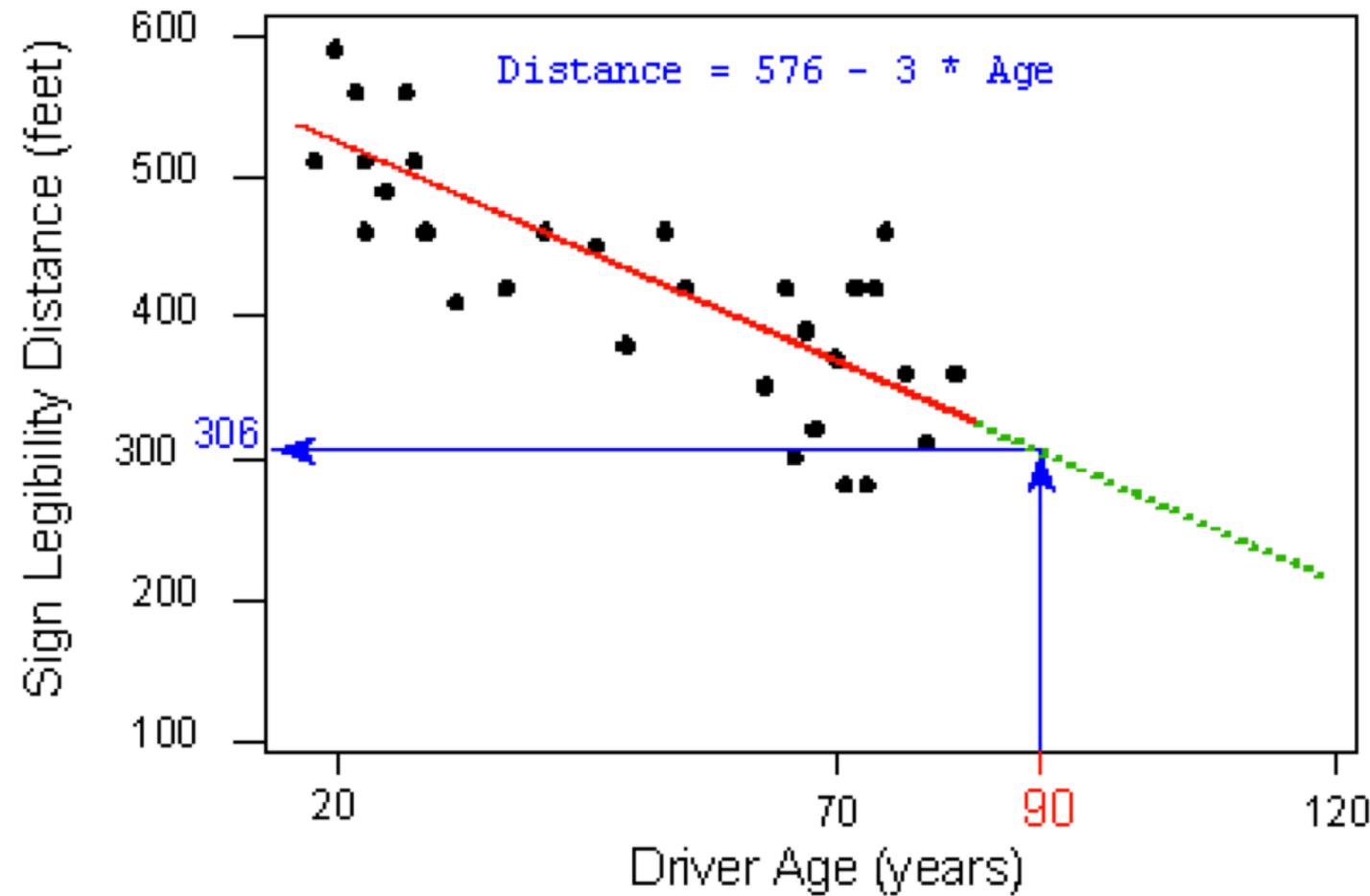
Machine Learning Algorithms (*sample*)

	<u>Unsupervised</u>	<u>Supervised</u>
Continuous	<ul style="list-style-type: none">• Clustering & Dimensionality Reduction<ul style="list-style-type: none">◦ SVD◦ PCA◦ K-means	<ul style="list-style-type: none">• Regression<ul style="list-style-type: none">◦ Linear◦ Polynomial• Decision Trees• Random Forests
Categorical	<ul style="list-style-type: none">• Association Analysis<ul style="list-style-type: none">◦ Apriori◦ FP-Growth• Hidden Markov Model	<ul style="list-style-type: none">• Classification<ul style="list-style-type: none">◦ KNN◦ Trees◦ Logistic Regression◦ Naive-Bayes◦ SVM

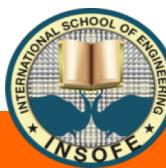
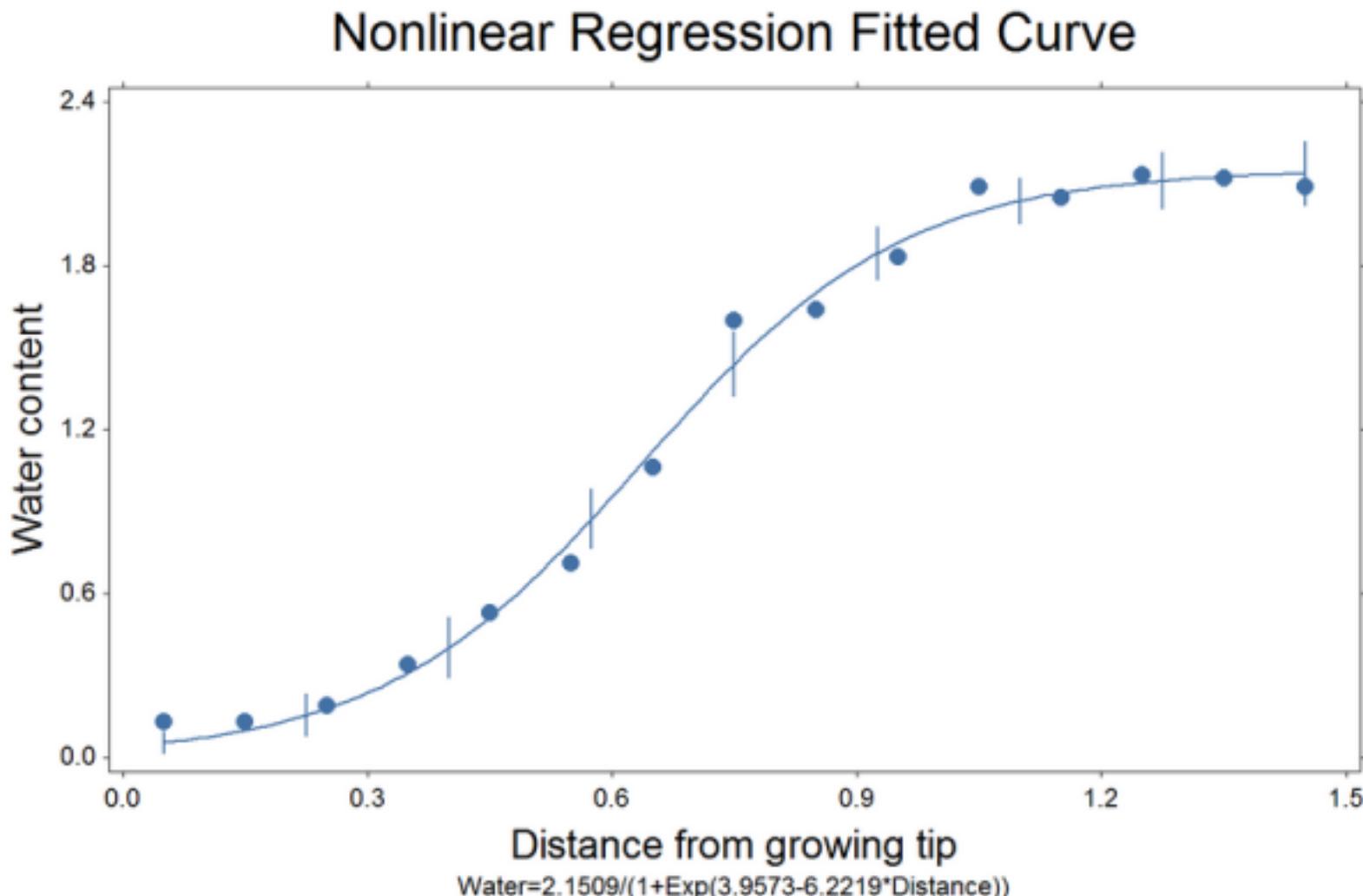
<http://www.differencebetween.net/technology/differences-between-supervised-learning-and-unsupervised-learning/>



Regression: Predicting a number based on some predictor variables



Regression can be non linear



Forecasting

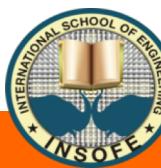


Source: <http://www.dailymail.co.uk/sciencetech/article-2120416/Twitter-predicts-stock-prices-accurately-investment-tactic-say-scientists.html>
Last accessed: August 4, 2014

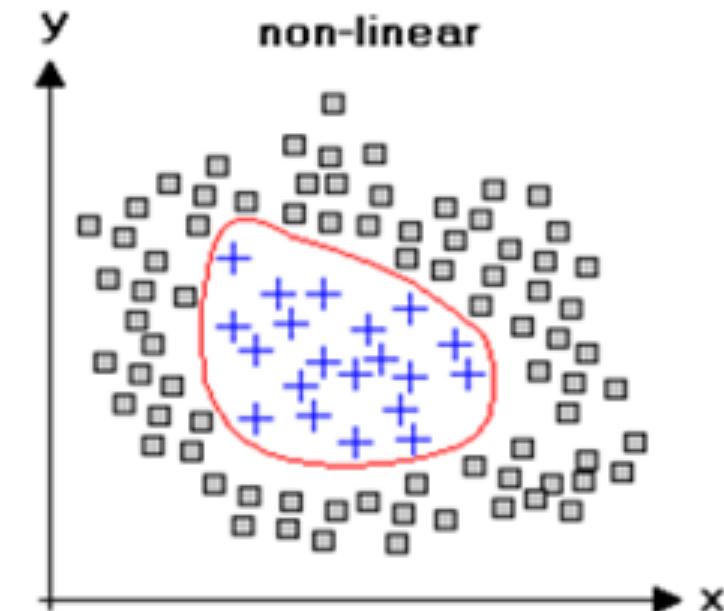
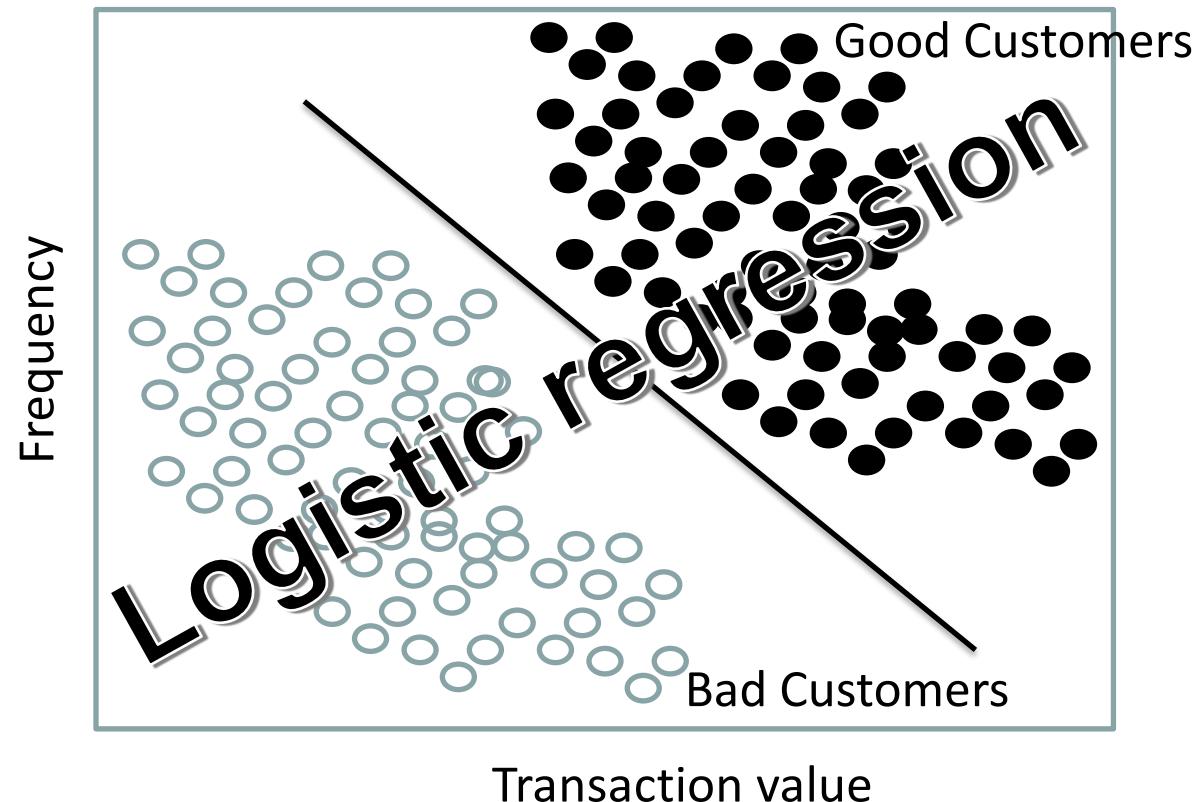
Classification

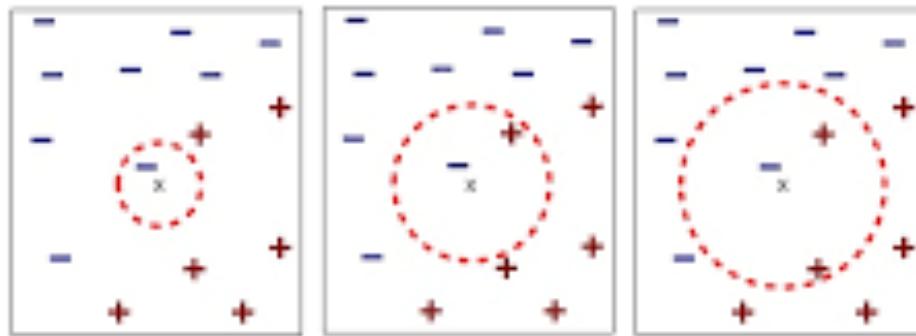


Predicting if an employee will quit based on available predictors such as Age, Recent promotion, performance review etc.



Classification





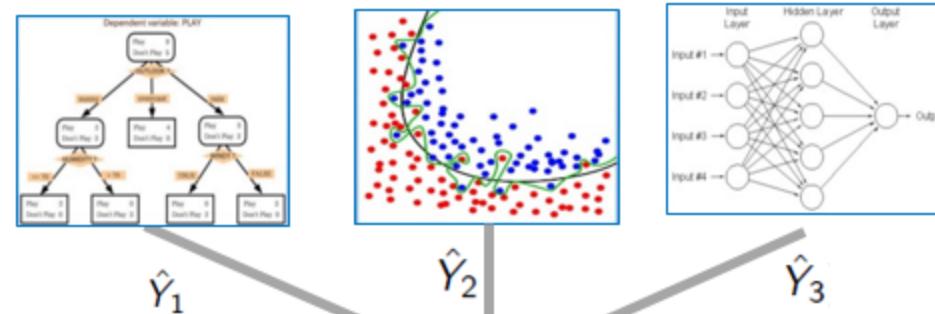
(a) 1-nearest neighbor

(b) 2-nearest neighbor

(c) 3-nearest neighbor

K-nearest neighbors of a record x are data points that have the k smallest distance to x

Mixture of experts



\hat{Y}_1

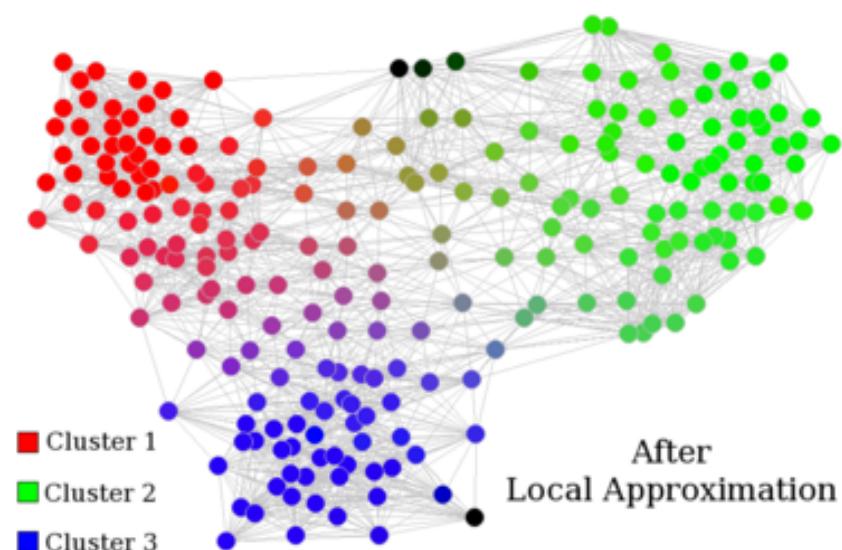
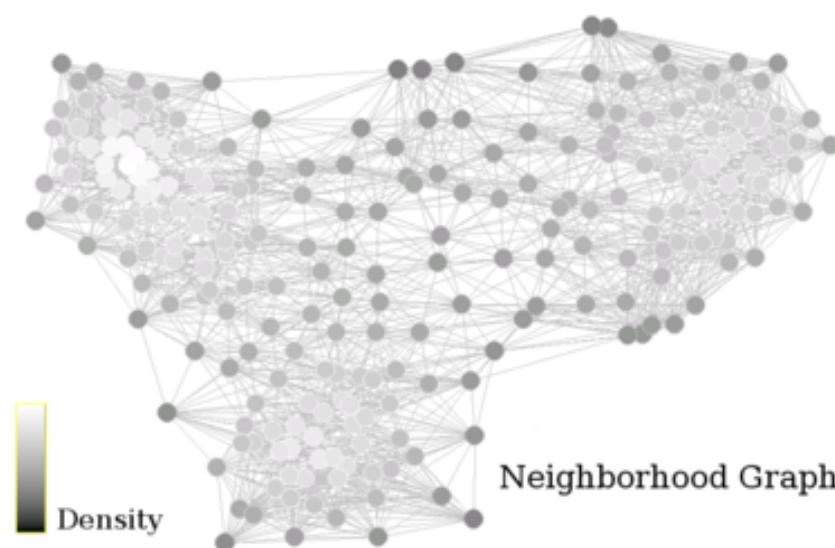
\hat{Y}_2

\hat{Y}_3

Second-level algorithm

New predictions

Clustering



Recommendation Engine



Write a Review 

Select Color 

Select Size (UK) 

SIZE CHART

Rs. 3499

Free home delivery

Sold By: WS Retail

1 OFFER

SAVE MORE: Save More: Shop for Rs.1199 or more and get 25% Off on Women's Clothing.See final price in cart. [View T&C](#)

[View details](#)

BUY NOW

Cash on Delivery may be available

30 Day Exchange Policy

Brand New 100% Original Pay Securely Easy Returns

Check your delivery options:
Enter Pincode

Standard delivery in 2-3 business days. ?
Faster Delivery may be available ?

CUSTOMERS WHO VIEWED THIS PRODUCT ALSO VIEWED



French Connection
Rs. 3499

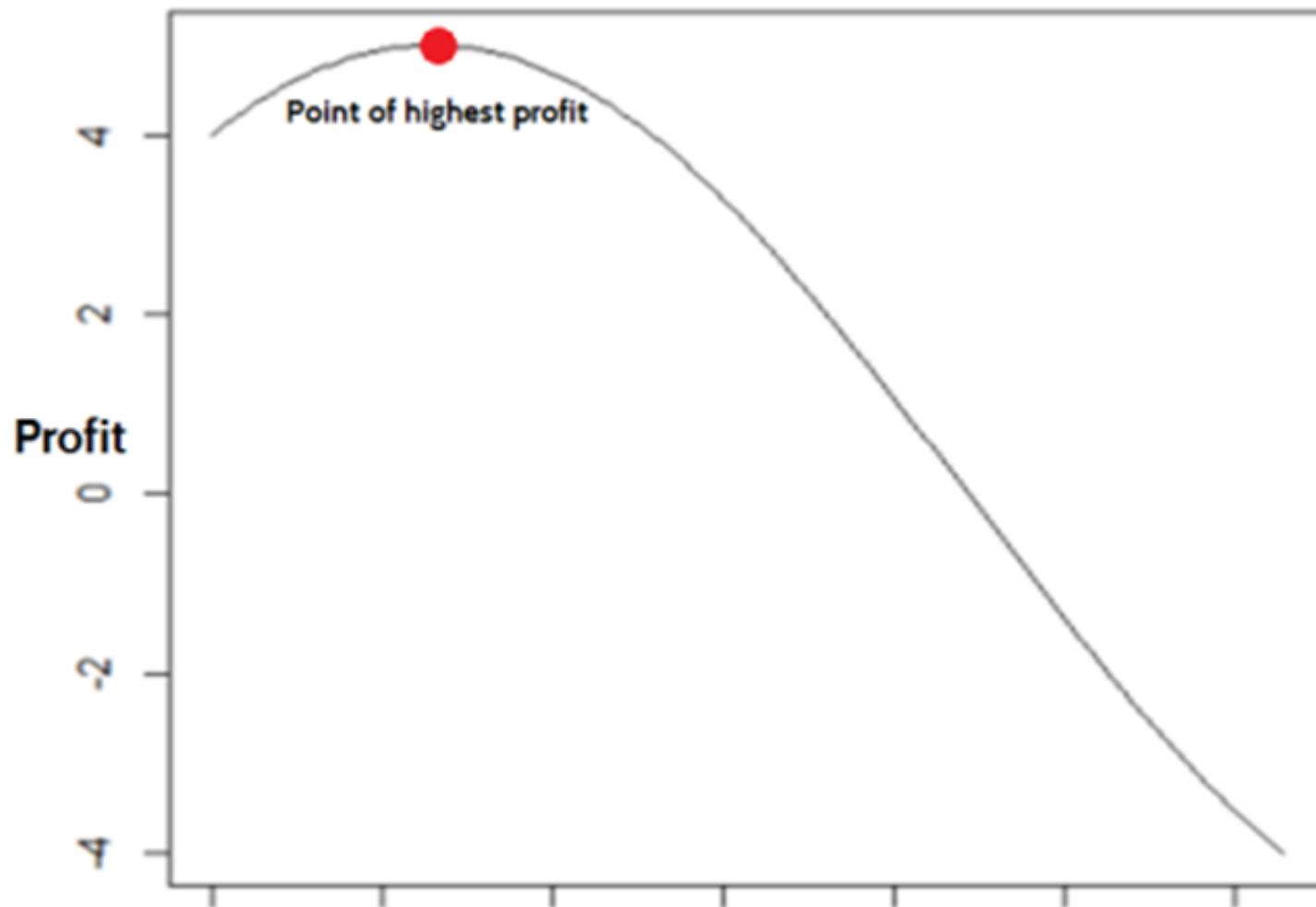
Flying Machine
Rs. 1699

Flying Machine
Rs. 1699

People
Rs. 1199

People
Rs. 1199

Optimization



Descriptive Analytics



What has happened?

Predictive Analytics



What could happen in the future based
on previous trends and patterns?"

Prescriptive Analytics



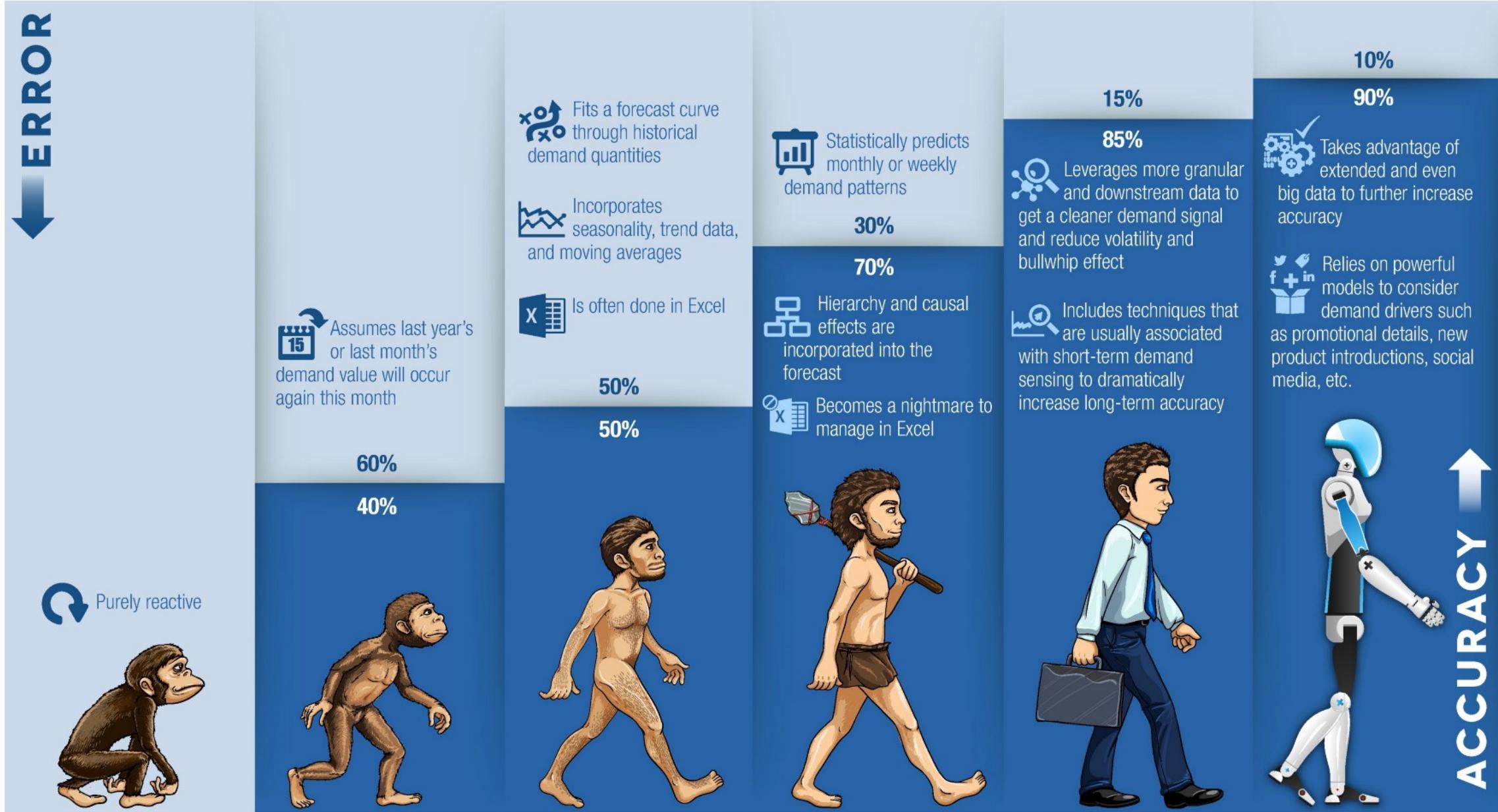
What should a business do?

The General Tasks of a Data Scientist

1. Get a little domain understanding
2. Define the problem statement well



ERROR ↓



Understanding the Problem



The Approach

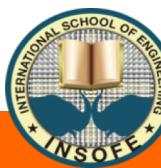


Results - The Streak!

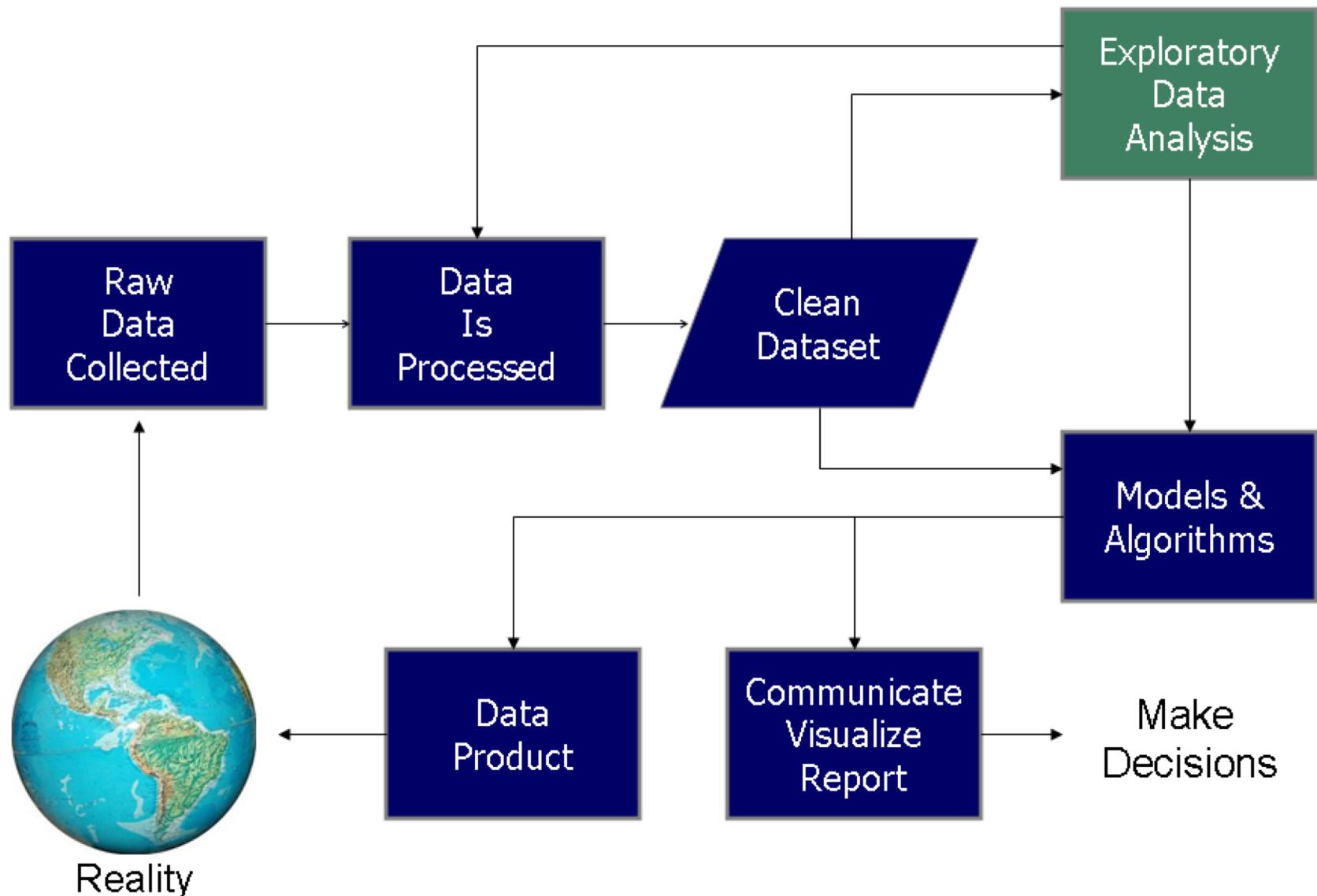


The General Tasks of a Data Scientist

3. Pre-process data to fix data issues like duplicates, missing values, etc.
4. Visualize data to the extent possible for better understanding and to see basic patterns
5. Identify what kind of a problem it is (Prediction/Forecasting, Classification, Optimization and/or Managing Big Data)
6. Identify appropriate modeling techniques and build models
7. Analyze results and iterate, as needed; DO NOT trust software outputs blindly
8. Visualize outputs and Communicate



Data Science Process



Iterative in nature

-From Wikipedia



Medical non-adherence is likened to be an invisible epidemic

Non-adherence = # of doses not taken or taken incorrectly that jeopardizes patient's therapeutic outcome

Note: Lack of adherence to non-pharmacologic treatments, such as recommended lifestyle changes, is not included



Stakeholders	Medical non-adherence impact	Est. \$\$\$	INDUSTRY ACTIONS
<ul style="list-style-type: none"> • Insurers • Employers • Patients 	<ul style="list-style-type: none"> • Increases healthcare costs due to disease-related complications 	<ul style="list-style-type: none"> • \$ 290 B Source: New England Healthcare Institute 	<ul style="list-style-type: none"> • Free drugs (<i>Netherlands</i>) • Lower co-pays (<i>Aetna</i>) • Reminders • Reward points (<i>HealthPrize</i>) • More targeted reminders (<i>RxAnte</i> uses algorithms)
<ul style="list-style-type: none"> • Pharmaceutical companies • Pharmacies • Pharmacy benefit managers 	<ul style="list-style-type: none"> • Erodes profits due to prescriptions never filled and medications not taken often enough 	<ul style="list-style-type: none"> • \$ 188 B Source: Capgemini 	

The entire medical industry is scrambling to address this issue !!!

Advanced analytics may provide the insights to help solve this puzzle

Source: http://pharmaceuticalcommerce.com/business_finance?articleid=26718

Stock Market Trends

Google searches can predict stock market crashes: Study

London, July 29: A rise in Google searches for terms relating to business and politics can predict a future stock market crash, researchers have claimed.

A team of researchers from Warwick Business School in the UK and Boston University in the US has developed a method to automatically identify topics that people search for on Google before subsequent stock market falls. Applied to data between 2004 and 2012, the method shows

that increases in searches for business and politics preceded falls in the stock market.

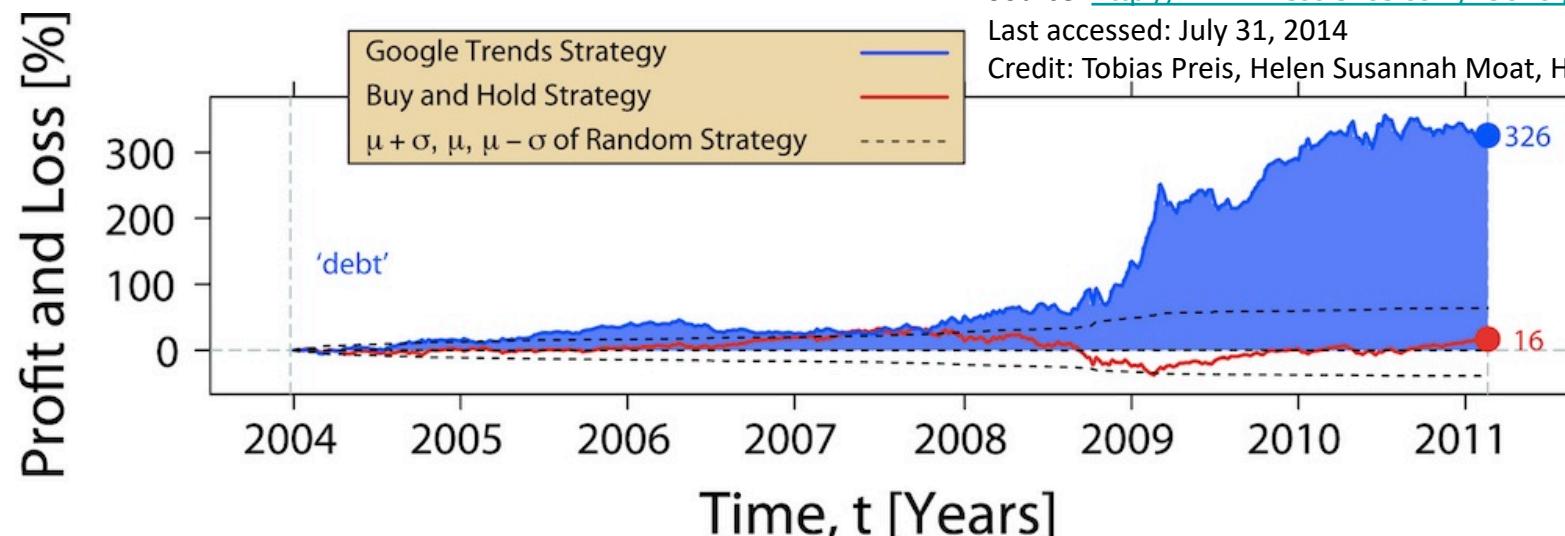
The researchers suggest that this method could be applied to help identify warning signs in search data before a range of real world events. "Search engines, such as Google, record almost everything we search for," said Chester Curme, research fellow at Warwick Business School and lead author of the study. "Records of these search queries allow us to learn about how people gather information online before making decisions in the real world. So there's potential to use these search data to anticipate what large groups of people may do," Mr Curme said.

In previous studies, Mr Curme and his colleagues, Tobias Preis and Suzy Moat of WBS, and H Eugene Stanley of BU, have demonstrated that usage data from Google and Wikipedia may contain early warning signs of stock market moves.

— PTI

Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=732781>

Last accessed: July 31, 2014



Source: <http://www.livescience.com/29016-google-predicts-stock-market.html>

Last accessed: July 31, 2014

Credit: Tobias Preis, Helen Susannah Moat, H. Eugene Stanley

Other Important Applications

- **Healthcare** – Non-compliance in taking prescription drugs
- **Banking and Insurance** – Fraud detection; Credit scoring; Detecting money laundering; Forecasting stock prices
- **Travel and Hospitality** – Improve customer experience
- **Politics** – Predict winners



Other Important Applications

- **Retail and Telecommunications** – Customer retention; Improving customer service quality; Planning store locations; Recommendation systems; Sales forecasting
- **Government** – Policy planning; Effective use of resources; Security against terrorist attacks; Effective policing by understanding crime patterns; Weather predictions; Calamity predictions



Other Applications

- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

Retail



- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

Manufacturing



- Alerts and diagnostics from real-time patient data
- Disease identification and risk stratification
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

Healthcare and Life Sciences



- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
- Energy demand and supply optimization

Energy, Feedstock, and Utilities



- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

Travel and Hospitality



- Risk analytics and regulation
- Customer Segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management
- Credit worthiness evaluation

Financial Services



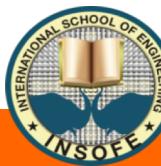


How best to make out of this program?

- Study 10-15 hrs during the week and prepare well for your exams
- Practice, practice, and practice is the only mantra.



Data Scientists Profiles



Seema Basantani – Sr. Data Scientist

Qualification: Masters in Computer Application

Work Exp.: 12+ years in IT industry, Worked with IBM India Private Limited and Aztecsoft

Highlights:

- Focus areas include Statistical Modelling, R and Python Programming, Machine Learning and Text Mining
- Worked with US client to provide support for building Chatbot solution
- Worked with UK Telecom client for solutioning of multiple projects related to Billing domain
- Recipient of IBM Bravo and Eminence & Excellence Awards





Mahidhar Vydyam - Data Scientist

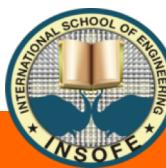
Qualification: M.Tech

Work Exp.: over 7 years worked as Asst. prof

Competencies: **Big Data,R,Python,Statistical Modelling, AI, and ML, and Optimization**

Key Data Science Projects:

- Worked on Big data and ML aspects of Customer and product segmentation for a German Farm technology client.
- Worked on forecasting sales of the products for a Customer using Time Series and ML.
- Worked on ML aspects of Chat Analytics for a American Clothing client.
- Worked on a Big Data Solution for Fraud Detection with a American Client





Chaithanya Kumar - Data Scientist

Qualification: B.Tech in Mechanical Engineering

Work Exp.: 4.5 years.

Worked for Amazon and Sutherland Healthcare Solutions.

Competencies: Statistical Modelling, ML, AI, Big Data, R, Python, Visualization, Power BI and Tableau.

Highlights:

- Worked on Classification of Insurance claims for Healthcare Industry (Denial Analytics).
- Worked on a Big Data based solution for Fraud Detection.
- Power BI dashboard development.
- Genetic Algorithm for Route optimization in python.





Tarun Kumar - Data Scientist

Qualification: B.Tech in Electrical Engineering

Work Exp.: 2 years.

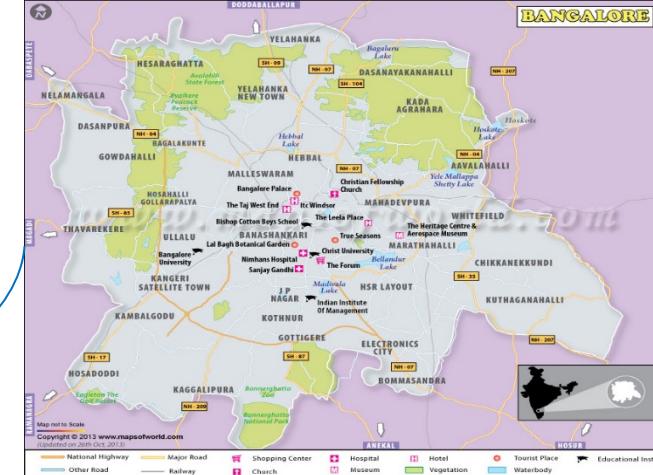
Worked for Larsen & Toubro – Marketing Solutions and Human Resourcing Products.

Competencies: Statistical Modelling, ML, AI, Big Data, R, Python, Visualization and Tableau.

Highlights:

- Worked on Product development for Automation of Human Resourcing (HR Analytics)
- Worked on a Text Analytics based solutions in Banking sector
- Business Analytics
- Shiny dashboard development for Retail Clients





HYDERABAD

2nd Floor, Jyothi Imperial, Vamsiram Builders, Old Mumbai Highway, Gachibowli, Hyderabad - 500 032
 +91-9701685511 (Individuals)
 +91-9618483483 (Corporates)

BENGALURU

L77, 15th Cross Road, 3rd Main Road, Sector 6, HSR Layout, Bengaluru – 560 102
 +91-9502334561 (Individuals)
 +91-9502799088 (Corporates)

Social Media

- Web: <http://www.insofe.edu.in>
- Facebook: <https://www.facebook.com/insofe>
- Twitter: <https://twitter.com/Insofeedu>
- YouTube: <http://www.youtube.com/InsofeVideos>
- SlideShare: <http://www.slideshare.net/INSOFE>
- LinkedIn: <http://www.linkedin.com/company/international-school-of-engineering>

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.