

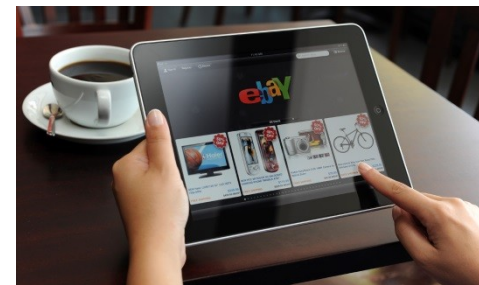
Data Analytics

Tushar B. Kute,
<http://tusharkute.com>



Data All Around

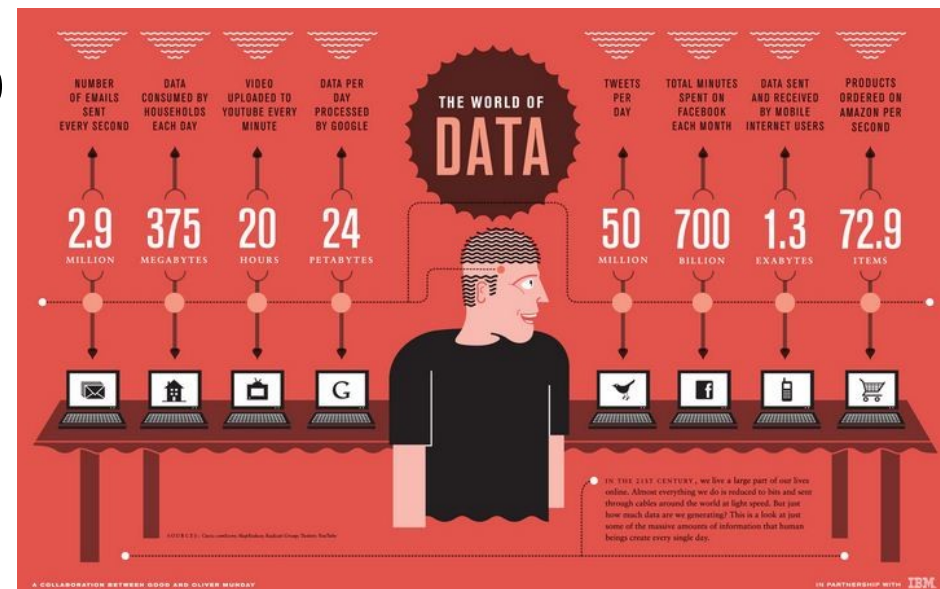
- Lots of data is being collected and war
 - Web data, e-commerce
 - Financial transactions, bank/credit transactions
 - Online trading and purchasing
 - Social Network
 - Cloud



How much data we have?

- Google processes 120+ PB a day
- Facebook has 80 TB of daily logs
- eBay has 16.5 PB of user data + 150 TB/day

- Cost of 1 TB of disk: Rs.4000
- Time to read 1 TB disk:
3 hrs (100 MB/s)



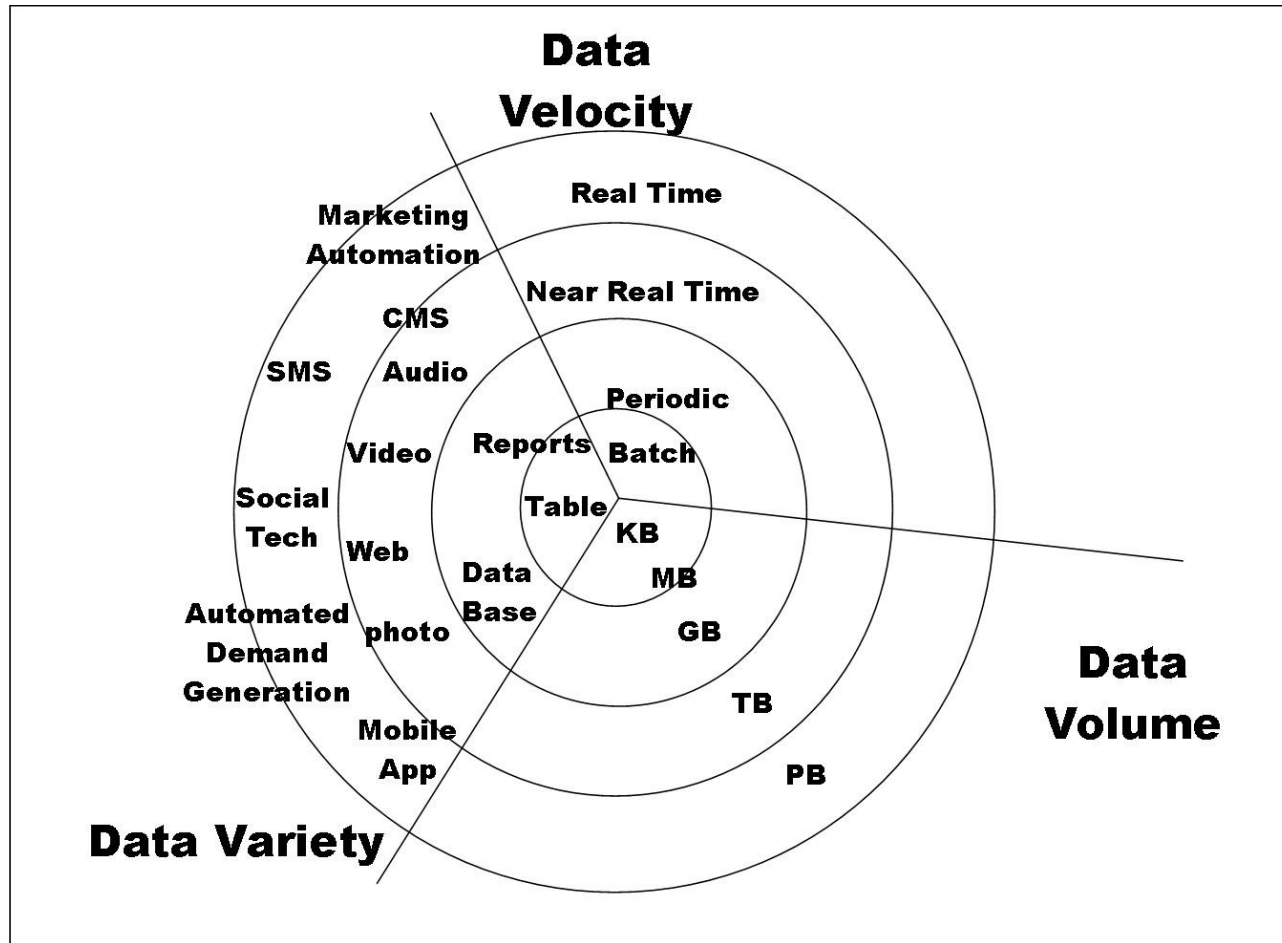
YouTube

- 100 hours of video are uploaded to YouTube every minute then have you ever wondered how can google store all such data and what is the youtube servers size in GBs?
- Factor for 720p: $(93.8614355\text{MB} - 25.9\text{MB}) / (93.8614355\text{MB}) = 0.72406132655$
- That would result in:
 - 191.141537 gigabytes/minute
 - 11.1996994 terabytes/hour
 - 268.792787 terabytes/day
 - 7.75157161 petabytes/month
 - 93.0188597 petabytes/year and according to research
- 15 years of youtube data is stored on lots of hard drives. Not only that but the same data is stored on multiple hard drives may be 100,000 hard drives stacked in server and size of the server would be 10,000,000,000,000,000,000 Bytes (That's 19 '0's, or 10 "Exabytes"). or 1 sextillion gigabytes of data

Big Data

- Big Data is any data that is expensive to manage and hard to extract value from
 - Volume
 - The size of the data
 - Velocity
 - The latency of data processing relative to the growing demand for interactivity
 - Variety and Complexity
 - The diversity of sources, formats, quality, structures.

Big Data



Types of Data

- Relational Data
(Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
- Social Network, Semantic Web (RDF), ...
- Streaming Data

What to do with this data?

- Aggregation and Statistics
 - Data warehousing and OLAP
- Indexing, Searching, and Querying
 - Keyword based search
 - Pattern matching (XML/RDF)
- Knowledge discovery
 - Data Mining
 - Statistical Modeling

What is Data Science ?

- An area that manages, manipulates, extracts, and interprets knowledge from tremendous amount of data.
- Data science (DS) is a multidisciplinary field of study with goal to address the challenges in big data.
- Data science principles apply to all data – big and small.

What is Data Analytics ?

- Data analytics is the science of analyzing raw data in order to make conclusions about that information.
- Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.
- Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

Processes in Data Analytics

- The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demographic, income, or gender. Data values may be numerical or be divided by category.
- The second step in data analytics is the process of collecting it. This can be done through a variety of sources such as computers, online sources, cameras, environmental sources, or through personnel.

Processes in Data Analytics

- Once the data is collected, it must be organized so it can be analyzed. Organization may take place on a spreadsheet or other form of software that can take statistical data.
- The data is then cleaned up before analysis. This means it is scrubbed and checked to ensure there is no duplication or error, and that it is not incomplete. This step helps correct any errors before it goes on to a data analyst to be analyzed.

Data Analytics Streams

- Theories and techniques from many fields and disciplines are used to investigate and analyze a large amount of data to help decision makers in many industries such as science, engineering, economics, politics, finance, and education
 - Computer Science
 - Pattern recognition, visualization, data warehousing, High performance computing, Databases, AI
 - Mathematics
 - Mathematical Modeling
 - Statistics
 - Statistical and Stochastic modeling, Probability.

Data Analytics Streams

- Descriptive analytics
- Diagnostic analytics
- Predictive analytics
- Prescriptive analytics

Data Analytics Streams

- Descriptive analytics describes what has happened over a given period of time. Have the number of views gone up? Are sales stronger this month than last?
- Diagnostic analytics focuses more on why something happened. This involves more diverse data inputs and a bit of hypothesizing. Did the weather affect beer sales? Did that latest marketing campaign impact sales?

Data Analytics Streams

- Predictive analytics moves to what is likely going to happen in the near term. What happened to sales the last time we had a hot summer? How many weather models predict a hot summer this year?
- Prescriptive analytics suggests a course of action. If the likelihood of a hot summer is measured as an average of these five weather models is above 58%, we should add an evening shift to the brewery and rent an additional tank to increase output.

Ways to use Data Analytics

- Improved Decision Making
- More Effective Marketing
- Better Customer Service
- More Efficient Operations

Improved Decision Making

- Companies can use the insights they gain from data analytics to inform their decisions, leading to better outcomes.
- Data analytics eliminates much of the guesswork from planning marketing campaigns, choosing what content to create, developing products and more.
- It gives you a 360-degree view of your customers, which means you understand them more fully, enabling you to better meet their needs. Plus, with modern data analytics technology, you can continuously collect and analyze new data to update your understanding as conditions change.

More effective marketing

- When you understand your audience better, you can market to them more effectively. Data analytics also gives you useful insights into how your campaigns are performing so that you can fine-tune them for optimal outcomes.
- You can gain insights into which audience segments are most likely to interact with a campaign and convert. You can use this information to adjust your targeting criteria either manually or through automation, or use it to develop different messaging and creative for different segments.
- Improving your targeting results in more conversions and less ad waste.

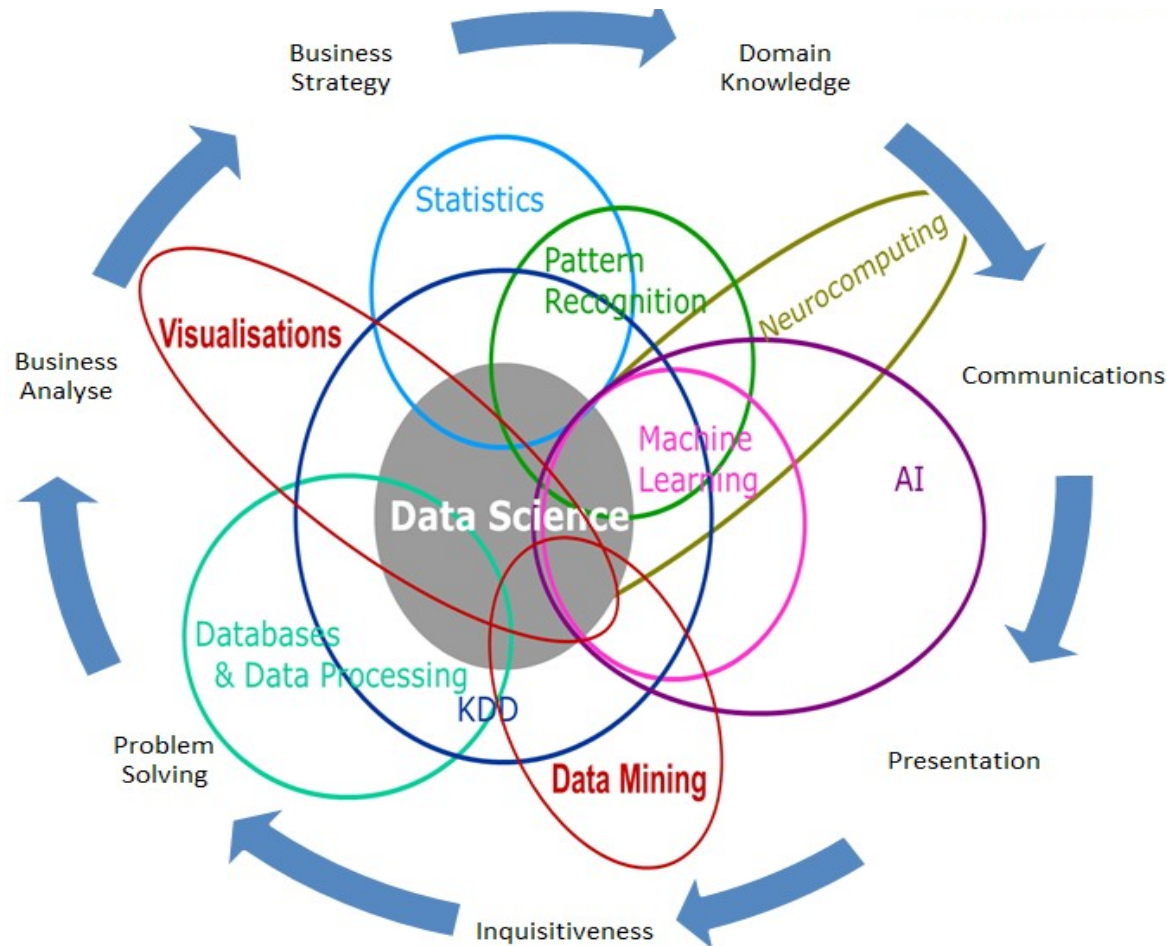
Better Customer Service

- Data analytics provide you with more insights into your customers, allowing you to tailor customer service to their needs, provide more personalization and build stronger relationships with them.
- Your data can reveal information about your customers' communications preferences, their interests, their concerns and more. Having a central location for this data also ensures that your whole customer service team, as well as your sales and marketing teams, are on the same page.

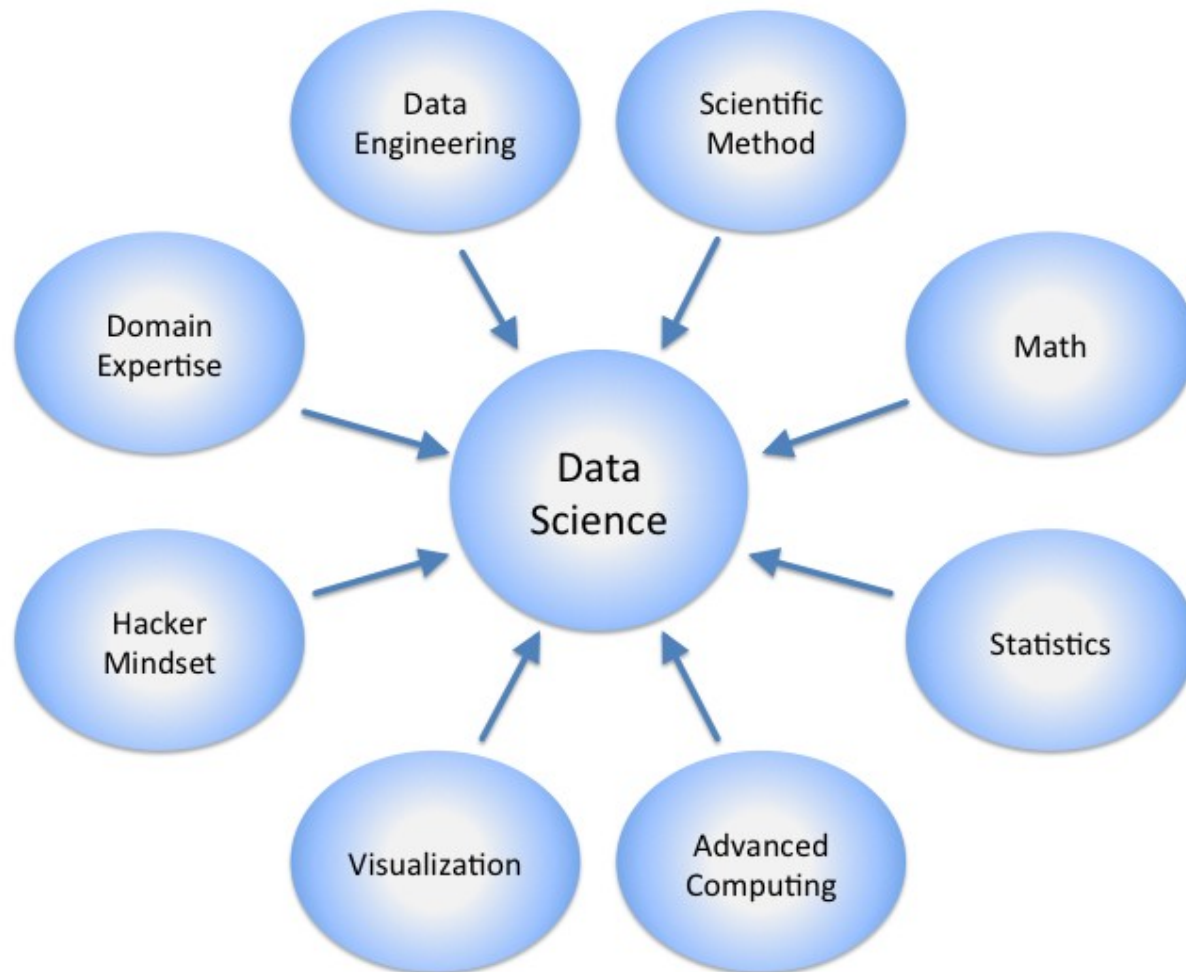
More efficient operations

- Data analytics can help you streamline your processes, save money and boost your bottom line. When you have an improved understanding of what your audience wants, you waste less time on creating ads and content that don't match your audience's interests.
- This means less money wasted as well as improved results from your campaigns and content strategies. In addition to reducing your costs, analytics can also boost your revenue through increased conversions, ad revenue or subscriptions.

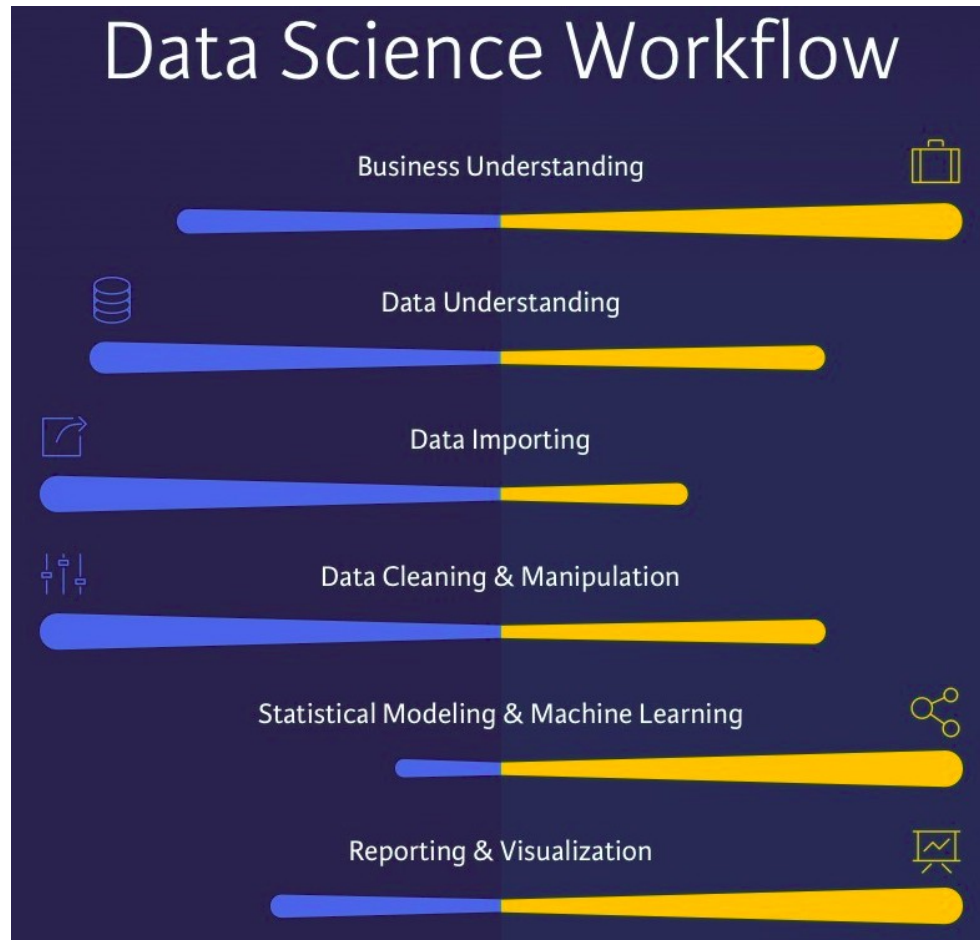
Data Analytics Disciplines



Data Science



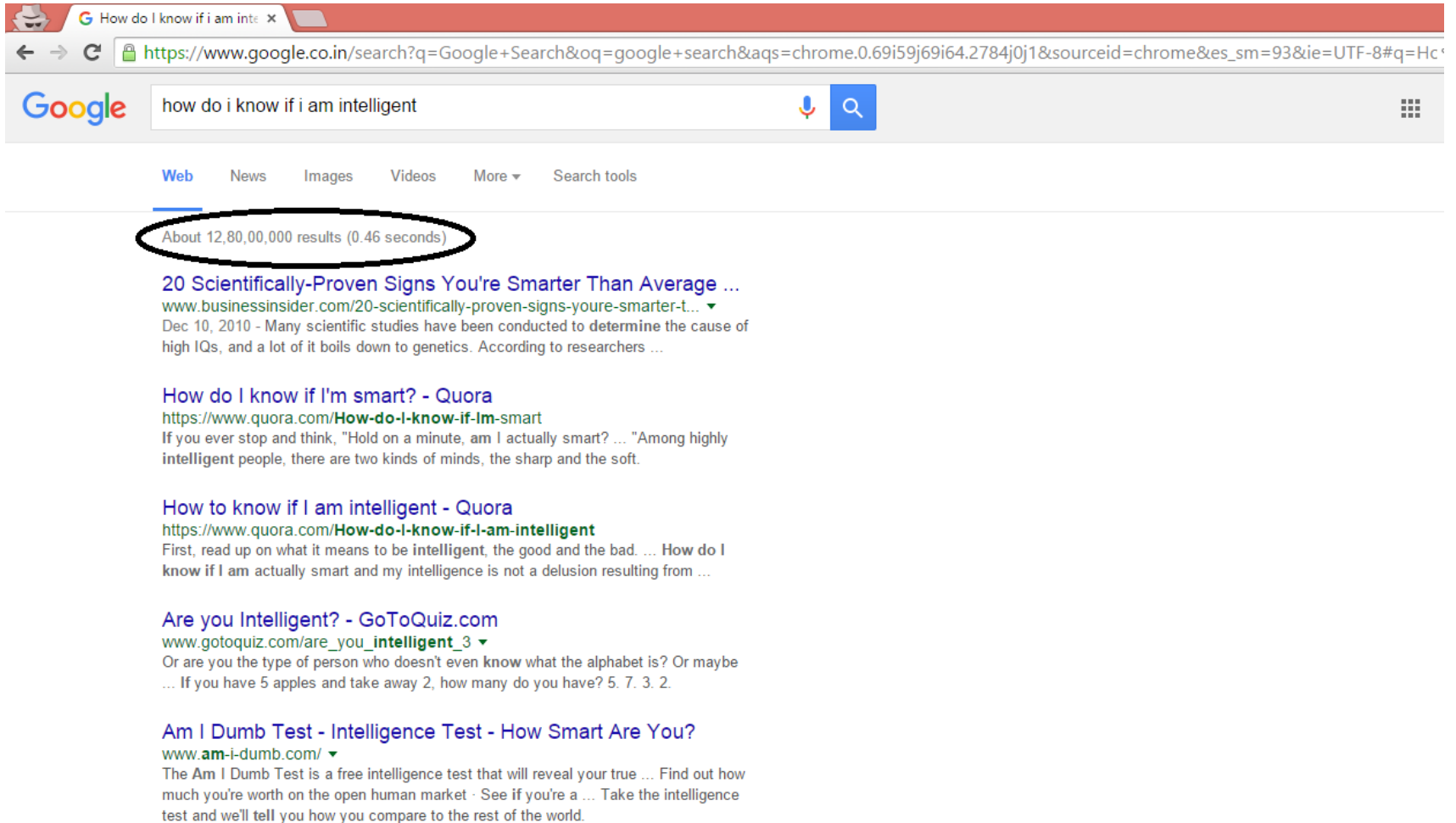
Data Science Workflow



Real Life Examples

- Internet Search
- Digital Advertisements (Targeted Advertising and re-targeting)
- Recommender Systems
- Image Recognition
- Speech Recognition
- Gaming
- Price Comparison Websites
- Airline Route Planning
- Fraud and Risk Detection
- Delivery logistics

Internet Search



The screenshot shows a Google search results page. The search bar contains the text "how do i know if i am intelligent". Below the search bar, the results are displayed. The first result is titled "20 Scientifically-Proven Signs You're Smarter Than Average ..." and is from the website "www.businessinsider.com". The second result is titled "How do I know if I'm smart? - Quora" and is from the website "https://www.quora.com/How-do-I-know-if-I-m-smart". The third result is titled "How to know if I am intelligent - Quora" and is from the website "https://www.quora.com/How-do-I-know-if-I-am-intelligent". The fourth result is titled "Are you Intelligent? - GoToQuiz.com" and is from the website "www.gotoquiz.com/are_you_intelligent_3". The fifth result is titled "Am I Dumb Test - Intelligence Test - How Smart Are You?" and is from the website "www.am-i-dumb.com/".

How do I know if i am inte x

https://www.google.co.in/search?q=Google+Search&oq=google+search&aqs=chrome.0.69i59j69i64.2784j0j1&sourceid=chrome&es_sm=93&ie=UTF-8#q=Hc

Google how do i know if i am intelligent

Web News Images Videos More Search tools

About 12,80,00,000 results (0.46 seconds)

20 Scientifically-Proven Signs You're Smarter Than Average ...
www.businessinsider.com/20-scientifically-proven-signs-youre-smarter-t...
Dec 10, 2010 - Many scientific studies have been conducted to determine the cause of high IQs, and a lot of it boils down to genetics. According to researchers ...

How do I know if I'm smart? - Quora
https://www.quora.com/How-do-I-know-if-I-m-smart
If you ever stop and think, "Hold on a minute, am I actually smart? ... "Among highly intelligent people, there are two kinds of minds, the sharp and the soft.

How to know if I am intelligent - Quora
https://www.quora.com/How-do-I-know-if-I-am-intelligent
First, read up on what it means to be intelligent, the good and the bad. ... How do I know if I am actually smart and my intelligence is not a delusion resulting from ...



Are you Intelligent? - GoToQuiz.com
www.gotoquiz.com/are_you_intelligent_3
Or are you the type of person who doesn't even know what the alphabet is? Or maybe ... If you have 5 apples and take away 2, how many do you have? 5. 7. 3. 2.

Am I Dumb Test - Intelligence Test - How Smart Are You?
www.am-i-dumb.com/
The Am I Dumb Test is a free intelligence test that will reveal your true ... Find out how much you're worth on the open human market - See if you're a ... Take the intelligence test and we'll tell you how you compare to the rest of the world.

Targeting Advertisement

GIFMaker.me

Home Resize Image Compress Image Password Generator MP3 Cutter Free Proxy PNG to ICO

**Hootsuite - 60 Days Free**

Enjoy 60 days free in celebration of our enhanced Instagram features.

Introduction

GIFMaker.me allows you to create animated gifs, slideshows, and video animations with music online freely and easily, no registration required.

flexjobs
FIND A BETTER WAY TO WORK

Telecommuting & FlexJobs

30% Off Today




Select multiple images:

Upload Images


4 images uploaded successfully

You can make your gif now

Drag the images to change the order:



Control Panel:



Canvas size:

100 % 1280 x 851 px

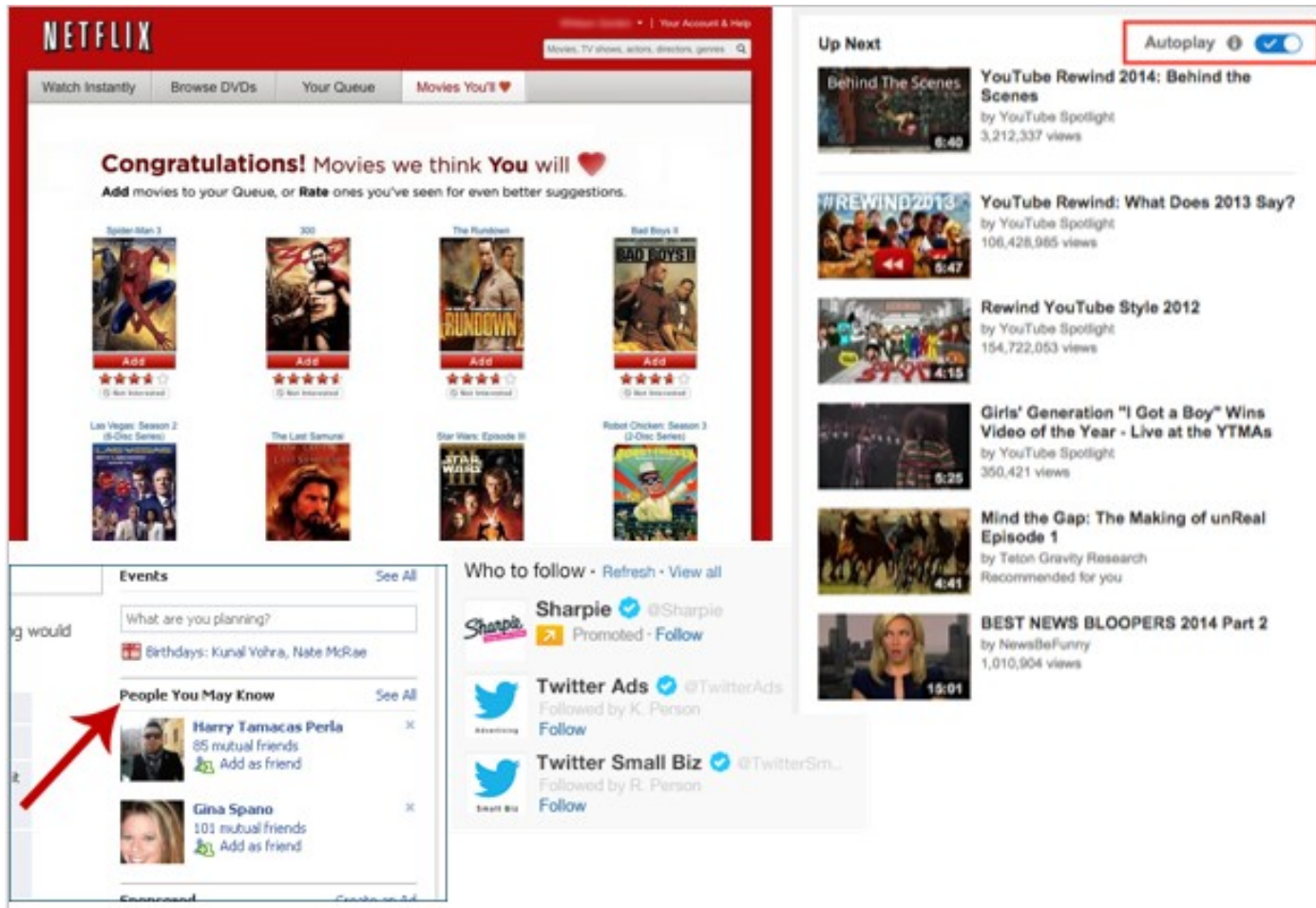
Animation speed:

500 milliseconds

Repeat times (0 = infinite loop):

0 times

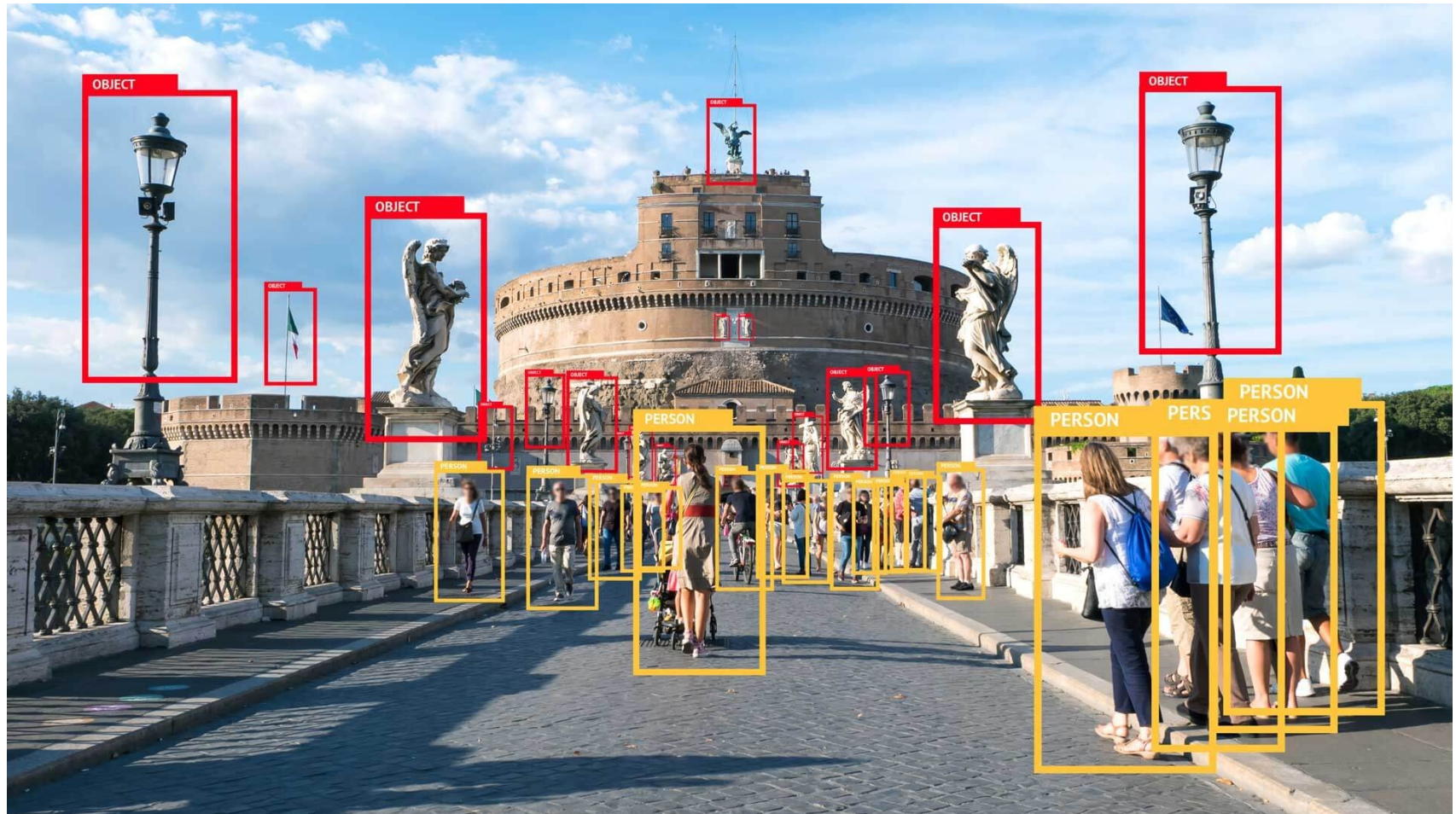
Recommender System



The collage illustrates three different recommendation systems:

- Netflix:** The top section shows a "Congratulations! Movies we think You will" recommendation. Below this, there are two rows of movie thumbnails with "Add" buttons and star ratings. The bottom section shows a "People You May Know" recommendation, with a red arrow pointing to the "Add as friend" button for Harry Tamacas Perla.
- YouTube:** The "Up Next" section shows a list of recommended videos, including "Behind The Scenes", "YouTube Rewind 2014: Behind the Scenes", "YouTube Rewind: What Does 2013 Say?", "Rewind YouTube Style 2012", "Girls' Generation 'I Got a Boy' Wins Video of the Year - Live at the YTMAs", "Mind the Gap: The Making of unReal Episode 1", and "BEST NEWS BLOOPERS 2014 Part 2". An "Autoplay" toggle is visible in the top right corner.
- Facebook:** The "Who to follow" section shows a list of recommended accounts, including "Sharpie", "Twitter Ads", and "Twitter Small Biz".

Image Recognition



Speech Recognition



Computer Games



Price Comparison Website

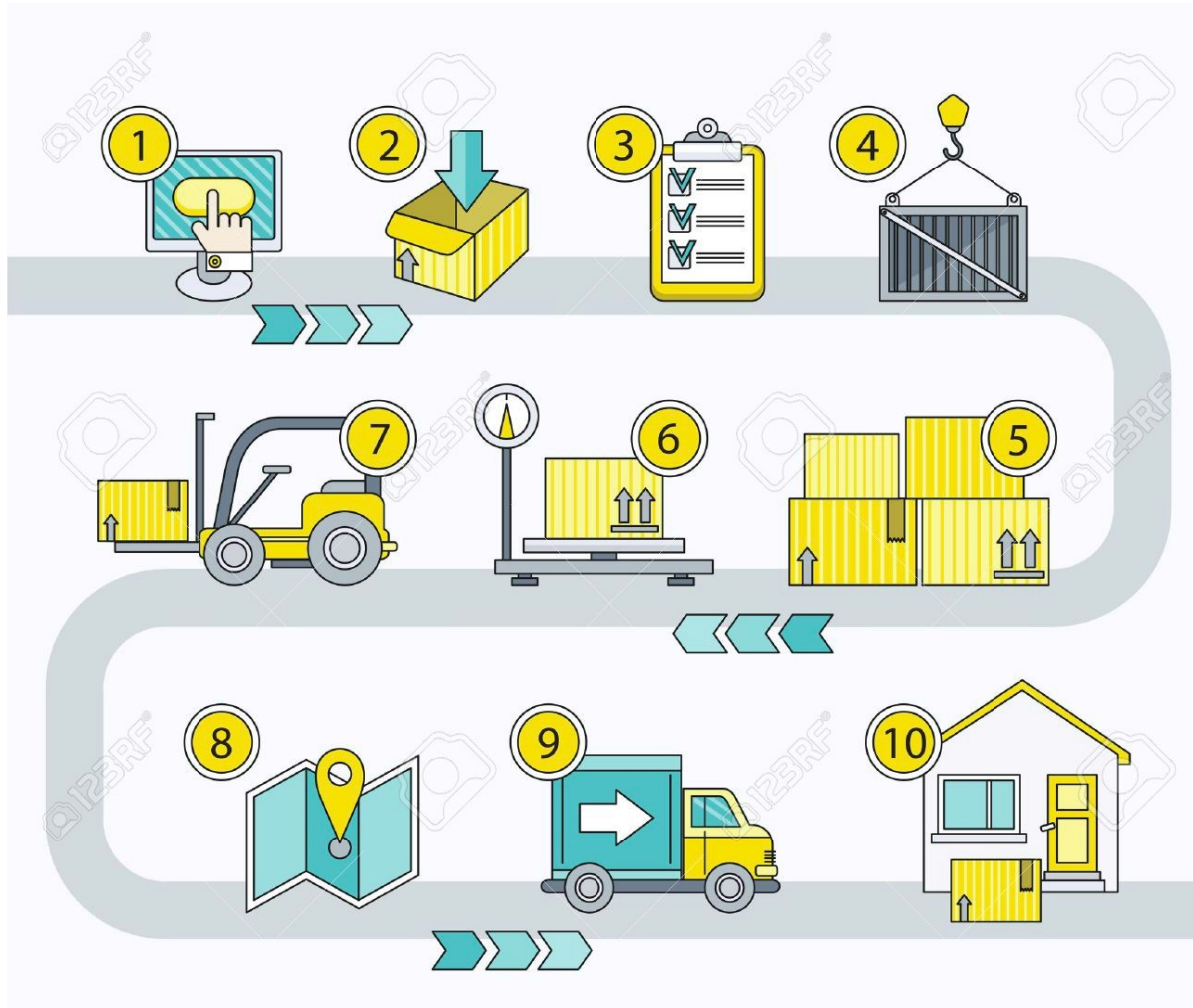




Fraud Detection



Delivery Logistics



What is Data Scientist ?

- Data Scientist
 - Most attractive Job of the 21st Century
- They find stories, extract knowledge from data and does the predictive analysis.



Data Analyst

- Data scientists are the key to realizing the opportunities presented by big data.
- They bring structure to it, find compelling patterns in it, and advise executives on the implications for products, processes, and decisions.

Working areas of data analyst

- National Security
- Cyber Security
- Business Analytics
- Engineering
- Healthcare
- And more

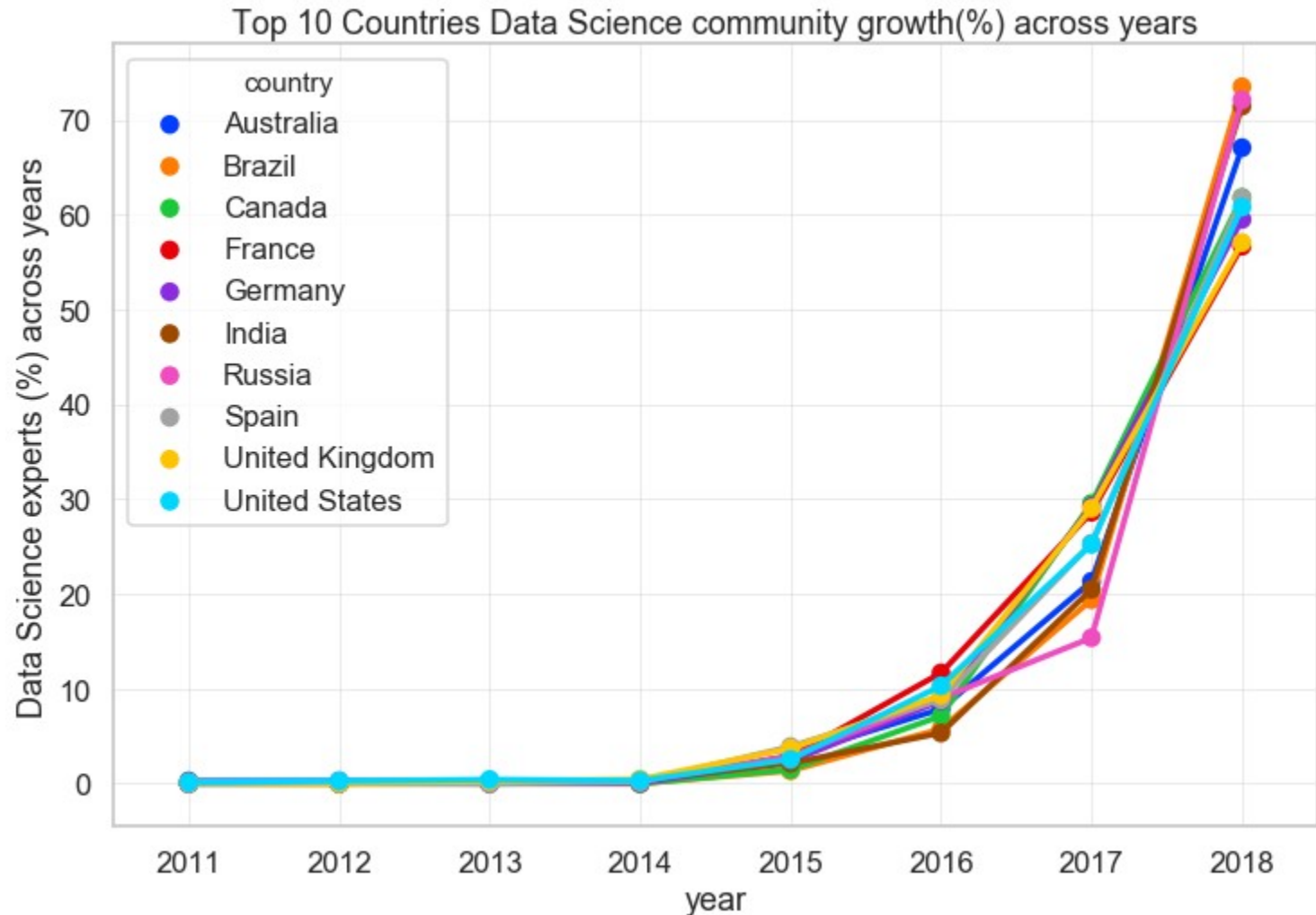
Data Analyst and Scientist Skill-set

Business	ML / Big Data	Math / OR	Programming	Statistics
Product Development	Unstructured Data	Optimization	Systems Administration	Visualization
Business	Structured Data	Math	Back End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian / Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

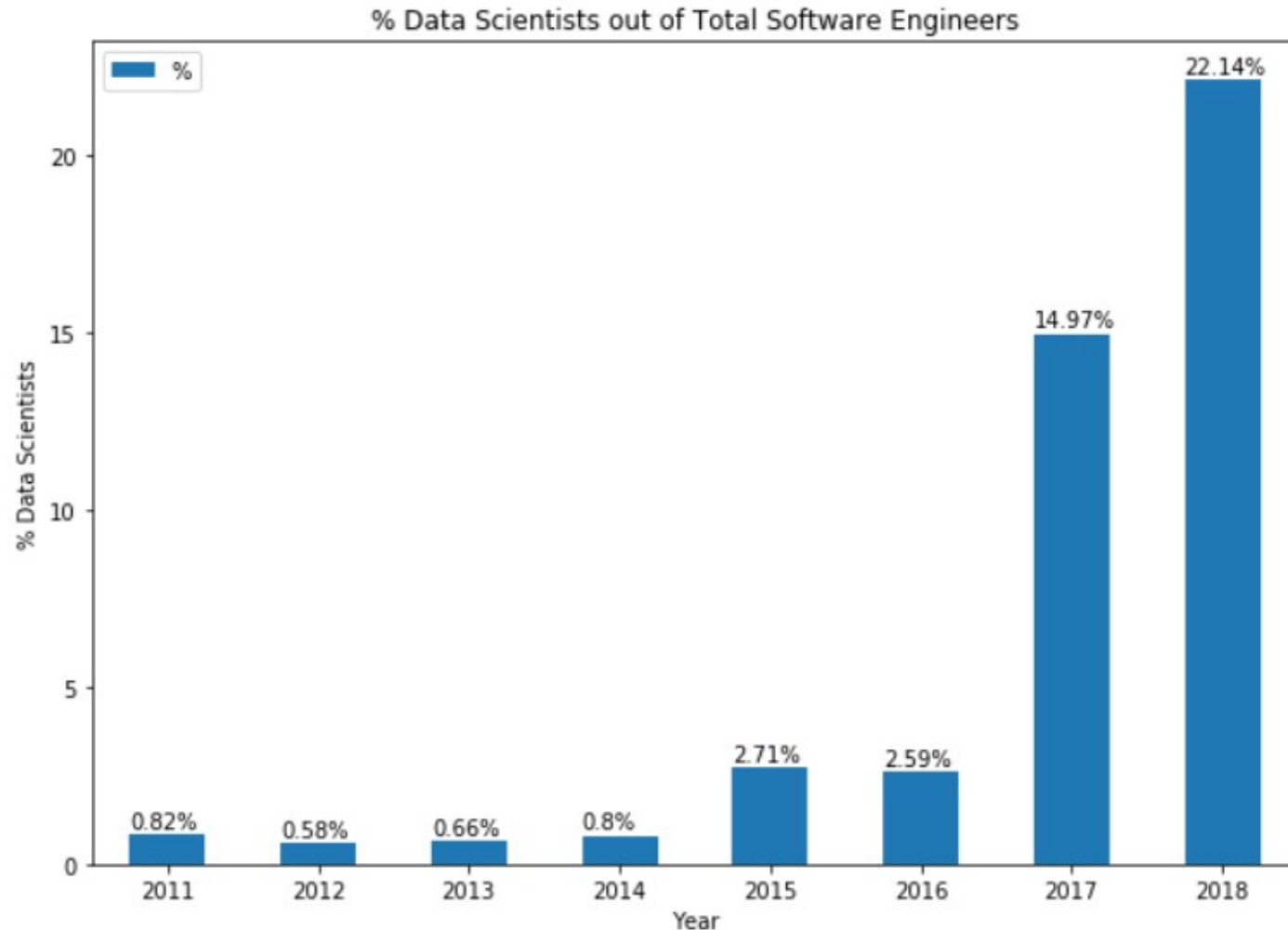
Data Science Jobs

JOB TITLES	
DATA SCIENTIST	DATA ANALYST
Data Scientist	Data Analyst
Senior Analyst	Business Analyst
Machine Learning Engineer	Data Architect
Research Data Scientist	Data Engineer
Director - Analytics	SQL Analyst
Big Data Scientist	Associate Data Center Analyst

Data Science Community Growth



Data Analyst and Scientist Growth



Data Science Today

WOMEN LEAD THE CHARGE AT ECONOMIC ORGANISATIONS

With the appointment of Gita Gopinath as chief economist at IMF, leadership positions in key global economic bodies are now held by women...

Gita Gopinath | CHIEF ECONOMIST, IMF

Christine Lagarde
MD, IMF

TIMES
THE TIMES OF INDIA

TCS doubles pay for fresh hires with new-age skills

Offers 1,000 Recruits ₹6.5L Instead Of Usual ₹3.5L Salary

Avik Das & Shilpa Phadnis | TNN

Bengaluru: Tata Consultancy Services (TCS) has offered about 1,000 freshers with new-age digital skills almost double the salary it normally pays those coming out of campuses. While the entry-level salary of Indian engineers in the IT industry has been stuck at about Rs 3.5 lakh per annum for the past decade, TCS is offering those with digital skills a starting salary of about Rs 6.5 lakh.

The selection of these candidates was based on their clearing a test focused on new digital areas. From this year, candidates who perform exceedingly well in the company's online National Qualifier Test (NQT), about which **TOI** reported recently, will also get an opportunity to take a shot at that digital skills-based examination.



TCS, one of the biggest recruiters from Indian engineering institutes, usually visits its accredited colleges to conduct a test followed by an interview. This process is going to be largely replaced by the NQT. "People who have done well in the NQT will get a chance to write another test for the digital talent pool, and if they clear and go through the interview, then they will get into the digital pool and their compensation will be differentiated," Ajay Mukherjee,

executive VP and head of global human resources, told **TOI** in an interaction.

The test involves programming with a higher degree of difficulty compared to the NQT. The test is longer and requires good coding skills. The move by TCS shows the lengths to which companies are ready to go to hire good talent. Employees armed with skills in the fields of machine learning, artificial intelligence (AI) and data analysis are getting better

appraisals across levels.

Such specialists are few and in much demand. Companies often have to spend significantly on training employees to acquire such skills. TCS's NQT, launched this year, has enabled it to reach out to a far larger student talent base, as also complete the recruitment process in three-four weeks compared to the three-four months it took under the traditional process.

Mukherjee would not comment on the number of people who would be hired this year as part of the digital pool, saying that the process is still on, but added that the numbers would be nearly the same or higher compared to the 1,000 it did in the last academic year. Apart from this process, TCS also does select hiring from the IITs and NITs, where it offers compensation packages that are even higher.

Data scientists now earn more than CAs

Those With 5-Yr Experience Make ₹75L +

Namrata Singh
@timesgroup.com

Mumbai: In 2012, Harvard Business Review named data scientist the "sexiest job of the 21st century". Glassdoor recently named it the "best job of the year" for 2016. In In-

INDIA TO FACE SHORTAGE

► So what's a data scientist? Generally speaking, practitioners are expected to know statistical analysis, predictive modeling and programming

► According to TeamLease, India will face a shortage

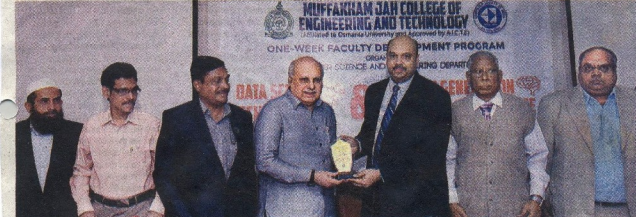


of 200,000 analytics professionals over the next 3 years

► Data scientists with 5 years' experience get over ₹ 75 lakh per annum as compared to ₹ 8-15 lakh for CAs and ₹ 5-8 lakh for engineers with the same work ex

Times of India - 08/05/2016 ... Read More

MUFFAKKHAM JAH COLLEGE OF ENGINEERING AND TECHNOLOGY
ONE-WEEK FACULTY DEVELOPMENT PROGRAM
ON DATA SCIENCE AND ARTIFICIAL INTELLIGENCE



'Data Science and Next Generation AI' event at MJCET

OUR BUREAU

Hyderabad: The Computer Science and Engineering department, MJCET is organising a one-week FDP on Data Science and Next Generation Artificial Intelligence from January 8-8 in collaboration with Computer Society of India (CSI), Chapter Hyderabad. The objective of the workshop is to give a comprehensive exposure of Data Science, Data Science Technologies, Next Generation Artificial Intelligence, its real time applications and thrust areas, current technology in practice through hands-on sessions.

Experts from industry (Tech Mahindra, Development Bank of Singapore, Deloitte, Lera Technologies, StratisBridge Advisors Pte Ltd (Singapore), and Yash Technologies) would be conducting the courseware of the FDP. About 100 plus participants including faculty, research scholars, PG students and industry persons from various different states of the country, to name a few - Andhra Pradesh, Chennai, Chattisgarh, Rajasthan - are attending the FDP.

The one-week FDP commenced with the inaugural function held on Thursday. Chief guest, Honorary Secretary, SUES Zafar Javed, Guest of Honour Firasat Khan, Associate Professor, and Graduate Programmes Director in the Management Information Systems Department at Metropolitan State University, Special Invitee K Mohan Raidu, Chairman, CSI Technological Council, Hyderabad, Advisor cum Director Basheer Ahmed, N Stharamaiah Incharge Principal MJCET, Head CSE A A Moiz Qywer and Coordinator of the FDP Dr. Mohammed Mahmood Ali graced the dais at the inauguration.

Welcoming the gathering, A A Moiz Qywer briefed the audience on the objective, relevance and scope of the FDP theme "Data Science Technologies and Next Generation Artificial Intelligence. Advisor cum Director Basheer Ahmed talked about the importance of the FDP and its theme and congratulated the CSE department for conducting the FDP. He shared his industrial experience of BHEL to explain the importance of Artificial Intelligence (AI) in Man-Machine Interface.

Special Invitee K Mohan Raidu threw light on the happening around regards Data Science, Artificial Intelligence and its mix from the industry perspective. Firasat Khan, guest of honor talked about the global perspective and present scenario of DS&AI. He emphasised on Data Science Technologies as the future of global business and employment opportunities for students in bulk.

The chief guest applauded the CSE department for coming up with FDP themes relevant to the current times both from industry and academia perspective and appreciated for the panel of experts from industry participating at the FDP. Lastly the inaugural concluded with vote of thanks, delivered by Mohammed Inran, Co-Coordinator of the FDP.

Data Science Today

Data scientist: the next hot job



Data scientists are responsible for discovering meaningful insights gleaned from data.

By Azuan Muda
azuanm@gmail.com/my

EMBARCING on a rewarding career with high demand in the information technology (IT) industry may be easy for those equipped with computer skills and knowledge in statistics.

The data scientist field is growing, providing graduates and professionals with a fulfilling career path. Tentspark founder Tang Siak Kwan's preference is to have a data scientist with communication and data modelling skills. "Communication is important in both understanding client requirements and having the ability to present data to clients. "Secondly, with data modelling and statistics knowledge, a data scientist should be able to produce predictive presentation to clients," he said.

The Malay Mail managed to catch up with Intel Malaysia data scientist Jenna Wan Jun. She spoke to us exclusively on what it is like to be a data scientist and how to become one.



Intel Malaysia data scientist Jenna Wan Jun.

Condensed Matter Lab.

She spoke to us exclusively on what it is like to be a data scientist and how to become one.

The Daily Life

Data scientists are responsible for discovering meaningful insights gleaned from data. Then, using all computational and inferential challenges, they provide automated predictions or recommendations with prescriptive capabilities.

They are also involved in the whole end-to-end process of cleaning and processing, managing and storing the data, exploring the data and communicating the results through visualisations and presentations.

The Entrance

There is no specific path in relation to academic qualifications. However, in her

opinion, a bachelor's degree in computer science at the minimum would be very helpful.

There is so much open-source material on data science online that there is almost no excuse to not pick up the necessary skills, if one really has the determination to learn.

A really good example of a self-made data scientist is Clare Corbelle, who gathered all the resources she could find online to create her own open source data science Master's programme.

She has a website called datasciencemasters.org which details a solid online syllabus for mastering data science.

"I would start out with some solid introductory data science courses in Coursera and then proceed to complete the machine learning offering by Stanford University Associate Professor of Computer Science Andrew Ng and the Learning from Data class by Professor Yaser Abu-Mostafa at Caltech," she said.

For paid options, the data science certification track MDeC offers on the Coursera platform is a great place to start.

Starting July 2015, several institutes of higher learning including Asia Pacific University, Malaysia Multimedia University, International Islamic University Malaysia, Sunway University, Monash University, University Institute Technology Mara and University Technology Petronas will offer undergraduate and post-graduate data science and computer science courses with data and business analytics specialisations.

Ideal Skills

• **Communication:** The value of data science depends on the ability of data scientists to communicate the outcomes

effectively to stakeholders.

• **Inquisitive:** Need to have the curiosity and thirst to keep learning.

• **Problem-solving:** Problems are constantly evolving and will require data scientists to employ new methods. There is a need to be creative in crafting out solutions and not fall into the trap of thinking in a linear way.

Core technical skills that would make a good data scientist is a combination of mathematical or statistical skills and computer science or programming skills.

Jenna emphasises the importance of a solid introduction to statistics, linear algebra and differential equations, which are the mathematical prerequisites for machine learning.

In addition, there should be some advanced programming courses, but the specifics could vary significantly.

Practical assignments and projects that reflect the complexities of real world problems and expose students to the large variety of approaches as well as the heuristic nature of data-driven problems provide valuable grounding to learners.

It would be helpful to gain experience in utilising big data tools and understanding or setting up big data server architectures, but the specific tools will keep evolving over time. Instead of getting hung up over specific tools, the focus should be to acquire the ability to innovate and adapt quickly.

Job Prospects

Multimedia Development Corporation (MDeC) targets producing 1,500 data scientists by 2020.

Professionals in the field may earn a median pay of RM37,547 a year (payscale.com).

Rajlakshmi Ghosh
@timesgroup.com

In a world where there is a need to pursue programmes, which equip students with skills that employers demand, the sharing of data resources, collaborative activities, and a culture of drawing from each other's work has become all too important. "An education system that does not equip social science students with tech-driven applications is greatly limiting their career prospects," says James Abbey, associate academic director of the University of London programmes at the London School of Economics (LSE). He explains, "Take the case of a Geography graduate for instance; there is a tremendous demand for Geographic Information Systems (GIS) analysts who are required to routinely deal with large datasets to process and analyse into user-focused displays such as graphics, maps and charts. This work requires extensive knowledge and experience



with GIS techniques, technology, and principles, as well as strong proficiency with computers, including R, Python, HTML and Microsoft Office software."

"Today, an Economics graduate will be highly valued as a compliance analyst, with responsibility for ensuring that compliance has been achieved internally as well as externally for a company. Similarly, a graduate who is aiming for public policy can conduct quantitative research using advanced computational and data science techniques such as machine learning (ML)," says Abbey.

While a social science degree can get one started on a wide array of careers in economics, management, social work, law, academia and po-

licymaking, the key here is to have had the relevant exposure to tech applications and data analysis tools needed for the job. "Superior career prospects, job retention, promotion prospects, as well as a marked higher pay bracket are a clear advantage for people with the right technical skills," adds Abbey.

DATA SCIENCE IN DECISION-MAKING

One of the key challenges for decision makers and managers is to understand what makes for good data science, and how the evidence from this field should be used in evaluation and decision-making. "Effective use of data and machi-

ne learning tools is critical in making adaptive and personalised policies that improve the standard of living and paves way for the development of society. If integrated well into the policy making process and understood well by policymakers, data analysis has tremendous potential to lead to better decisions. Policy-makers in state departments, for instance, can avail data science to tackle social issues such as traffic, road safety and crime," says Abbey.

"Specific data science tools like R and Python, along with Tableau, will do wonders for various tracks of Social Science. Then there is Deep Learning, a part of the ML family. At a 'deeper' level, it helps to understand hierarchies, and patterns that can help a system to learn complex functions mapping the input to the output directly from data, without depending entirely on human-made features," says Abbey.

He feels social scientists must shed their inertia and train themselves in data science techniques because these are, ultimately, languages and can be acquired with a finite amount of effort. More importantly, the data science techniques can help take the investigation of social science questions much deeper; hence facilitate better and smoother social decision-making.



Indian focus on data science



"The Indian education system has traditionally been slow in keeping up with changing industry trends. Thus, the direct integration of data science within social science curriculum is still nascent. Some institutes have taken steps such as the School of Management and Labour Studies of the Tata Institute of Social Sciences (TISS) that offers an Executive PG Diploma in Analytics (EPGDA), and Indian School of Business & Finance (ISBF), New Delhi that offers a PG Diploma Programme in Data Science. Academic content for the latter comes from the London School of Economics (LSE)," says Yavar Ehsan, associate professor of Management and Information Systems, ISBF.

themselves in data science techniques because these are, ultimately, languages and can be acquired with a finite amount of effort.

More importantly, the data science techniques can help take the investigation of social science questions much deeper; hence facilitate better and smoother social decision-making.

Data Science Today

SATURDAY, APRIL 18, 2015 THE AGE

MY CAREER 27

Feature

Is data scientist sexiest job of the century?

This scientific work is a fast-evolving profession, **Karla Dondio** writes.

A few years ago, *The Harvard Business Review* hailed the burgeoning role of data scientist "The sexiest job of the 21st century". With big data technology driving the change, how does the new role stack up?

"It's probably fair to say it's a lot sexier than any of its earlier incarnations," laughs Peter Davenport, who laments the title of "statistician".

But if job satisfaction is a barometer for sexiness then Davenport, a data scientist at Sensis, is hard-pressed to find any drawbacks to his role.

Sensis is a marketing services company that includes well-known brands such as Yellow Pages and White Pages.

Dubbed the "oracle" by work colleagues, Davenport assists the company's sales and marketing teams to understand the behaviour of businesses to develop strategies to then target those businesses.

"It helps the teams to sell but it

also provides value back to those businesses," says Davenport.

Big data is distinguished by the (3Vs) volume, velocity and variety of information that can now be captured and used by businesses to deliver a unique customer experience. Subsequently, many IT specialists are switching their job titles to reflect how big data technology has changed the nature of their work.

"The data scientist term resonates because you start to use lots of data: data from open sources and data coming from new technologies," Davenport says. "It feels like a pure science. It's the breadth that's amazing." The breadth is exactly why data scientists' tools are integral to their work. Davenport uses R, an open-source program for data mining with SPSS modeller, and MapInfo for geographic analysis. Other popular data software includes GenStat and Hadoop.

Davenport believes that the data scientists who excel in their field are lateral thinkers who can make



Crunch time: Peter Davenport loves his job, but don't call him a statistician.

sense of the mass of data available today. The ability to communicate complex ideas in a concise and logical manner is a highly valued skill that Davenport brings to Sensis.

Echoing the sentiment, Dr

Nicola Powell, founder of and data scientist at DataGen, says that communication is the key to giving customers value.

"You might be a great data scientist in terms of being able to identify trends, but if you can't

communicate that to your clients they're not going to be able to answer their business questions," Powell says. "The other great skill set is curiosity. I've always asked questions and I continue to ask questions."

Powell started her career working in agriculture as a research scientist, but says she did not see lab work as a long-term proposition. Her love of data, research and science combined led her to starting her business a few years ago.

A typical day for Powell might see her spending long hours in front of her computer trying to decipher large amounts of numbers and text. Nonetheless, the work is varied, covering everything from customer behaviour to campaign effectiveness to social media analytics to PR. Powell says her clients' end goals are always at the forefront of her work.

"It's all about interpreting results to every level of business," says Powell.

So is data scientist the sexiest job in the 21st century?

"Of course!" Powell says gleefully. "It's definitely a profession to watch out for."

Decoding Data Science



Towards New Horizon

Data Science Foundation (DSF), a non-profit entity engaged with the advancement of Data Science (analytics and big data) organized the first annual summit, titled "Data Science Summit in association with Economic Times on August 11 at The Park, Kolkata. Sponsored by the department of Information Technology & Electronics, Government of West Bengal, NASSCOM, the summit was inaugurated by chief guest Birendra Basu, minister-in-charge, department of IT & Electronics, Govt of West Bengal.

Along with the additional chief secretary, Govt of West Bengal, Mihail Gusev, deputy consul general, Russian Federation, Gautam Banerjee, president, Data Science Foundation and Haspam Choudhury, president, NASSCOM, graced the event among others. At the event, Ambardip Dasgupta, former national consulting head of KPMG, was felicitated by Ananta Choudhury, executive director, REERIS and also presented the award to the awardee for his contribution and positively building the ecosystem.



of Data Science
The summit
form for both
to deliberate
in the new
research and
with the latest
ing, IoT, deep
the panel dis-
data driven
session starts
applications
domains of
Nirupam Cha
Highlight

Science, Biju said that the state will soon launch National Institute of Data Science. Sen, on the other hand pointed that

benefits through data science techniques." Anshuman Paul, VP & executive director, First American, said, "The company is try-

Several other innovation and hurdles that the young organisations are facing in regards to Data Science and the way they are dealing with it were discussed by eminent panelists that included Somnath Banerjee, director & CEO of Business Billion, Shantanu Sorn, CEO, Somnatics, Sanjoy Chatterjee, CTO, Ideal Analytics and Sanjib Bose, director of Probita. After the panel discussion, Probita Mukhopadhyay, CA

त्यापार

द्वितीय मसौदा, नाशिक, रविवार, १५ सप्टेंबर २०१९

divyamarathi.com

जॉब स्प्रीक इंडेक्सनुसार आयटीतील नियुक्त्यांमध्ये ३८% वाढ, नॅसकॉमनुसार बिग आर्टिफिशियल इंटेलिजन्समध्ये १.४० लाख पदे रिवत
आयटी उद्योगात बिग डेटा अ‍ॅनॅलिटिक्स आणि मशीन लर्निंगला सर्वात २०१८-१९ मध्ये देशातील इन्फर्मेशन
जास्त मागणी. समस्यांची उकल करण्यासाठी उत्तम कौशल्याची गरज टेक्नॉलॉजी उद्योग १२.८५ लाख
कोटी रुपयांवर जाणार

२०१८-१९ मध्ये देशातील इन्फॉर्मेशन
टेक्नॉलॉजी उद्योग १२.८५ लाख
कोटी रुपयांवर जाणार

एचआर-interview



आष्टी उदात्तागत भात जगातील महारक्ती
मनुज ओढवणेत जसो, या क्षेत्रात
टॅटोची गुलकत व उद्योगाची परज यात
मोठा फरक आहे. विशेष निमित्त हे १५६
खंड कोटी रुपयव्यय काजत भांडवलावर
हिसरी सवती मोठी आष्टी कंपनी आहे.
हा परक कमी करपायझीती कंपनी उपाय
करते. नेहमी करीअरच्या कृतीने
हिल मारतीने विशेषे मुळा मनुजवळ
अधिकारी सौरभ गोविंद वाघार्याची चर्चा
केली.

• तप्यत्वा स्थितं युक्तान्ताली अक्ती क्षेत्रात कोणत्या प्रकारचे रोजमार उपलब्ध आहेत?

समुप्रांत अर्थो क्षेत्राच्या विविध भागांत डिजिटल स्किल डेव्हलपिंग मोटो मागणी आहे. कन्सल्टिंग, डिजिटल, ऑनलाईन वॉरल्ड इंटरनेट (कृत्रिम बुद्धिमत्ता) अशी काय्युअल रिझॉल्यूशनसहच्या विभागांमधोल भरतील येणारे पंडा होत आहे. बिग डेटा अनीलिटिक्स (डेटा विशलेषक), जाका फुल स्केक, मोबाईल, क्लाउड कॉम्प्युटिंग आणि डेव्हॉयससहच्या स्किलसचीदेखील मोटो मागणी आहे.

• विप्रोमये नोक्तीसहटी कोणत्या प्रकारची पात्रता आणि कौशल्याची गरज असते ?

असा कोणताही प्रेक्षारंज्याचे ऑनलाईन कल मांडू असेल, कम्युनिकेशन स्कूल चांगले असेल आणि शिक्षकाची इच्छा असेल, प्रशिक्षणांतर प्रवेश पाठ्यवीर विद्यार्थ्यांचे येऊ शकतात. प्रेक्षारंज्याची आपली प्रोव्हिडिंग असेसमेंट सुरू केले आहे.

• आवटी क्षेत्रात अजिंक्य परिसरितीत फेरर्स आणि अनुभवी व्यावसायिकंसाठी शिक्ती रांदी आहेत ?
आवटी कंपनी मोठ्या प्रमाणावर भरती करत

देशातल्या आयटी क्षेत्रातील नोकऱ्यांत 3८% वाढ झाल्याची नोंद

- जर्ब सन्धे होइकसुमार अर्धौ खोवाथाल
- मिर्कसुवाथाल ३८% खाइ होइत अछि
- मासुवर मिर्कसुवाथाल, आगजोटी अणि विंग डेटामे २०२० परित १४ लाख नैक-नैक
- मिर्कसुवाथाल जर्ब नैकसुमाथे २४% खाइ होइत अछि
- जातिक अर्धौ व्यवसाया २०१०-१८ मने भारतल कर्मनैकषा ५५% बढा भ्वाइ १३.१३-१३.७ लाख कोटी रु. होला
- २०१८-१९ मने दोहातल अर्धौ, आगजोटीसुमार उडान १८.५ लाख रु. झाल
- नैकसुमा अडवाकसुमार अडिफिकेशनल इटिलिजिजसुमार १.३० लाख परि रिक्त अछि
- दोहातल कसकड कर्मनैकषा बजारोइ २०२०परित अडवा ५०.०० कोटी रुपयावर जालक, कस्यो तो १०,२० कोटी रुपयाचि ओ

आज, तंत्रज्ञानाचे वेगवेगळे परलोल होत आहे. युवक ऑनलाइन क्लिपस तयार करून त्यांचा स्वतःचा नोंदीबरोबरचा वाटा वाचू शकतात. अनेकवेळा व्यावसायिकांची समस्या जटिल आहे. विविध कलत्रात, साधने, समसूची, डिजिटल ऑनलाईनस आणि ऑनलाईन आणि ऑनलाईनसाठी तयार झालेले तंत्रज्ञानाचे वेगवेगळे प्रकारही आहेत.

टॅक्निकल इंटरव्ह्यूचे आयत्या काळाचा अनुभव घेऊन कायदा आणि उमेदवाराकडून अपेक्षा असणे प्रेक्षासमोरील त्यांचे विवेचन आणि प्रीप्रिपिशन समायोजन आहे. विवेचनसह असलेल्या उमेदवारांनी उमेदवाराच्याच चांगले कामगिरीतून क्लिपस पोस्टिंगसह अडिटरव्हू आणि शिकवण्याचे प्रत्येक वेळीही जाणे.

• इयुक्तांनी मुलासाठीही करी करावी ?
• विद्यार्थ्यांच्या आयुष्यात दोन प्रकारचे इंटरव्यू घेणे.
• विल्लू देवनाथल आणि दुसरा किराविल्लू.
• विद्यार्थ्यांच्या संपर्कातून जस्त नेकाचा दिवस
आहेत. तुम्हाला कधीही नेकाची भिल्लवण्यासाठी काय
पद्धत आहे?

या क्षेत्रात आहेत रोजगाराच्या संधी

- [illegible]

विद्येने वसुधासमने टॅलेटला ऑनकेशन/सॉफ्टसेस डेव्हलपमेंट या त्यांच्या सात्यासभूत क्षेत्रात नोकऱ्या दिल्या आहेत. विद्येने आरवल्या वेबसाइटवर करार पेज टांगले आहे, ज्या पदासाठी नियुक्ती करायची त्याची माहिती येथे दिली जाते. इच्छुक उमेदवार येथे जाईसाठी अर्ज करू शकतात. सर्वोत्तम टॅलेट ह्यावर करण्यासाठी आम्ही एम्पलीमेंट रेफरल सूचक भर देतो.

• **विप्रो उद्योजकतेला प्रोत्साहन देते का?**
विप्रोमध्ये आम्ही उद्योजकतेला प्रोत्साहन देतो.

आम्ही इन्व्हेलन्ससाठी एक फंड तयार करत आहो. दरवर्षी नवीन बिझिनेस आयडिया घेऊन येणाऱ्या टीमला कंपनी दरवर्षी या फंडातून रक्कम देते.

• **अजिंक्ये मती प्रथित वदत इत्यतः का?**
डिजिटल टेक्नालोजीचा भारती प्रतियोग्य खूप परिणाम झाला आहे. उमेदवार कंपनी जवळपास नवीन आपल्या वेबसाईटला आहत हे काय शकता. डिजिटल एडवर्टाईसिंगा माध्यमातून मोठ्या प्रमाणावर उमेदवारांची भारती केली जाते. फ्रेशर्सच्या भारतीसाठी आम्ही द्या नॅशनल टॅलेंट हंट कायदेशर लॉच केली आहे. यामध्ये क्षेत्र, कॉलेज व ईजिनिअरिंग स्ट्रुमला वेगळे ठेवून प्रतिभाका उमेदवारांना सोयी मिळते.

• भविष्यत अवर्ती प्रयोगत नोकरीच्या काय संघी
आवतील? वृद्धांनी काय बदल घडवावा ?

सध्या आषाढी उद्योगात हजारो लोक काम करत आहेत. भविष्यात कंपनीकडून प्रकाशचे विकल सेट निवडतो त्यावर हार्वर्स पॅटन अवलंबून असेल. काळाबरोबर तंत्रज्ञान बदलत आहे. त्यामुळे आषाढी कंपनीनां किर्लुष्क असलेले आणि अनेक स्टोक होल्डर्सबरोबर काम करू शकण्याच्या लोकांची गरज भासेल.

Prime Disciplines of Data Analysis

- Machine Learning
- Artificial Intelligence
- Deep Learning
- Natural Language Processing

Machine Learning

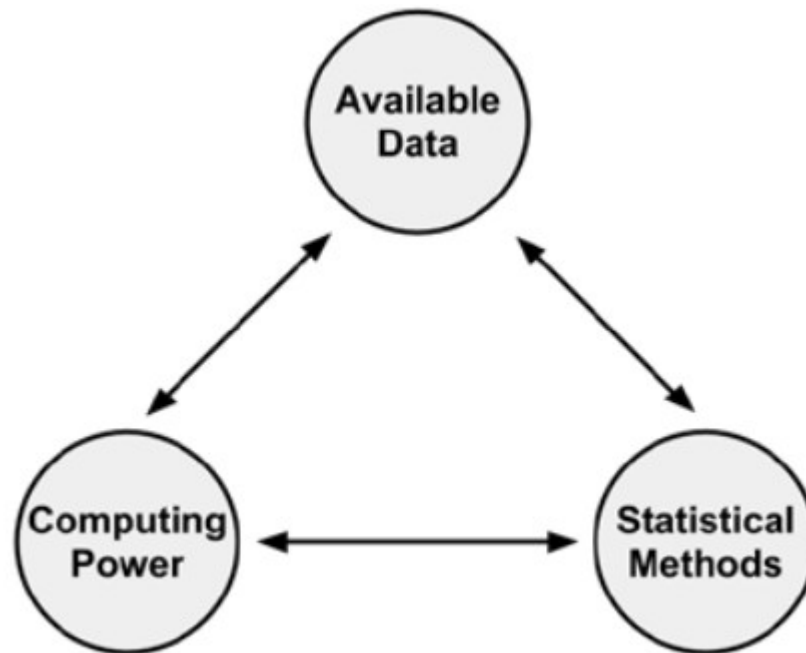
- Machine learning is an application of **artificial intelligence** (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.
- The process of learning begins with **observations** or data, such as examples, **direct experience**, or **instruction**, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.
- The primary aim is to allow the computers learn automatically **without** human intervention or assistance and adjust actions accordingly.

Origins of Machine Learning

- The earliest databases recorded information from the observable environment.
- Astronomers recorded patterns of planets and stars; biologists noted results from experiments crossbreeding plants and animals; and cities recorded tax payments, disease outbreaks, and populations. Each of these required a human being to first observe and second, record the observation.
- Today, such observations are increasingly automated and recorded systematically in ever-growing computerized databases.

Machine Learning

- The field of study interested in the development of computer algorithms for transforming data into intelligent action is known as machine learning.



Data Mining

- A closely related sibling of machine learning, data mining, is concerned with the generation of novel insight from large databases (not to be confused with the pejorative term "data mining," describing the practice of cherry-picking data to support a theory).
- Although there is some disagreement over how widely the two fields overlap, a potential point of distinction is that machine learning tends to be focused on performing a known task, whereas data mining is about the search for hidden nuggets of information.

Useful web resources

- www.mitu.co.in
- www.pythonprogramminglanguage.com
- www.scikit-learn.org
- www.towardsdatascience.com
- www.medium.com
- www.analyticsvidhya.com
- www.kaggle.com
- www.stephacking.com
- www.github.com

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License



@mitu_skillologies



/MITuSkillologies



@mitu_group



/company/mitu-
skillologies



MITUSkillologies

Web Resources

<https://mitu.co.in>

<http://tusharkute.com>

contact@mitu.co.in

tushar@tusharkute.com