

**Data Science  
Transforming the ICT domain**

Feb 2018



# Contents

Foreword	4
Message from Aegis School of Business	5
Data Science in ICT	6
Benefits Realized through Data Science	11
Pointers for ICT companies	15
The Future	17
Concluding Remarks	20
About Aegis	21
Acknowledgments	22
Contacts	22
References	22

# Foreword

Information and Communication Technology (ICT) as a sector has seen a phenomenal growth in the past decade. These companies have access to a great amount of data generated by digital footprints left behind by customers which is increasing exponentially. ICT as a sector has gone an overhaul due to this boom of data generation. Data storage capacity of devices across the globe to store and receive data over network and compute has increased more than ten-folds in a very short duration. All this has driven adoption of new ICT technologies like Cloud Computing, Internet of Things, etc. This certainly demonstrates the emergence of Data Science to convert data into business insights.

The credit for optimization achieved by these technologies in managing data using distributed storage and computation capabilities goes to advancement in Data Science and its close relationship to developments in the ICT Sector. The collaboration between industry and academics has been a driving force behind the exponential rise in the usage of ICT devices. Both these technologies, Data Science and ICT, have formed an interdependence which helps both these domains to aid development.

Data Science has become crucial to understand the external and internal forces impacting the business through data generated from social media, search engines, government portal and can be leveraged in widespread business

applications. The rapid development in the domain of Internet of Things and mobile devices has helped us realise once highly ambitious projects of Autonomous Cars and Smart Homes. The demand for people to work in this interdisciplinary domain of Data Science and ICT is an upward trend, far outgrowing the supply, which defines the need for skilled data scientists with ample knowledge about emerging information technologies and the ability to implement business solutions effectively.

Having set the complexities of data science, application of these technologies has become simple and one does not require high end development skill in order to adopt data-science skills. With the help of easily available libraries and user-friendly tools, business users can develop and adopt data science technologies. Easy access to web-hosting, cloud computing and AI tools has enabled business users to focus more on business logic and achieve quicker business insights and outcome.



**Hemant Joshi**

# Message from Aegis School of Business

Taking decisions based on Data not only makes instinctive sense, but it makes strong commercial sense too! Over the past few years, there's been a lot of media hype about data science, Big Data, Machine learning, deep learning etc. And every organization is trying to transform itself into a data driven organization, however, struggling how to do it, not able to fully understand what these fields are. This white paper is an attempt to clear some air around this newly defined field called Data Science. In our practical experience at "Data Science Delivered" at Aegis School of Data Science dealing with various clients in solving problems, few things we have learned that Data science is nothing but an approach of solving problems using data; making sense out of data and automating the process of decision making among many.

For solving the churn problem of leading telecom operators, we had to combine massive data coming from multiple sources like CDR, CRM and billing to predict the churn, this was a big data and large scale machine learning problem. Traditional data analysts and business analysts were not skilled enough to make sense of big data coming from different sources and in different forms: structured, unstructured, text, image, video, machine data, ERP, CRM, email, social media, blogs IOT devices etc. and perhaps that's what led Dr. DJ Patil to coin Data Scientist term in 2008 with Jeff Hammerbacher and crown themselves

with Data Scientist title to define their complex data jobs at LinkedIn and Facebook, respectively.

Per Gartner's report, "AI and machine learning will increasingly augment and extend virtually every technology enabled services, thing or application."

Creating intelligent systems that learn, adapt and potentially act autonomously rather than simply execute predefined instructions is the primary battleground for technology vendors through at least 2020" which offers tremendous opportunities and challenges for the ICT sector. The application of AI, ML, Deep learning, Data Science is growing in India and around the world in almost every industry, telecom, IT, Insurance, manufacturing, healthcare, banking, retail, media, consulting, e-commerce, oil & gas, automobile, airline, Govt, NGOs and startups and every functional area. With our own experience at Aegis School of Data Science of addressing to over 35,000 professionals across the country in last three years, we have realized there is a huge appetite to learn data science primarily driven by high salaries, demand, fear of losing jobs, and off course the intellectual challenges this field offers.

The demand for data scientists with depth of knowledge and applied skills in various areas like math, stats, AI coding, AI, ML, Deep Learning, NLP, Big Data etc.

which is growing at rapid pace however Indian academia and industry is not able to fulfill the need.

As India has the largest pool of people with math and coding skills, we can become the world's largest hub of skills for Data Science, ML, AI and Deep learning. To accomplish this mission Aegis and IBM have created one of the best programs in Data Science, Business Analytics and Big Data. And now NVIDIA one of the best technology providers for Deep Learning has joined hands in this journey.

Hope this white paper helps unravel the complex world of data science.



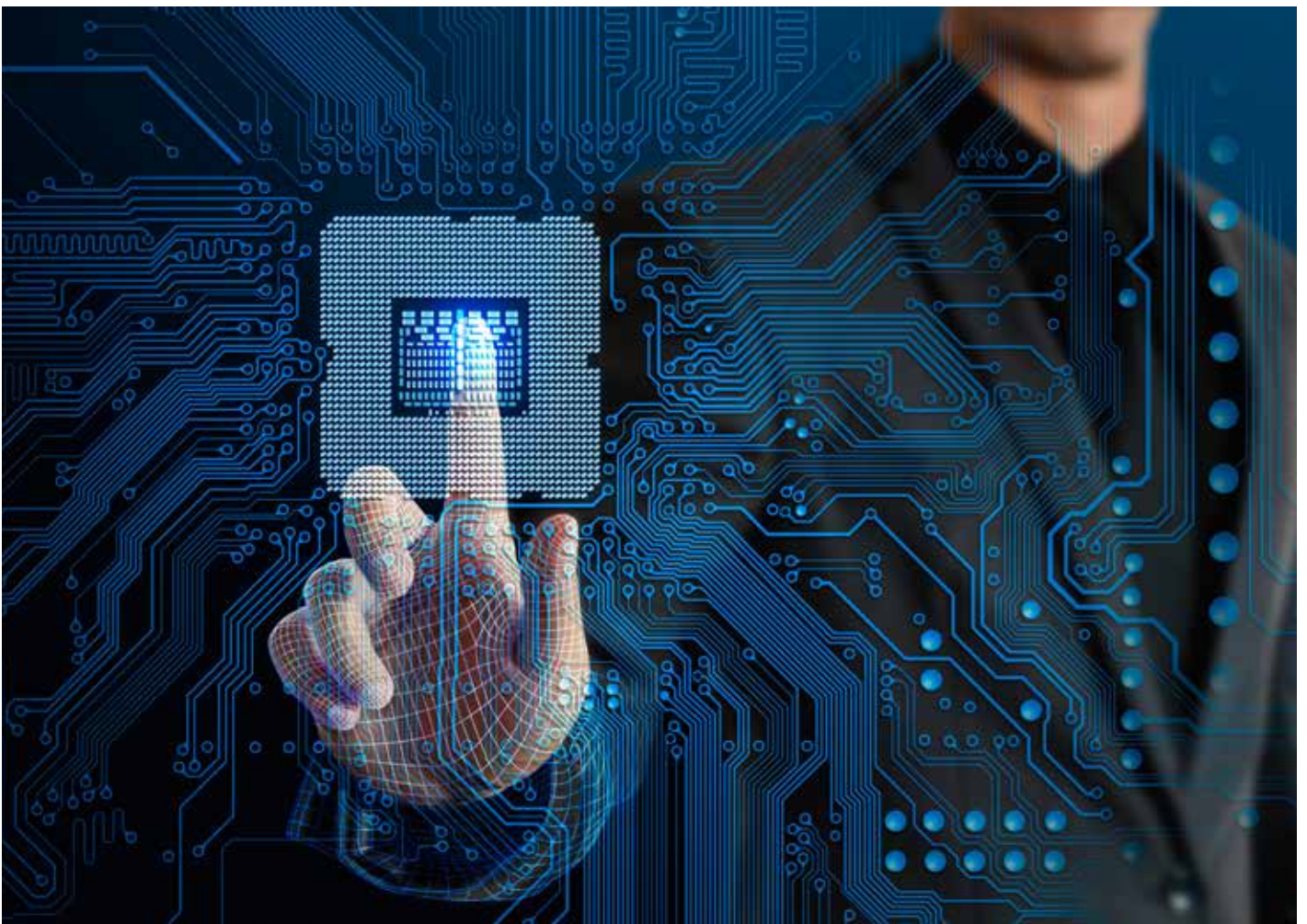
**Bhupesh Daheria**  
 Founder CEO, Aegis School of Data Science  
 Partner, Data Science Delivered  
 Founder, Data Science Congress  
 Founder, Aegis Graham Bell Awards  
 Founder, mUni

# Data Science in ICT

Data-driven Science, or Data Science, as it is popularly known, is an interdisciplinary field of scientific methods, processes, and systems to extract knowledge (insights) from data in various forms, either structured or unstructured.

Data science is "a concept to unify statistics, data analysis and their related methods" in order to "understand and

analyse actual phenomena" with data. It employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, information science, and computer science. In particular the subdomains of machine learning, classification, cluster analysis, data mining, databases, and visualization are used extensively in realizing the potential for data science.



### Technology Pillars

**Statistics:** Many critical academics and scholars believe that data science is just a fancy term for statistics, and that statistics is at the very core of data science study. All these critiques are true to say that statistics forms a very important pillar in the field of data science, providing the very basic as well as complex metrics to solve and evaluate the analysed set of data. Some commonly and everyday used metrics include type of data (Continuous vs Discrete), statistical distributions (Poisson, Binomial, etc.), Probability and Cumulative density functions and accuracy analysis using ROC curves. Some statistical theorems form the base for various data science algorithms like Bayes Theorem, KNN Algorithm and Bagging/Bootstrap aggregating. Thus, we can say that knowledge of statistics is very important from the point of view for real world application of data science.

### High Level Programming Language:

Python and R, are two prominent open source, high-level versatile programming languages used for almost all purposes over the breadth of the industry. The open source community is very robust which has given rise to a variety of inbuilt packages which makes these languages very easy-to-use and an important tool for data science. The growing abundance of libraries and packages to

realise data science goals and support for the same has made Python and R, both – popular languages in the market. Emergence of open source licensing platforms like GitHub has helped to grow the number of contributors which has led to development of various packages and API Frameworks for ease of use by the academia as well the industry to implement Data Science Solutions in their businesses.

**Domain Knowledge:** Domain knowledge is used all the time in Data science applications (sometimes without knowing that you are doing it). A good example is feature extraction, how do you know that these features are important for the model which you are building? For e.g. Internet Speed may be good features in knowing about how much time a customer spends on online content streaming. Domain Knowledge helps to identify which data set make sense to business and how they are going to be consumed. All that is needed to build a data science model is a dataset which consists of examples or records in rows and attributes or features in columns. A Data Science model needs a response feature which is what would be predicted when the other attributes are known only with the help of domain knowledge.

“The exponential growth in the adoption of digital technologies by new age and traditional organizations presents numerous opportunities for people looking to build a career in this space. Notwithstanding concerns that automation, artificial intelligence and machine learning will result in job loss, there are whole new areas of work where we are witnessing an acute shortage of skilled people.

The edge of innovation is an ever expanding line. Acquiring analytics and data science skills across the spectrum of business, mathematics and technology will enable the workforce of today to stay relevant, capitalize on the opportunities and ride the growth wave.”

**Rajan Sethuraman,**  
Chief People Officer at Latent View Analytics

### Evolving Ecosystem

In the Indian telecom Industry, players such as Reliance Jio, Vodafone and Airtel have actively implemented data science to the large amount of data they gathered from their customer base to improve their services and network. Also, the insights lent from this data has enabled companies to launch specific plans and offerings based on the location and age of users. The entry of Reliance Jio in the telecom sector and the phenomenal market capture strategy by offering 4G services at minimal cost is all due to the study and analysis of the demands of the consumers. This competition has forced other market players to match offers from Jio, leading to a wider user-base

connecting to the internet and further contributing more data to the domain for future development. The active use of consumer data and their digital footprint for analysis can certainly help the companies to develop their services and products in and around users for optimal market penetration and maximum returns for their businesses.

How can organizations use existing data set which is already available with them? How has technology changed over the years which makes it much easier and simpler to work on large data set without making huge investments in terms of technology?<sup>1</sup>



## 01

### Cloud Computing



Simply put, cloud computing is delivery of computing services—servers, storage, databases, networking, software, analytics and more—over the Internet, where users need only to pay for what they use. The service became popular after Amazon introduced its Elastic Compute Cloud in 2006. In cloud computing, all the resources are allocated on demand and distributed among multiple users. The goal is to maximize computing speed and increasing efficiency. This allows small normal users and smaller companies to rent large-scale units for better computation and storage as and when required on the go, supported by the emergence of platform-as-a-service (PaaS) for Data Science that helps small companies deploy and administer their clusters with reduced prices and low efforts.

## 02

### Internet of Things (IoT)



The Internet of Things (IoT)—the practice of capturing, analysing, and acting on data generated by networked objects and machines—is among the hottest technology topics in business today. While a growing number of companies are creating business value with IoT applications, the technology is still in its early days. Two trends will dramatically expand IoT possibilities in the enterprise, multiplying practical applications while potentially lowering costs. The emergence of new wireless communication networks designed specifically for IoT applications, which can lower the cost and extend the reach of connected applications. The arrival of “edge computing” IT infrastructure, which facilitates analysing and acting on IoT sensor data close to the source, making applications more responsive to rapidly changing local conditions while avoiding communications bottlenecks.





03

### Artificial Intelligence



Artificial Intelligence (AI) is concerned with designing intelligent systems that exhibit characteristics associated with human intelligence. Areas stemming from AI include neural networks, time series prediction, classification, evolutionary computation, genetic programming, vision, robotics, expert systems, speech processing, planning, and natural language processing. Majority of the AI methods stated above are used extensively in performing analytics on the enormous amount of data in this rapidly changing world. Applying AI to Data analytics is helping companies make relevant sense of data, detect correlation between factors and disruptions better, deal smartly with the lightning speed at which information is being generated, and gain insights from the data they have.

04

### Networked systems



A large number of ICT technologies (e.g. IoT, cloud services, and media distributions) that are involved in Data Science are basically part of complex networked systems. For Big Data enterprises that experience ever increasing workload, both in scale and complexity, to ensure that network traffic issues are solved and insights are delivered regularly by completing the workload; networked systems will play a very important role.

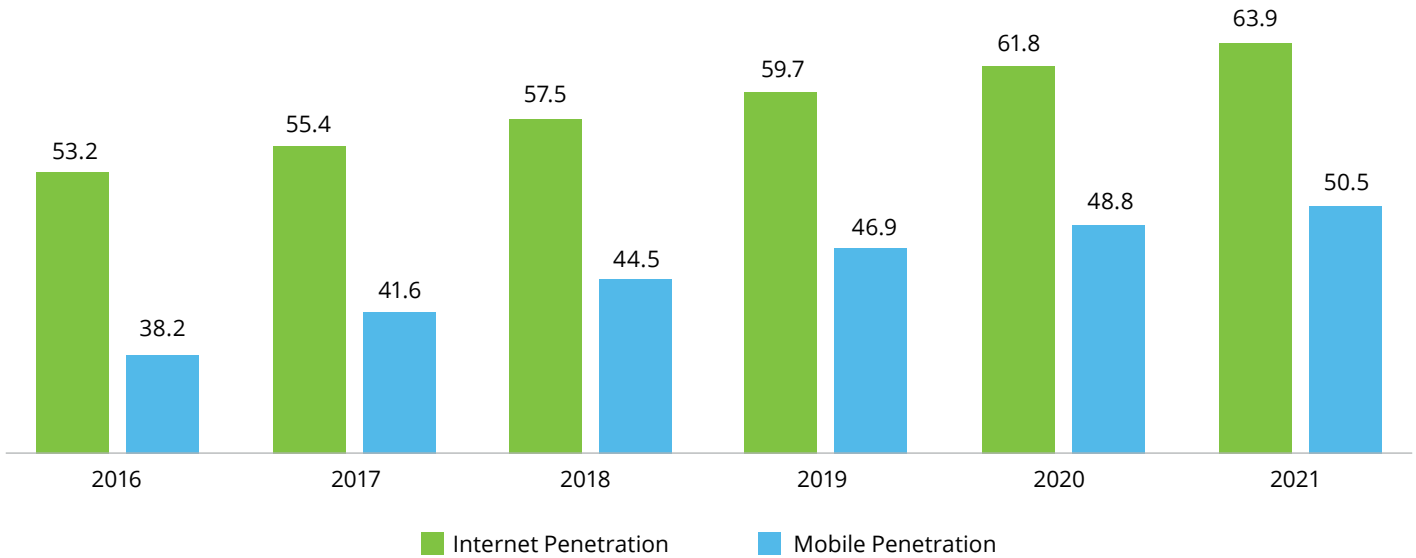
05

### Mobile services



Mobility of humans allows us to encounter more people and go to different places which greatly affects the experiences that we gather. The mobility of the people and the personal devices that we use help in sharing or locations and experiences about numerous situations that we engage in. Sensor industry will continue to benefit on a large scale from mobile devices to build sensors that will be part of the wireless network of communication to provide coverage in biology, air pollution, weather, health, moisture, and motion. These units will be installed on mobile devices, in the environment or attached to the body and will produce a large amount of data about body activities, user movements and user interactions with other people.

Figure 1: Share of world Population connected to Internet



Source: Statista Research

“The field of Data Analytics has been around for the last 20-30 years across sectors like telecom, banking, insurance, etc. with not much focus on the area of Data Science for Governance. The challenges for Data Science in governance are quite different, where at times there isn't enough available data to make better governance or policy decisions. DSC 2017 has planted the seeds for developing the domain of "Data Science for Public Good" in India and I wish in the following years there are more in-depth deliberations on various governance aspects.”

**Avik Sarkar - Head, Data Analytics Cell, NITI Aayog at Government of India**

# Benefits Realized through Data Science

## Potential for Data Science

ICT companies are actively seeking to intensify their use of data science in order to improve current services and create new ones. Data science opens up opportunities for better understanding of their customers, which in turn leads to improved sales and marketing opportunities. ICT companies can work on different type of price model for varying customer needs and market competition.

Fixed and mobile telecommunication network operators, including Internet Service Providers (ISPs), are an important source of data. Most telecommunication data can be considered as the result of an action undertaken such as making a call, sending an SMS, accessing the Internet or recharging a prepaid card.

Data from mobile operators have the greatest potential to produce representative results and reveal



developmental insights on the population, since the service with the widest coverage and greatest uptake and popularity is the cellular service. Not surprisingly, the data for development initiatives have mainly drawn on mobile-network data rather than on those from fixed-line telephone operators or ISPs.

### Competitive Advantage for Early Starters for adopting Data Science

**Customer Profiling:** Thanks to the phenomenal progress on the technology front, ICT companies can capture a wide gamut of behavioural data about their customers. These profiles include details about customers' mobility patterns, social network activity and personal preferences. Collectively, these digital breadcrumbs enables the companies to segment their customers based on a variety of parameters. Depending on the geographical region, there may be different privacy and data regulations governing the manner in which the telecom companies can gather or use this data. This affects the working of the operators to gather insights by behavioural profiling to a great extent.

### Network Planning and Management:

ICT companies can optimize the network routes and improve their Quality of Service by continuously analysing the network traffic in real time. The use of real-time Deep Packet Inspection (DPI) enables gather details of the current traffic volumes, including the geospatial distribution of demand, and to manage their network connections effectively using optimal resource allocation.<sup>2</sup>

Operators can adapt their resource allocation to ensure that more resource is allocated in high-revenue regions where most active customers reside by utilizing the geospatial information from their devices. This is a niche area of great

importance to companies as they seek to understand the demands placed on their networks by the use of popular Over the Top (OTT) services.

### Product Recommendation:

Data Science-developed product recommendation systems learn behavioural shopping patterns, such as purchasing similarities between customers or relatedness of search items, to predict customers' preferences towards new items. These prediction-based recommendations lead to higher sales revenue by exposing customers to additional products of interest and encouraging upgrades to more expensive products

### Pricing Strategies:

Data Science models are commonly used to develop optimal pricing strategies to maximise profits. Dynamic price optimisation is a revenue management tool widely used in retail, automotive, mobile communication and electricity industries, and is generated using data on variability in customer preferences and buying patterns.

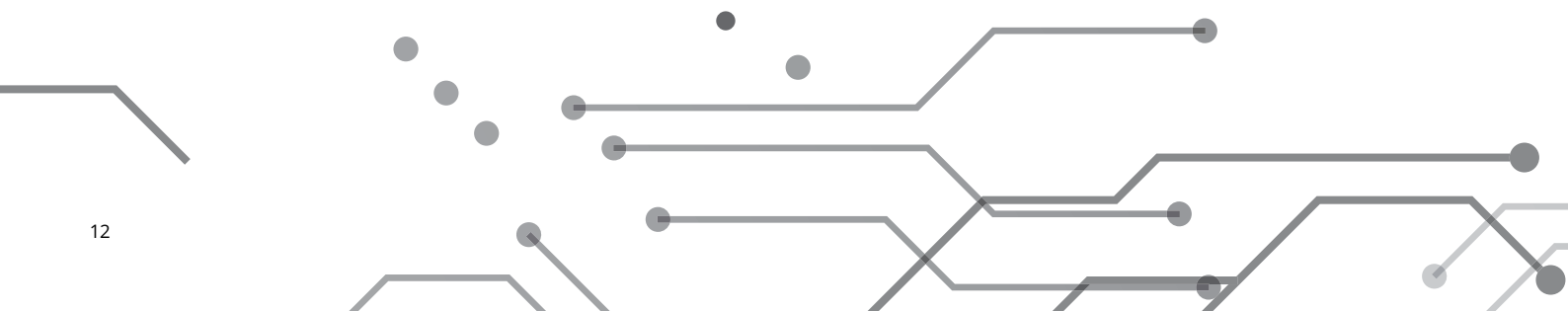
### New Business Lines:

It is natural for operators to leverage the data they hold for better insights to increase their revenue streams. The customer insights obtained paves the way for creation of new business lines, either through innovation or by partnering with other businesses, including credit-scoring and other financial services. One example is of a US-based big data start-up which obtains data from telecom operators and financial firms to build customer portfolios and in turn evaluate the creditworthiness.<sup>2</sup> Cross-promotions with brick-and-mortar businesses are a potentially high-growth area in which the detailed mobility profiles available to operators are leveraged.

**Cyber Security:** The biggest advantage with data science is that it can, indeed, assist security analysts in detecting actual threats more quickly and allow organizations to act proactively. This is achieved through in-depth historical analyses of security data.

Machine data is not just logs, but comprehensive records of behavior of end-users, server, networks, applications, transactions and mobile devices. It's not limited to API data, machine configurations, message queues, events, CDR (call detail records), IoT (Internet of Things) data, sensor data from industrial machines, automation and many others. Consequently, in cyber security, machine data is useful for fraud detection, artificial intelligence and recommendations. Data Science also detect changes over time that render network behavioral profiles of normal vs. abnormal traffic without manual intervention.

**Gaining Competitive Advantage:** In studies conducted up to five years ago, researchers found that companies that use data science in their decision-making were 5% more productive and 6% more profitable than competitors. 17 other studies showed that firms with these capabilities were also five times as likely to make decisions faster than competitors and three times as likely to have faster execution on those decisions.



“In the age of digitalization and artificial intelligence how can we avoid dehumanization of society by machines?

We come from the stance that algorithms with no base in human thinking or solutions that derogate and enslave people are not intelligent in human sense. We believe that only cognitive algorithms that traceably emulate real good human heart logic and moral norms should interact with us humans. Because our mission as the scientific society, individuals and companies is to enable progress of values that make us human.”

**Nikola Sucevic**  
Algorithms/Analytics - Data Science  
Head at Reliance Jio

### Revenue and Growth Benefits

Businesses can use data insights derived using data science to improve customer engagement by better identifying, understanding and responding to customers. For example, speech and text analytics techniques can be applied to document live customer interactions to follow up on potential sales leads. Data science can improve efforts at customer segmentation to predict the most profitable, or most risky customers. In the insurance industry, predictive models that identify high risk customers help insurance businesses minimize losses and develop appropriate premiums. For online businesses with large product ranges, machine-learned ranking (MLR) can improve product search using personalization based on a user’s search and previous purchasing history, which means customers can find and purchase the right products faster.

The powerful combination of business expertise and data science drives better strategic revenue growth decisions. Data Analysis models are commonly used to develop optimal pricing strategies to maximize profits. Dynamic price optimization is a revenue management tool widely used in retail, automotive, mobile communication and electricity industries, and is generated using data on variability in customer preferences and

buying patterns. Data science models are increasingly necessary to handle the volume of telecom data in real-time. By using Data science to make use of all available data and generate customized offers for the right customer at the right time, telecom companies can increase the probability of sales and generate higher revenue.<sup>3</sup>

### Capital Savings

There are many ways data science help businesses make better use of their physical assets and budgets. Optimizing different stages of the production process, from inventory management to quality control can deliver substantial savings to ICT companies. These characteristics are inherent for a leader in data science.

Other ICT companies can certainly use predictive methods of analysis for optimization to save on capital expenditure by following leaders. Models can be trained to utilize user usage data and accordingly optimize the placements of towers and laying of network lines saving the companies a lot in capitals savings. Also, study of the usage can help the companies to take decision with regards to either setting up own services or rent the components from

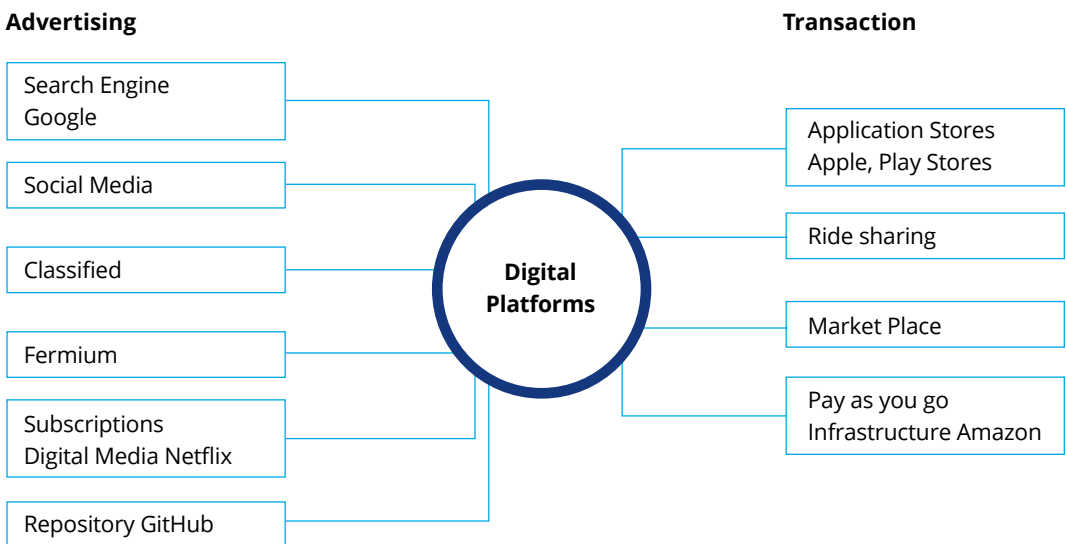
outside. We see that many companies provide services to customers from their competitors so as to save on capital costs and get the same benefit for their customers in return. Thus, data science can play a vital role in optimization of network planning and service delivery for these telecom companies.<sup>3</sup>

**Time and Efficiency Benefits**

Businesses can achieve significant time and efficiency benefits by using Data science applications to cut down on costs and shift human resources to

higher value activities. Data Science techniques aid in the development of applications that automate tasks and augment existing processes, leading to productivity improvements and cost savings for ICT companies. Many routine internal processes can be done quicker and more systematically by machines than people. Models can provide preliminary structure to raw data, saving people from performing routine tasks that are highly time consuming, letting these high-skilled workforce to concentrate on high productivity tasks.<sup>3</sup>

**Figure 2: Early starters extensively uses different digital platform for competitive Advantage**



Source: Statista Research

# Pointers for ICT companies

Data science requires significant investment in terms of time and money. These costs are highly dependent on how a business decided to procure and implement these applications. Leaders have learnt the art of spending in successful projects and follow defined methodology for successful returns.

Deloitte Access Economics' research found that development of successful Data Science applications for small projects generally costs in the order of a few hundred thousand dollars, while larger enterprise-level projects can cost a few million dollars to scale and implement. These costs include the costs of data scientists, obtaining and storing data, and implementation costs.



“Data science is an integral component of content driven web destinations. In late 2000s, data science emerged as a tool to optimize cost and improve efficiency of existing products and services, e.g., how do I recommend more contextual items to users became a pinching problem for most of the e-commerce companies. Data science since then has been applied in many domains such as retail, travel, manufacturing, health-care etc. However as we enter into late 2010s, it has gone beyond just being a tool for optimization, and has become an essential component to build new products and services. It is a very exciting time to be in the field of data science, every industry is going to get disrupted with data driven decision making, and it is going to make the future of humanity a lot simpler and better.”

**Vijay Gabale,**  
Co-founder and CTO at Huew

### Best Practices in Data Science

#### Investment in Resources:

Mix of talented highly skilled cross-functional resources of creative, user experience, analytics, technology, and industry experts who provide leading-class support and development – faster, cheaper, better – empowered.

#### Scrum Development:

Iterative and collaborative development approach, using design-thinking to deliver prioritized Proofs of Concept quickly and frequently in order to maximize responsiveness to dynamic business needs.

#### Reusable Solution

**Components:** Leverage industry best practices, third party data sets, accelerators, and vetted algorithms that aim to save data processing, development, and time.



#### Lab Locations & Lab

**Teams:** Three state-of-the-art Lab locations dedicated to promote innovation and collaboration, in addition to trained “Lab-in-a-Box” teams that are deployable at any location.

#### Next Generation Tools

**& Data:** A modern technology infrastructure that decreases costs and accelerates the value of data science through our Platform-as-a-Service

#### Ecosystem of Partners

**and Vendors:** Partnering with startups and by adopting leading practices of industry based on the latest technology and business innovations.

Eric Schmidt & Jared Cohen, in their book “The New Digital Age” describe a typical future morning for a professional like this:

“There will be no alarm in your wake-up routine – at least, not in the traditional sense. Instead, you’ll be roused by the aroma of freshly brewed coffee, by light entering your room as curtains open automatically, and by a gentle massage administered by your high-tech bed. You’re most likely to awake refreshed, because inside your mattress there’s a special sensor that monitors your sleeping rhythms, determining precisely when to wake so as not to interrupt a REM cycle.”



# The Future

How is data science going to shape up the future business strategies? Let us understand how Netflix and Amazon are using content streams service to drive revenue.

## Understanding data science impact on content delivery network

Have you ever wondered where your video comes from when you watch Amazon Prime, Netflix or Jio TV? Netflix serve video streams out of our own

content delivery network (CDN), which is tailored to one specific application: delivering internet TV to its members around the world. This system alone is responsible to serve 100% of Netflix content, over 125 million hours every day, to 100 million members across the globe!<sup>4</sup>

Such online content providers are the future of content consumption, all thanks to the ease of use and on-the-go access to all your content from multiple



devices. Internet and advances in the communication sector has certainly helped such companies to flourish and garner a huge user-base for their services. But, it is not just the ease of access that has led to the popularity of content streaming service providers; one of the highly evolving field of data science and its pro-active use is at the helm of the growing traffic of consumers that these service provides witness. **It won't be wrong to say that Netflix and Amazon are two completely data-driven**

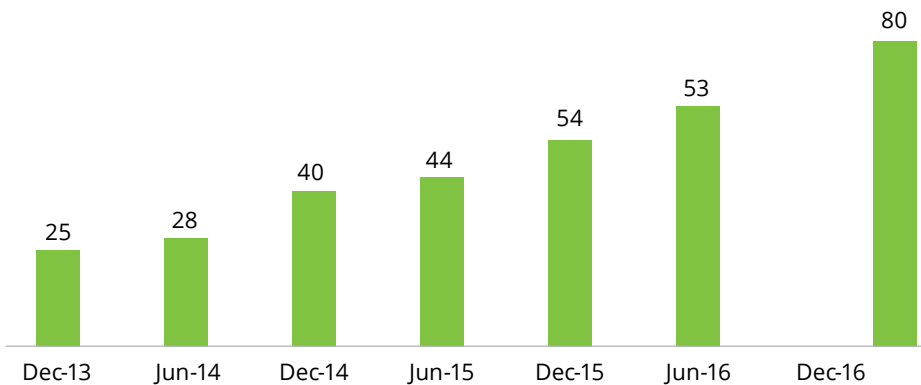
**companies, which use the humongous amount of data they have in the best possible way to improve their services and generate more revenue every day.**

The core job of data science here also is to gain insights into the customers, to optimize and deliver a better product. Data science enables these businesses to make informed decisions and improve their services considerably relying on the quantitative aspect.

Netflix has more than a 100 million customers worldwide and it actively uses the data generated by these users to help them take crucial decisions.<sup>5</sup> Keeping a track of the content watched, completion rates of a particular TV show, age demographics and region-wise indexes, Netflix takes crucial decisions regarding buying licenses and production of Netflix original shows. For example, if Netflix can derive from data that more than 60% of users watched a particular show to completion, it might think to revive or restart the TV show again with a new season.

Netflix captures a lot of data in terms of when someone pauses, rewinds the video, the place of watching, time and day of watching, ratings, user search data, type of device used, etc. It uses all these data points to derive various insights and implement a better, dynamic, robust and more personalized recommendation system for each individual user. This enables them to retrieve their customers for a long term, as more relevant content for each user obviously means that they would continue their subscription.

Figure 3: Amazon Prime Users (In Millions)



Source: Consumer Intelligence Research 2016, Statista Digital Economy

**Future: Long-Term Capacity Planning**

These Content Streaming Channels are very fast, and their content delivery systems are highly dynamic. Some of the many phenomena that can change the system behaviour are catalogue changes, member growth, encoding advances,

and the dynamic electronics ecosystem. Each of these factors greatly affects the way traffic is served from a network location, also what hardware systems will be compatible and effective in the long run with respect to the dynamic nature of changes.

The major data science challenge is to combine these various factors into medium- and long-term forecasts. The work involves a combination of demand forecasting, system modelling—combining all complex factors together to build an efficient performance model—and resource analysis to identify domains of over and under-utilization now and in the near future continuously.

#### Use Cases<sup>6</sup>

##### **Kofera: Digital Marketing in Asia**

Kofera uses data science learning to help e-commerce and online retailers build, optimize and monitor their advertising and marketing campaigns. Based in Indonesia, Kofera's clients range from small and medium businesses (SMBs) to large businesses with over a million products.

Data Science driven data-analytics is the future of marketing optimization. One example involves creating predictive

models that perform dynamic budget allocation. The model is trained on historical data and provides suggestions of how businesses should allocate their marketing budget between different campaigns.

As these budget allocations are updated, the model is retrained on incoming data and can continue to optimize its recommendations. This translates into real cost savings for businesses. Kofera's clients can save over 15% on marketing costs by using campaign monitoring systems which optimize campaigns every 15 minutes instead of weekly or monthly.

These benefits are quickly realizable too – benefits from cleaning data and using NLP are virtually instantaneous, and benefits from marketing optimization are realized in as little as one to two months.

##### **Airbus: Operational Efficiency**

Airbus Defense and Space utilizes data science in a number of applications. One of these is the detection and correction of satellite images with imperfections such as the presence of clouds. For example, it can be challenging to detect the difference between clouds and snow on images by eye. Data Science techniques allow

a previously time consuming task that was prone to human error to be made more efficient and accurate. This allows for analysis of higher quality satellite images and provides information on cloud location so satellite systems can be reprogrammed as needed.

Airbus also uses data science to extract information from satellite images for big data-style analysis, for use in other applications, including for agricultural, engineering or environmental purposes. Prior to using automated analysis techniques, data scientists had to develop rule-based algorithms to extract information in a geometric way; for example specifying that buildings are likely to be rectangular. By picking up on patterns on their own, data science algorithms are an improvement on this, and more accurate.

Airbus expects future applications to continue to achieve improvements. They have already seen a significant reduction in the time to build algorithms, from several years down to a few weeks, and this reduces time to trial so that new applications can be developed.



# Concluding Remarks

The potential business outcomes delivered through data science can be an attractive prospect for ICT sector. However, there are considerations that organizations need to recognize and evaluate to maximize the returns from investment. As with any technology, the successful development and deployment of Data Science applications within an organization requires various capabilities and skills. In developing technology and capability, there are considerations in terms of business outcome, Data Lake, technology architecting and process digitization.

- Business Outcome: Translating a business problem into a testable hypothesis – Defining how to measure successful outcomes from tests – Establishing baselines to enable assessment of incremental benefits delivered by data science use cases – Quantifying success, and translating insights into format that is understood by less technical colleagues – Commercially minded (cost vs benefit etc.).
- Technology Architecting and Implementation: Technology engine which collect, consolidate and nurture data analytics on real time/scheduled jobs for applying data science.

- Process Digitization: Ensure existing processes and data collection channels are digitized and creating data inputs that are relevant for applying data science.
- Data Lake: With the exponential increase in various type of data (internal data, external data, partner's data, competitor data, business process, social data, and people data) organization need to build data lake to solve various data science problems. The Data Lake excels at utilizing the availability of large quantities of coherent data along with deep learning algorithms to recognize items of interest that will power real-time decision analytics.

Implementation and change management is an important final step in a Data Science project. Data science transformations are not short term endeavors, and subsequently require organization-wide transformation. Effective implementation is crucial to realizing the full benefits of Data Science applications; and constant support of the organization and executives are an important part of such endeavors.

# About Aegis

Aegis School of Business, Data Science, Cyber Security and Telecom was founded in the year 2002 with support from Bharti Airtel to develop cross functional technology leaders. In 2015 Aegis and IBM collaborated to launch, India's first Post Graduate Program (PGP) in Data Science, Business Analytics and Big Data and later in 2017 PGP in Cyber Security. These programs are jointly certified and delivered by Aegis School of Business in association with IBM. IBM has set up high end Business Analytics and Cloud Computing Lab at Campus. Also Aegis and NVIDIA partnered for Deep Learning and applied AI courses. Aegis is no. 1 School of Data Science and among top 5 in Business Analytics. Aegis takes up various industry projects, research and consulting assignments in the field of data science under its initiative "Data Science Delivered" and "Data Science for social good", and helping organizations for devolving skills on data science, ML, DL, Big Data, Analytics etc.

For more information, please visit:  
[www.Aegis.edu.in](http://www.Aegis.edu.in) &  
[www.mUniversity.mobi/Aegis](http://www.mUniversity.mobi/Aegis)

Aegis has also initiated worlds largest innovation award for Telecom, Internet, Media and Edutainment (T.I.M.E) and Social, Mobility, Analytics and Cloud (SMAC) know as Aegis Graham Bell Awards for developing an ecosystem for fostering innovation in India. This award in organized with support of Cellular Operators Association of India (COAI); Convergence India; Deloitte and Telecom Centre of Excellence (TCOE).

For more information, please visit  
[www.BellAward.com](http://www.BellAward.com)

In 2017, Aegis launched Data Science Congress to create a vendor-neutral platform where different stakeholders including policy makers, industry, experts, data scientist, CIO, decision makers can share their knowledge, best practices, innovations, products, uses cases and establishing a dialogue among the practitioners, users and tech vendors.

For more information, please visit  
[www.DataScienceCongress.com](http://www.DataScienceCongress.com)

# Acknowledgments

**Hemant Joshi**

**N Ramu**

**Senthilvel Kaliyamurthy**

**Payal Agarwal**

**Amit Kumar**

# Contacts

## **Aegis School of Business, Data Science, Cyber Security & Telecommunication**

7th Floor, CETTM, Technology Street, Hiranandani Gardens,  
Powai, Mumbai - 400076  
Telephone: +91 22 2570 2815  
Website: [www.aegis.edu.in](http://www.aegis.edu.in)  
e-mail: [info@aegis.edu.in](mailto:info@aegis.edu.in)

## **Deloitte**

7th Floor, Building 10, Tower B,  
DLF Cyber City Complex,  
DLF City Phase - II  
Gurgaon, 122 002, India  
Telephone: +91-0-124 679 2396  
e-mail: [inideas-tmt@deloitte.com](mailto:inideas-tmt@deloitte.com)  
Website: [www.deloitte.com/in](http://www.deloitte.com/in)

# References

<sup>1</sup>What Does Big Data Analytics Need from ICT to Develop? Feb 2016, See: <http://blog.agroknow.com/?p=4807>

<sup>2</sup>Measuring the Information Society Report 2014, UN Global Pulse, See: <http://www.unglobalpulse.org/sites/default/files/Pages%20from%20MIS2014%20-%20Big%20Data%20Chapter.pdf>

<sup>3</sup>Deloitte Access Economics – Business Impact of Machine Learning

<sup>4</sup>How Data Science Helps Power Worldwide Delivery of Netflix Content, May 2017, See: <https://medium.com/netflix-techblog/how-data-science-helps-power-worldwide-delivery-of-netflix-content-bac55800f9a7>

<sup>5</sup>Number of Netflix streaming subscribers worldwide from 3rd quarter 2011 to 4th quarter 2017 (in millions), Statista, See: <https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>

<sup>6</sup>Deloitte Access Economics – Business Impact of Machine Learning





Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee (“DTTL”), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as “Deloitte Global”) does not provide services to clients. Please see [www.deloitte.com/about](http://www.deloitte.com/about) for a more detailed description of DTTL and its member firms.

This material is prepared by Deloitte Touche Tohmatsu India LLP (DTTILLP). This material (including any information contained in it) is intended to provide general information on a particular subject(s) and is not an exhaustive treatment of such subject(s) or a substitute to obtaining professional services or advice. This material may contain information sourced from publicly available information or other third party sources. DTTILLP does not independently verify any such sources and is not responsible for any loss whatsoever caused due to reliance placed on information sourced from such sources. None of DTTILLP, Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, the “Deloitte Network”) is, by means of this material, rendering any kind of investment, legal or other professional advice or services. You should seek specific advice of the relevant professional(s) for these kind of services. This material or information is not intended to be relied upon as the sole basis for any decision which may affect you or your business. Before making any decision or taking any action that might affect your personal finances or business, you should consult a qualified professional adviser.

No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person or entity by reason of access to, use of or reliance on, this material. By using this material or any information contained in it, the user accepts this entire notice and terms of use.