Module 2- Introduction to Digital India

Module Overview

This module aims to introduce Digital India. It aims to explain Digital Skills and Data science in detail. With this, you will further understand different roles in data science industry.

Module Objective

At the end of the module, you will be able,

- To understand the meaning and importance of Digital India
- To understand Digital Skills.
- To understand Data Science and its processes.

Introduction to Digital India

Digital India can connect rural areas with high-speed Internet networks, allowing e-commerce to reach parts of India, which are currently served only by unorganized retail.

It was not very long ago that people who shopped online in India were considered 'ahead of time'. But, times have changed, and in today's world the question isn't "Have you ever shopped online"; it's "How often you shop online." We can say with certainty that e-commerce has indeed arrived in India, bearing in mind that we still have a long way to go.

Notwithstanding the phenomenal growth during the last two years, online shopping still represents a minuscule 1-2 per cent of overall retail sales in India. By 2020, India's e-commerce market is expected to reach $100 billion, and there is no doubt this sector will greatly contribute to the Indian economy. The e-commerce industry in India is still in a nascent stage. To start with, in India there are 180 million Internet users and a small portion of them has the access to data connections. The e-commerce industry will grow further with rising disposable incomes, affordable access to data, and higher mobile penetration.

It is time to re-look at regulations for digital companies from a different lens. The policies and frameworks that were designed for brick-and-mortar businesses many years ago need a fresh evaluation for digital companies. For instance, online shopping sites face myriad tax regimes and, in some cases, double taxation too because they deliver goods to customers across different states in

India. It is crucial to implement a standardized tax structure so that issues like double taxation can be avoided. Hopefully, the expected announcements related to the Goods & Services Tax (GST) will result in simplification and standardization of the tax regime for digital companies.

Just like a new house needs a strong foundation, the e-commerce industry still needs a well-developed ecosystem in India. Last mile delivery is one of the weakest links that is limiting the reach to rural India. Therefore, any tax subsidies for logistics companies that would help develop this ecosystem would in turn help fuel the growth of e-commerce in India. In the area of fashion e-commerce, one specific opportunity is around standardization of sizes of clothing products. This would make shoppers more comfortable buying online. However, this type of initiative requires investment from the government into an independent body that sets such standards across the industry

Introduction to Digital Skills

As we begin, it is important to have a common understanding of what is meant by digital skills.

The worldwide expansion of the digital economy and digital society requires us to be equipped with an array of digital skills that will allow us to succeed in work and life. In the world of work, digital skills not only qualify us for jobs in conventional sectors, but also open doors to participate in emerging sectors and even to start our own businesses. People with more advanced digital skills can take advantage of an even wider range of opportunities brought about by ongoing advances in digital technologies, platforms, and devices. Digital skills are particularly important when considering the changing nature of the work environment, including the sharp growth in the use of freelancers and people participating in the gig economy, as well as broader structural changes that will profoundly impact the jobs of the future.

People with relevant digital skills can safely access news and information, communicate with friends and family, and access important services related to e-health, e-government, digital finance,

agrotech, smart transportation, and otherwise enjoy the many benefits of participating in the global knowledge society.

The kinds of digital skills required to succeed are dramatically different today from those required even just five years ago. We used to be able to identify a discrete set of digital skills and have confidence that training programs would equip citizens with those skills. These typically covered topics like basic hardware and software operations, email, and search.

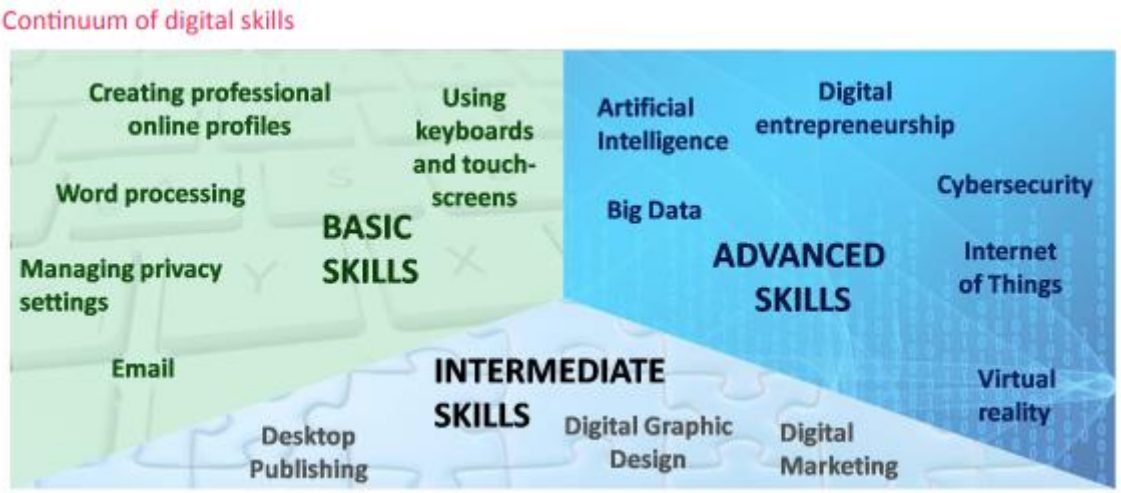
Today, we need continually to review and update those digital skills being taught as a result of new technologies and innovations – **artificial intelligence, big data, block chain, cloud computing, Internet of Things (IoT), machine learning, and mobile applications**. This fast-changing backdrop makes it important for countries with existing digital skills training programs to update their strategies and for those countries who have yet to launch a national digital skills program to take action.

What are Digital Skills?

- Using email appropriately and effectively
- Developing a range of online research strategies
- Using statistical software
- Using word processing packages to produce, format and present written work professionally
- Optimizing use of presentation packages to support the development and delivery of presentations
- Managing personal online identity
- Using a range of tele-communication technologies
- Using technology to support collaborative working
- Utilizing online bookmarking tools to improve online productivity
- Creating or editing video and image

Digital skills are becoming increasingly important and a priority for many industries, include:

- Coding and programming
- Developing and using robotic and automation technologies
- Leveraging information and communication technologies (ICT) skills in business
- Exploring the world of cloud computing and the Internet of Things.



21st century skills

Digital skills take their place within a broader framework, often referred to as '21st century skills'.

According to a World Economic Forum report, 21st century skills are comprised of three 'pillars': foundational skills, competencies, and character qualities, as set out in the graphic below. Sometimes

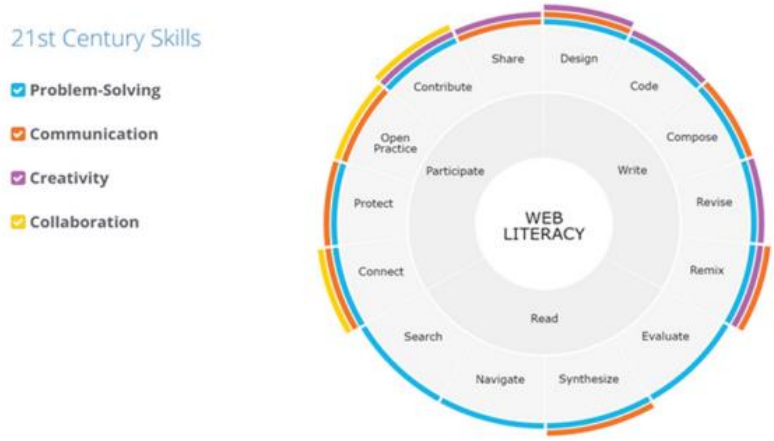
21st century skills are called 'soft skills'. As shown here, digital skills (referred to in the graphic as 'ICT literacy', fall under the Foundation category. This underlines the great importance of the connection between digital skills and other competencies and character qualities – all within an overall system of lifelong learning

21st century skills

Mozilla is one organization that has explicitly designed its framework to embrace 21st century skills.

As shown below, each of the skill areas is connected to one or more 21st century skills. Each skill area contains a number of curricular offerings where, for instance, one develops problem-solving and creativity competencies while learning how to code.

Web literacy



Emerging and specialized skills

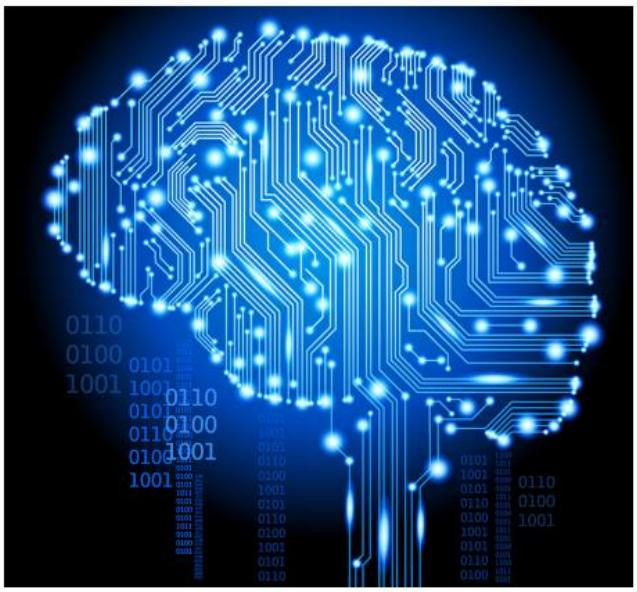
In many respects, emerging and specialized skills represent the direction in which digital skills programmes are heading, and thus offer future-oriented opportunities for countries to consider. Research undertaken in 2016 underlines how critically important the teaching of such skills is at school level: estimates of numbers graduating from coding bootcamps, while encouraging, are falling way below what is needed to bridge the anticipated shortfall in digital skills. Thus if we are to ever achieve scale on developing intermediate and advanced digital skills, such skills need to be integrated into schools' curricula around the world. The topics are: computational thinking, data literacy and mobile literacy.

Computational thinking and coding

Recent years have witnessed burgeoning interest in the teaching of computational thinking as a basic literacy, on a par with reading and writing. **Computational thinking** is 'a problem solving process that includes a number of characteristics and dispositions. Computation has also been described as 'a way of solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computer science. To flourish in today's world, computational thinking has to be a fundamental part of the way people think and understand the world.'

In terms of content, **computational thinking** involves 'problem solving, examining data patterns, decomposing problems, using algorithms and procedures, making simulations, computer modelling, and reasoning about abstract objects.'

In practice, teaching computational thinking can begin when children are of primary school age, and many countries are doing this as described later in the toolkit. Thus, computational thinking starts at a basic level, but rises through to advanced topics. Teaching computational thinking can be introduced with exercises that require no technology and which then gradually add the use of computers and other devices. As with all skills included in this toolkit, there are many online resources that can be used to teach computational thinking such as Google's Computational Thinking for Educators.



Data literacy

People who can derive meaningful information from data are in high demand in every sector, a trend driven by the global explosion in big data and the proliferation of sophisticated tools to manage, analyse, and visualize data. Some observers suggest that in 10 years, data experts will have replaced computer experts as being a desirable profession. As such, people with data skills will be found in every industry -- from SMEs to large corporations -- just as people with computer skills are now needed across all job sectors. While data scientists with advanced mathematical and statistical expertise represent the top tier, many organizations are demanding data-based literacy and specific skills from employees:

- Knowing what data is appropriate for a particular purpose
- Interpreting data visualizations, such as graphs and charts
- Thinking critically about information yielded by data analysis
- Understanding data analytic tools and methods, and when and where to use them
- Recognizing when data is being misrepresented or used misleadingly
- Communicating information about data to people lacking data literacy, an ability sometimes referred to as data storytelling

Mobile literacy

Of the five billion people with mobile phones, nearly half (47 percent) “mainly use their device to place a voice call or send a text message.”

As increasing numbers of these people move from feature phones to smartphones, and as the next billion comes oncloline, many will skip the stage of using personal computers (PCs) and feature phones entirely, moving directly to powerful handheld computers, i.e. smartphones. There is a critical need to bridge the gap between using a phone for basic functions and using sophisticated

smartphones. Mobile literacy has emerged to meet this need by identifying the necessary digital skills smartphone users need in order to optimize their mobile internet experience.

Activity

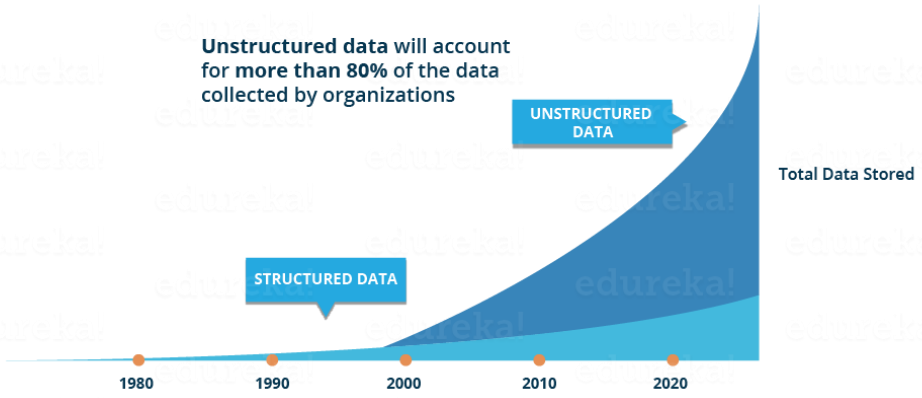
Trainer will divide the participants into five groups and ask them to prepare and present chart paper activity on Digital Skills. Post this activity, each group will present chart in front of the class. After the trainer's feedback, participants will write their understanding in the space given below

NOTE

Introduction to Data Science

Let's Understand **Why We Need Data Science?**

Traditionally, the data that we had was mostly structured and small in size, which could be analyzed by using the simple BI (Business Intelligence) tools. Unlike data in the traditional systems which was mostly structured, today most of the data is unstructured or semi-structured. Let's have a look at the data trends in the image given below which shows that by 2020, more than 80 % of the data will be unstructured.



This data is generated from different sources like financial logs, text files, multimedia forms, sensors, and instruments. Simple BI tools are not capable of processing this huge volume and variety of data. This is why we need more complex and advanced analytical tools and algorithms for processing, analyzing and drawing meaningful insights out of it.

Let's dig deeper and see how Data Science is being used in various domains.

How about if you could understand the precise requirements of your customers from the existing data like the customer's past browsing history, purchase history, age and income. No doubt you had all this data earlier too, but now with the vast amount and variety of data, you can train models more effectively and recommend the product to your customers with more precision. Wouldn't it be amazing as it will bring more business to your organization?

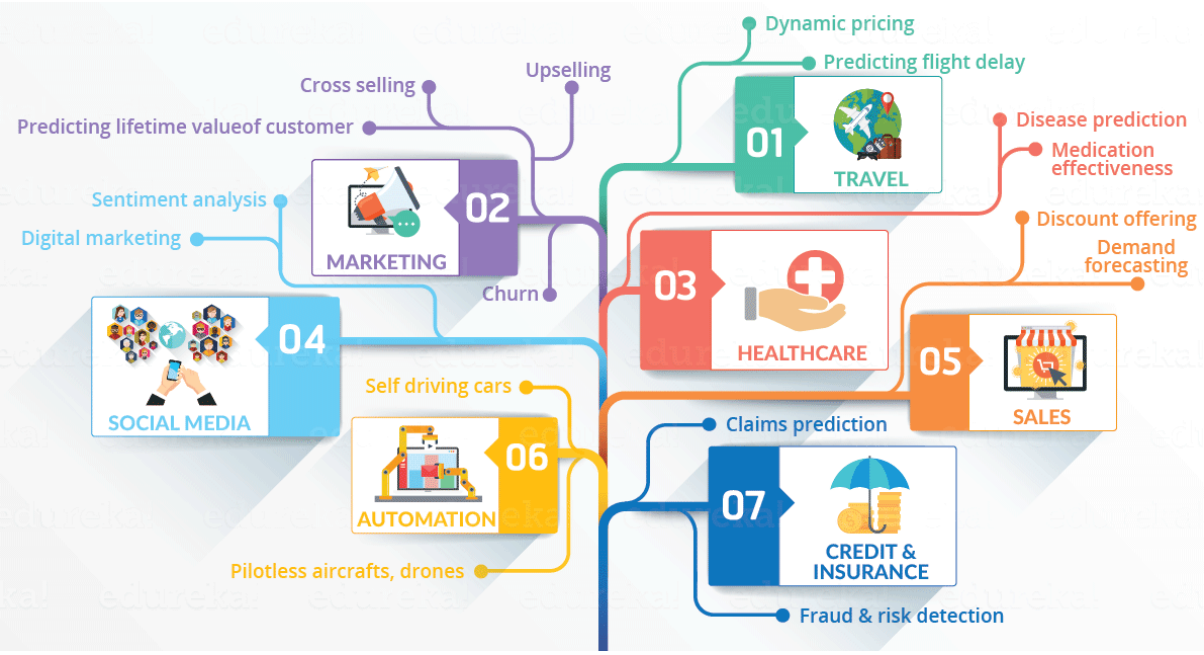
Let's take a different scenario to understand the role of Data Science in decision making. How about if your car had the intelligence to drive you home?

The self-driving cars collect live data from sensors, including radars, cameras and lasers to create a map of its surroundings. Based on this data, it takes decisions like when to speed up, when to speed down, when to overtake, where to take a turn – making use of advanced machine learning algorithms.

Let's see how Data Science can be used in predictive analytics. Let's take weather forecasting as an example.

Data from ships, aircrafts, radars, satellites can be collected and analyzed to build models. These models will not only forecast the weather but also help in predicting the occurrence of any natural calamities. It will help you to take appropriate measures beforehand and save many precious lives.

Let's have a look at the below infographic to see all the domains where Data Science is creating its impression.



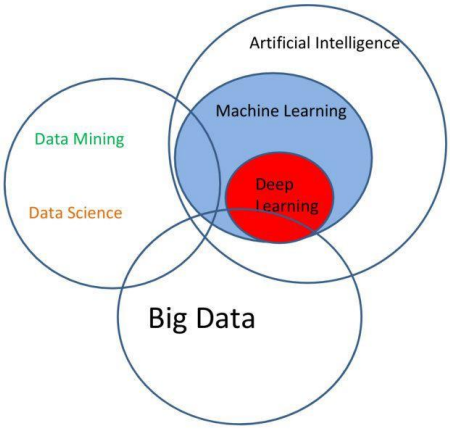
Different parts within data science can be as following, the topics will be explained in detail.

- **Big Data:** Collecting and processing any data which is huge in volume, arrival/processing rate or invariant in structure.
- **Data Mining:** Process of finding out hidden patterns in the structured data and find hidden information in the data
- **Data Analytics:** It is a process which is one step above data mining. Data analytics identifies the type of the analysis to be performed within which data mining techniques will be performed.

- **Data Analysis:** It is a more general approach of finding insights out of the raw data by forming a hypothesis and proving them using statistical tests.

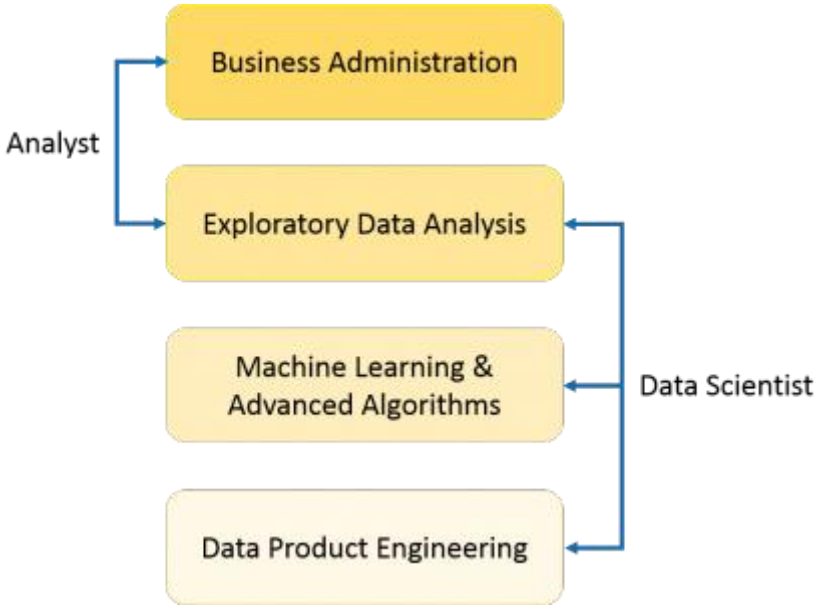
- **Data Science:** It defines the process of understanding the business problem to deliver the solution

- **Machine Learning:** It is a tool used in data analytics to predict/find out a hidden layer of information in data. An example can be predicting the attrition rate of an organisation/whether an employee will stay in the organisation or leave it.



What is Data Science?

Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data.



As you can see from the above image, a Data Analyst usually explains what is going on by processing history of the data. On the other hand, Data Scientist not only does the exploratory analysis to discover insights from it, but also uses various advanced machine learning algorithms to identify the occurrence of a particular event in the future. A Data Scientist will look at the data from many angles, sometimes angles not known earlier.

Data Science is primarily used to make decisions and predictions making use of predictive causal analytics, prescriptive analytics (predictive plus decision science) and machine learning.

- **Predictive causal analytics:** If you want a model which can predict the possibilities of a particular event in the future, you need to apply predictive causal analytics. Say, if you are providing money on credit, then the probability of customers making future credit payments on time is a matter of concern for you. Here, you can build a model which can perform predictive analytics on the payment history of the customer to predict if the future payments will be on time or not.

- **Prescriptive analytics:** If you want a model which has the intelligence of taking its own decisions and the ability to modify it with dynamic parameters, you certainly need prescriptive analytics for it. This relatively new field is all about providing advice. In other terms, it not only predicts but suggests a range of prescribed actions and associated outcomes.

  The best example for this is Google's self-driving car which we had discussed earlier too. The data gathered by vehicles can be used to train self-driving cars. You can run algorithms on this data to bring intelligence to it. This will enable your car to take decisions like when to turn, which path to take, when to slow down or speed up.

- **Machine learning for making predictions**: If you have transactional data of a finance company and need to build a model to determine the future trend, then machine learning

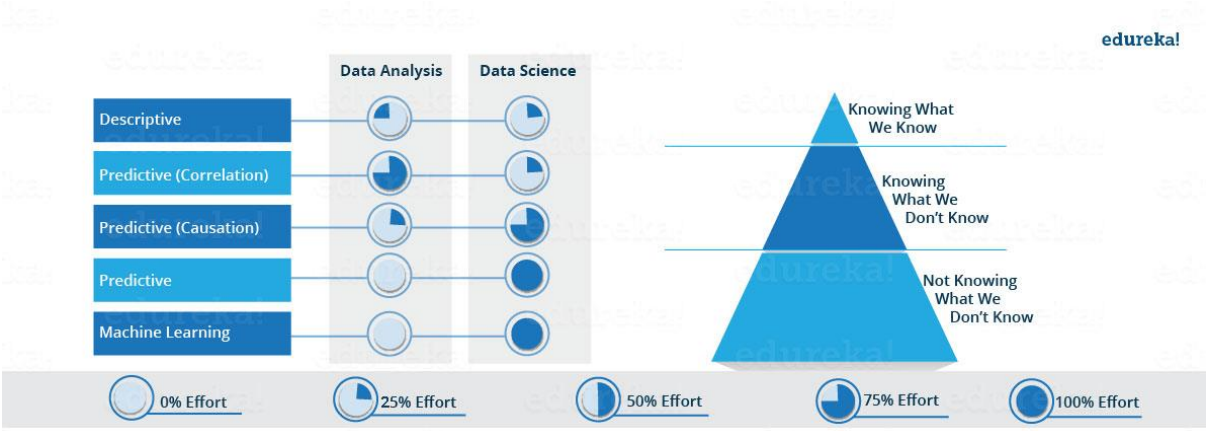
algorithms are the best bet. This falls under the paradigm of supervised learning. It is called supervised because you already have the data based on which you can train your machines.

For example, a fraud detection model can be trained using a historical record of fraudulent purchases.

- **Machine learning for pattern discovery:** If you don't have the parameters based on which you can make predictions, then you need to find out the hidden patterns within the dataset to be able to make meaningful predictions. This is nothing but the unsupervised model as you don't have any predefined labels for grouping. The most common algorithm used for pattern discovery is **Clustering**.

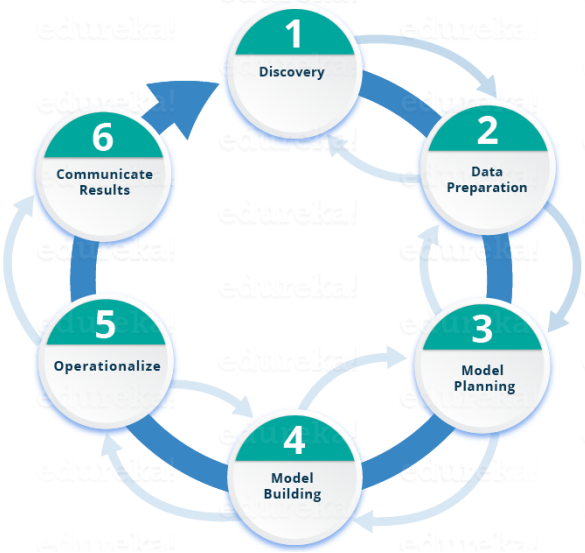
Let's say you are working in a telephone company and you need to establish a network by putting towers in a region. Then, you can use the clustering technique to find those tower locations which will ensure that all the users receive optimum signal strength.

Let's see how the proportion of above-described approaches differ for Data Analysis as well as Data Science. As you can see in the image below, Data Analysis includes descriptive analytics and prediction to a certain extent. On the other hand, Data Science is more about Predictive Causal Analytics and Machine Learning.



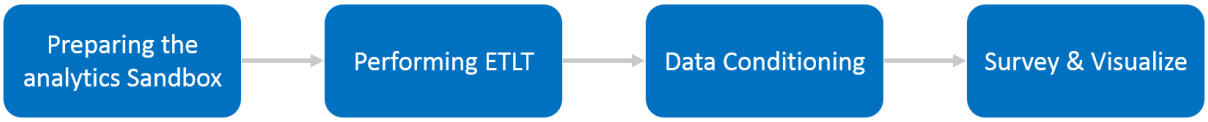
Lifecycle of Data Science

Here is a brief overview of the main phases of the Data Science Lifecycle:

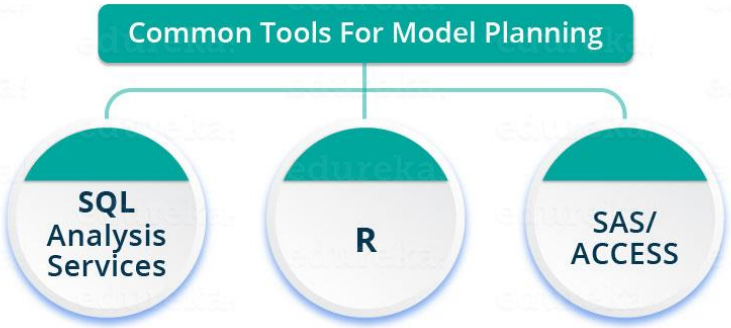


Phase 1—Discovery: Before you begin the project, it is important to understand the various specifications, requirements, priorities and required budget. You must possess the ability to ask the right questions. Here, you assess if you have the required resources present in terms of people, technology, time and data to support the project. In this phase, you also need to frame the business problem and formulate initial hypotheses (IH) to test.

Phase 2—Data preparation: In this phase, you require analytical sandbox in which you can perform analytics for the entire duration of the project. You need to explore, pre-process and condition data prior to modeling. Further, you will perform ETLT (extract, transform, load and transform) to get data into the sandbox. Let's have a look at the Statistical Analysis flow below.



Phase 3—Model planning: You will determine the methods and techniques to draw the relationships between variables. These relationships will set the base for the algorithms which you will implement in the next phase. You will apply Exploratory Data Analytics (EDA) using various statistical formulas and visualization tools. Let's have a look at various model planning tools.

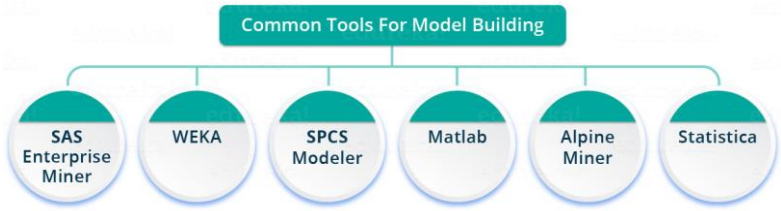


- **R** has a complete set of modeling capabilities and provides a good environment for building interpretive models.

- **SQL Analysis** services can perform in-database analytics using common data mining functions and basic predictive models.
- **SAS/ACCESS** can be used to access data from Hadoop and is used for creating repeatable and reusable model flow diagrams.

Phase 4—Model building: In this phase, you will develop datasets for training and testing purposes. You will consider whether your existing tools will suffice for running the models or it will need a more robust environment (like fast and parallel processing). You will analyze various learning techniques like classification, association and clustering to build the model.

You can achieve model building through the following tools.



Phase 5—Operationalize: this phase, you deliver final reports, briefings, code and technical documents. In addition, sometimes a pilot project is also implemented in a real-time production environment. This will provide you a clear picture of the performance and other related constraints on a small scale before full deployment.

Phase 6- Communication results: Now it is important to evaluate if you have been able to achieve your goal that you had planned in the first phase. So, in the last phase, you identify all the key findings, communicate to the stakeholders and determine if the results of the project are a success or a failure based on the criteria developed in Phase 1.

Different Roles in the Data Science Industry

There are multiple roles a professional can take in the data science industry which are in a lot of demand too. These roles all deal with data in some way or the other but are different from each other depending on what you do with data.

The Data Scientist

He/she masters a whole range of skills and talents going from being able to handle the raw data, analysing that data with the help of statistical techniques, to share his/her insights with his peers in a compelling way. No wonder these profiles are highly wanted by companies like Google and Microsoft.

The Data Analyst

He/She is a master of languages like R, Python, SQL and C. Main responsibility is collecting, processing and performing statistical data analysis.

The Data Engineer

The data engineer often has a background in software engineering and loves to play around with databases and large-scale processing systems.

The Data Architect

With the rise of big data, the importance of the data architect's job is rapidly increasing. The person in this role creates the blueprints for data management systems to integrate, centralise, protect and maintain the data sources. The data architect masters technologies like Hive, Pig and Spark, and needs to be on top of every new innovation in the industry.

The Data Statistician

The historical leader of data and its insights. Although often forgotten or replaced by fancier sounding job titles, the statistician represents what the data science field stands for: getting useful insights from data

The Machine Learning Engineer

Artificial intelligence is the goal of a machine learning engineer. They are computer programmers, but their focus goes beyond specifically programming machines to perform specific tasks. They create programs that will enable machines to take actions without being specifically directed to perform those tasks. An example of a system a machine learning engineer would work on is a self-driving car. They take the key role of providing the intelligence to the work done by analysts, for example, forecasting sales of products, segmenting different types of customers based on their habits and traits etc

The Business Analyst

Less technically oriented, the business analyst makes up for it with his/her deep knowledge of the different business processes. (S)he masters the skill of linking data insights to actionable business insights and is able to use storytelling techniques to spread the message across the entire organization.

Trainer will take the participants to the computer lab and ask the participants to do internet research on different types of Data Science Technologies.