jims
Sector-5, Rohini, Delhi

October 2016, Issue-2

IT Flash Newsletter

DATA SCIENCE "PROCESS AND ITS TOOLS"

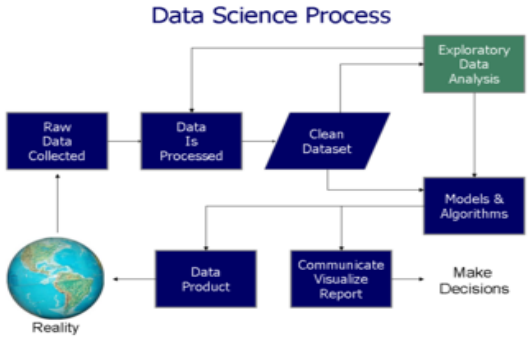
Data is increasingly cheap and ubiquitous. We are now digitizing analog content that was created over centuries and collecting myriad new types of data from web logs, mobile devices, sensors, instruments, and transactions. IBM estimates that 90 percent of the data in the world today has been created in the past two years.
The rise of "big data" has the potential to deepen our understanding of phenomena ranging from physical and biological systems to human, social and economic behavior.

Let's understand the term "DATA SCIENCE"

Data science is an investigation of where data originates from, and how it can be transformed into a significant resource, how it can be beneficial to the fields like business examination that consolidates software engineering. Data Science includes utilizing techniques and numerous algorithms to investigate massive measures of information and to concentrate learning from them. Mining a lot of structured and unstructured information can help an organization get control over costs, increase operational efficiency, perceive new market trends and increment the organization's upper hand.

"DATA SCIENCE PROCESS"

In data science, various components are involved such as organisation of data, packaging of data and delivering of data .Organising is the first most phase in which physical location and structure of the data is planned and then executed. Packaging is where prototypes are constructed, the statistics are performed and the visualization is produced. Delivering is the last most phase where results gets placed and value of the various operations are obtained. However, what separates data science from all other existing roles is that they also need to have a continual awareness of What, How, Who and Why. A data scientist needs to know what will be the output of the data science process and must have a clear vision to this output. A data scientist needs to have a clearly defined plan as to how this output is required to be achieved within the constraints of available resources and time. A data scientist needs to deeply understand who are the people that will be involved in creating the output and most of all the data scientists must know why there is a motivation behind attempting to manifest the creative visualization.



In this highly competitive world of business, data is an invaluable resource. Almost all major business are data-driven i.e. they're collecting the enormous amount of data and the conclusions they are drawing out of that data. This whole process starts with the process of collection of raw facts and figures generated from the business operations. As the data is huge and scattered, it is processed or mined so that cleaned data set is obtained. After the cleaning process, data is clustered and analyzed so that these data patterns could make sense. After analyzing the data, data is then visualized in the form of statistical reports and predictive analysis is done on the basis of these reports. Thus, these statistics are used to base the company's decisions, define market scope, products for the business or may be suitable or targeted customers for the business.

TOOLS AND TECHNIQUES USED IN DATA SCIENCE

Data that is handled by the data scientists is not small, we're talking here about big data (both structured and unstructured). Thus, to handle this data, multiple technologies are used as data science is not a single step process and has different technical aspects and requirements at every step. Some of the technologies and techniques used in data science are:

A) Hadoop –Basically, Hadoop is an open source software designed for managing the storage and utilization problems of the big data by bringing the analytics to the data for processing by using powerful functionalities like MapReduce. Prominent users of Hadoop are Facebook and Yahoo.

B) Spark – Spark is a framework based on clustered computing. It has the ability to run over Hadoop file system and overcome the speed related problems of traditional MapReduce by providing data parallelism and in memory storage instead of disk making analysis faster.

C) Python – Python is a high level language which provides excellent libraries for data science. Some of these libraries are Scipy and Numpy for statistic modelling, Orange and Pattern for data mining, Scikit library for machine learning, etc. Python has attracted users like EverNote and Spotify.

.

D) R Language – R is a famous software package which is used to perform mathematical operations and the visualization of data after the cleaning and mining of data is done. It is very famous among data scientists as it provides thousands of extension packages that are needed for statistical modeling of data. Companies related to finance, health care, marketing business, etc. use R language for analysis. Currently companies like Bing, Facebook and Bank of America are using R for analysis.

E) Scala – Scala programing language has become increasingly popular in data mining because it can perform both object oriented (OOP) and functional programming. Companies like FourSquare, LinkedIn, Siemens and Twitter are using Scala language.

F) Excel – Excel is a widely available tool that can perform highly sophisticated analysis of data. Though it is not capable of analysis of massive data as compared to Hadoop but it is powerful enough for small scale analytic projects which include clustering, optimization and predictive analysis using supervised machine learning.

G) SAS – SAS or Statistical Analysis System is a data mining software suite used for data science features like data management, advanced analytics and social media analytics. It uses its analytics features for quality predictive modeling of data and visualization. Despite being very costly tool, it is very famous among the data scientists due to its integration with open source tools like R, Python and Hadoop. Thus, making it a very powerful tool for data science

Apart from these techniques and technologies, many other technologies are used like IBM SPSS, A/B Testing, Optimizely, Maxymyser, Clojure language, Java language, Adobe Target, etc.

-Komal Jaiswal and Nitin Jain-MCA- I and IInd Year

"AI- NIGHTMARE MACHINE KNOWS WHAT SCARES YOU"

The idea of artificial intelligence (AI) — autonomous computers that can learn independently. Those individuals probably wouldn't find it reassuring to hear that a group of researchers is deliberately training computers to get better at scaring people. The project, appropriately enough, is named "Nightmare Machine." Digital innovators in the U.S. and Australia partnered to create an algorithm that would enable a computer to understand what makes certain images frightening, and then use that data to transform any image, no matter how harmless-looking, into the stuff of nightmares.

Images created by Nightmare Machine are unsettling. Iconic buildings from around the world appear eroded and distorted within shadowy settings or amid charred and smoldering landscapes, glimpsed through what appears to be murky, polluted water or toxic gas clouds.
Nightmare Machine's faces are equally disturbing. Some of the subjects are almost abstract, but subtle — creepy suggestions of hollow eyes, bloody shadows and decaying flesh still cause unease. Even the lovable Muppet Kermit the Frog emerges from the process as a zombie-like creature that would terrify toddlers and adults, too.

The primary reason for building Nightmare Machine was to explore the common fear inspired by intelligent computers. They wanted to playfully confront the anxiety inspired by AI, and simultaneously test if a computer is capable of understanding and visualizing what makes people afraid.

The designers used a form of artificial intelligence called "deep learning" — a system of data structures and programs mimicking the neural connections in a human brain.

"Deep-learning algorithms perform remarkably well in several tasks considered difficult or impossible," The research group's main goal is to understand the barriers between human and machine cooperation.

Nightmare Machine could use your help to learn how to become even scarier.

"DDoS ATTACK THAT DISRUPTED INTERNET WAS LARGEST OF ITS KIND IN HISTORY"

Friday, 21st October 2016

On Friday, 21st October 2016, When people of Europe and America having their prime time for casual news, reading, tweeting etc.. but some of them had trouble in accessing their usual sites and services that day. Similar problems were reported across the globe where their servers of Dyn, a company that controls much of the internet's domain name system (DNS) infrastructure were accessed and on investigation, it was found that the problem was due to to **Distributed Denial of Service attack (DDoS).**

A Distribute Denial of Service (DDos) attack is an attempt to make an online service unavailable by overwhelming it with traffic from multiple sources.

The cyber-attack that brought down much of America's internet on 21st October, 2016 was caused by a new weapon called the Mirai botnet and is likely the largest of its kind in history.

It was remained under sustained assault for most of the day, bringing down sites including Twitter, the Guardian, Netflix, Reddit, CNN and many others in Europe and the US.

The cause of the outage was a Distributed Denial of Service (DDoS) attack, in which a network of computers infected with special malware, known as a "botnet", were coordinated into bombarding a server with traffic until it collapsed under the strain

Mirai was the "primary source of malicious attack traffic".

Unlike other botnets, which are typically made up of computers, the **Mirai botnet is largely made up of so-called "Internet of Things" (IoT) devices such as digital cameras and DVR players.**

Because it has so many internet-connected devices to choose from, attacks from Mirai are much larger than what most DDoS attacks could previously achieve. Dyn estimated that the attack had involved **"100,000 malicious endpoints", and the company, which is still investigating the attack,** said there had been reports of extraordinary attack strength of 1.2Tbps.

Sumedha Sen and Tanisha Jain
BCA-IInd Year and MCA-Ist year

"ONLINE PRIVACY MAY BE BOOSTED BY US FCC DATA RULES"

28th October 2016,

Earlier, customer's personal information can be easily shared by the Internet Service Provider with the Ad agencies without seeking their consent and customer starts getting emails of the product which is not of their interest. Now, customer can easily get rid of from at least some online tracking easily by new approved rules. On 27th October 2016, The Federal Communications Commission approved rules that Internet service provider like **Comcast, AT&T and Verizon** need to ask customers' permission to use or share much of their data with third parties like advertisers.

According to the FCC's new rules, The internet service providers will be required to get a customer's explicit consent before they can use or share "sensitive" personal information. Sensitive data includes browsing history, mobile location data, TV viewing history, call and text message records, and mobile apps data we use.

The rules also require service providers to spell out to consumers what data they collect and why. It requires service providers to notify customers of data breaches. It's the consumers' information, not the network's, unless the consumer gives permission"- FCC Chairman Tom Wheeler said during a press conference. **"Internet Service Providers shouldn't be able to sell something that isn't theirs without customer's permission"**. This could potentially make it harder for Ad-agencies to build advertising businesses that could serve as competitors to Google and Facebook**. Google and Facebook-digital-ad behemoths are not covered by the new FCC rules.**

Cable and phone companies want to increase revenue from ad businesses of their own - AT&T has said increasing advertising tailored to customers' preferences is one of its goals with its $85.4 billion purchase of HBO, CNN and TBS owner Time Warner. Verizon has bought AOL and agreed to buy Yahoo in order to build up a digital-ad business.

FCC officials approved the said rules in the last week of October, 2016 and these rules will pose new challenges for companies like HBO, CNN, and Time Warner.

"MAKE YOUR PASSWORD HACK PROOF BY SENDING IT THROUGH YOUR BODY"

When wireless signals are sent through the air, there is a possibility that they can be intercepted by a hacker. If this wireless communication contains sensitive information like a password, this is particularly concerning.

When we send a wireless WiFi signal, it's broadcasting every-where and any eavesdropper wifi channel can hack into the stream and try to break that encrypted password.

"**Fingerprint sensors have so far been used as an input device. It can be shown for the first time that fingerprint sensors can be re-purposed to send out information that is confined to the body,"** said Dr. Shyam Gollakota, UW Assistant Professor of Computer Science and Engineering.

These **"On Body Transmissions"** offers a more secure way to transmit authenticated information between devices that touch parts of the body such as a smart door lock or wearable medical device and a phone or device that confirms your identity by asking you to type in a password.

Let's say if one wants to open a door using an electronic smart lock then one can touch the doorknob and touch the fingerprint sensor on his/her phone and transmit his / her secret credentials through his /her without leaking that personal information over the air."

The research team tested the technique on iPhone and other fingerprint sensors as well as Lenovo laptop trackpads and the Adafruit capacitive touchpad.

The research team from the UW's Networks and Mobile Systems Lab systematically analyzed smartphone sensors to understand which of them generates low-frequency transmissions below 30 megahertz that travel well through the human body but don't propagate over the air.

The researchers found that fingerprint sensors and touchpads generate signals in the 2 to 10 megahertz range and employ capacitive coupling to sense where your finger is in space and to identify the ridges and valleys that form unique fingerprint patterns.

"UNKNOWN FACTS IN IT"

- The Password for the computer controls of nuclear tipped missiles of the U.S was 00000000.
- The first 1 GB hard disk was announced in 1980 which weighed about 550 pounds and had a price tag of $40, 000.
- The first microprocessor created by Intel was the 4004. It was designed for a calculator.

- The 30th of November is known as "Computer Security Day".
- Hewlett Packard was started at garage in Palo Alto in 1939.
- The First Computer mouse was invented by Doug Englebart in around 1964 and was made of wood.

Incharge: Ms. Chandni Kohli

Editorial Team: Dr. Praveen Kumar Gupta
Mr. Sanjive Saxena