SIGNIFICANCE OF DATA SCIENCE APPLICATIONS IN FINANCIAL FUNCTIONS

Prof Pankaj Roy Gupta

C4i4 & KIAMS

C4i4 LAB CENTRE FOR INDUSTRY 4.0®

Session Coverage

- Section-I : Understating the Data Science
- Section-II : Data Science Applications in Business
- Section-III: Data Science Applications in Financial Functions

- Q & A
- Conclusion

SECTION 1: About Data Science

What is Data?

- Data is a **collection of raw (un-processed) facts and figures or (processed) information** from which conclusions are extracted.
- Data is collected as observations in the form of **audio & sound, images** and **videos** by using Information Technology and stored in computer systems for extracting intelligence for decision making.
- There are two general ways to represent data: **analog** and **digital**.
- Analog data are continuous. They are 'analogous' to the actual facts they represent.
- Digital data are discrete, broken up into a limited number of elements.
- Nature is analog, while computers are digital. Many aspects of our natural world are **continuous** in nature.
- Facts (as measurements and statistics) used as a basis for reasoning, discussion or calculation
- The number 17 million is an example of data – with no meaning
- Example of Data – The data shown below are Mark's scores on five Math tests conducted in 10 weeks of Ritesh.
  45, 23, 67, 82, 71
  The data helps us compare his scores and learn his progress

Types of Data (13 different Forms) as per Forbes for Data Science

S.N.	Type of Data	Description
1	**Structured, Unstructured and Semi-structured Data**	
2	**Time stamped Data**	data captured (event time) or collected (processed time)
3	**Machine Data**	is the digital data created by the systems, technologies and infrastructure
4	**Spatiotemporal Data**	location and time in the same event
5	**Open Data**	is freely available to anyone in terms of its use and rights without any restrictions
6	**Dark Data**	is digital data and information that is not being used and lies dormant in some form
7	**Real Time Data**	it can help with everything from deploying emergency resources in a road crash to helping traffic flow more smoothly during a citywide event
8	**Genomic Data**	involves analyzing the DNA of patients to identify new drugs and improve care with personalized treatments
9	**Operational Data**	business execution operational data using ERP, SCM, CRM, B2B, Banking Retail Application etc.
10	**High Dimensional Data**	is data related with facial recognition technologies, image processing
11	**Unverified Outdated Data**	is data that has been collected, but nobody has any idea whether it's relevant, accurate or even of the right type
12	**Translytic Data**	Data collected using technology architecture and hybrid transactional analytic database systems, which are enabled by the in-memory technology
13	**BIG Data**	data that will not practically fit into a standard (relational) database for analysis and processing caused by the huge volumes of information being created by human and machine-generated processes

What is Science ?

1. **Science** came from a Latin word **“Scientia”** which means knowledge, a knowing, application, expertness, or experience or a corpus of human knowledge.

2. **Science** is defined as the observation, identification, description, experimental investigation, and theoretical explanation of natural phenomena, related with earth, nature, surroundings, society, business etc

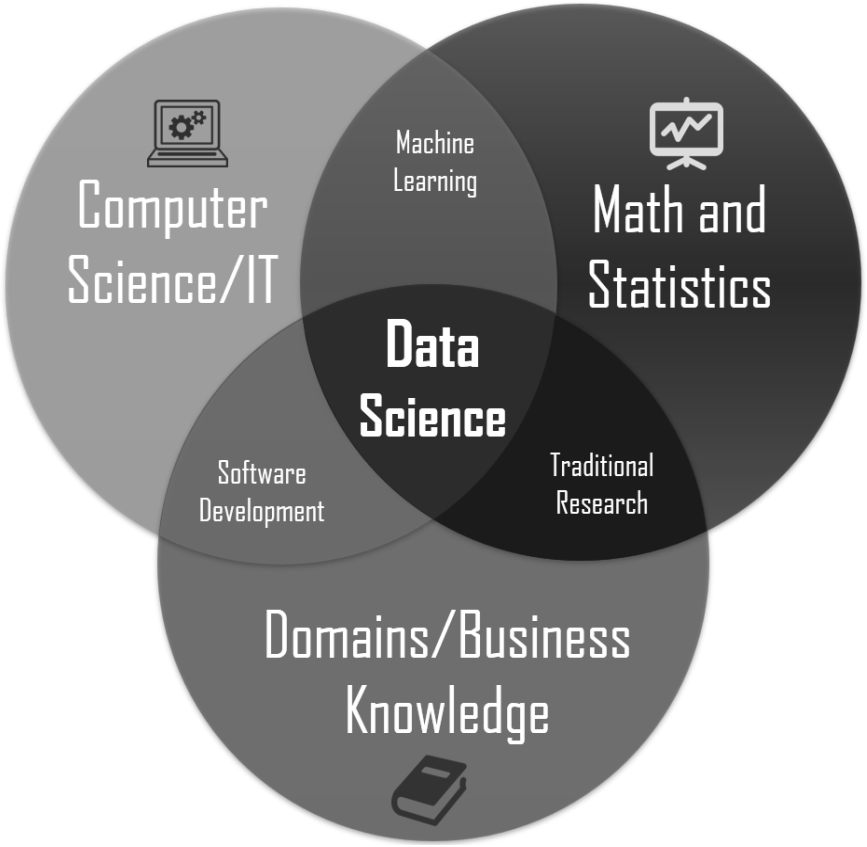
3. **Science** is the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence. Scientific methodology includes... Evidence. Experiment and/or observation as benchmarks for testing hypotheses etc.

What is Data Science ?

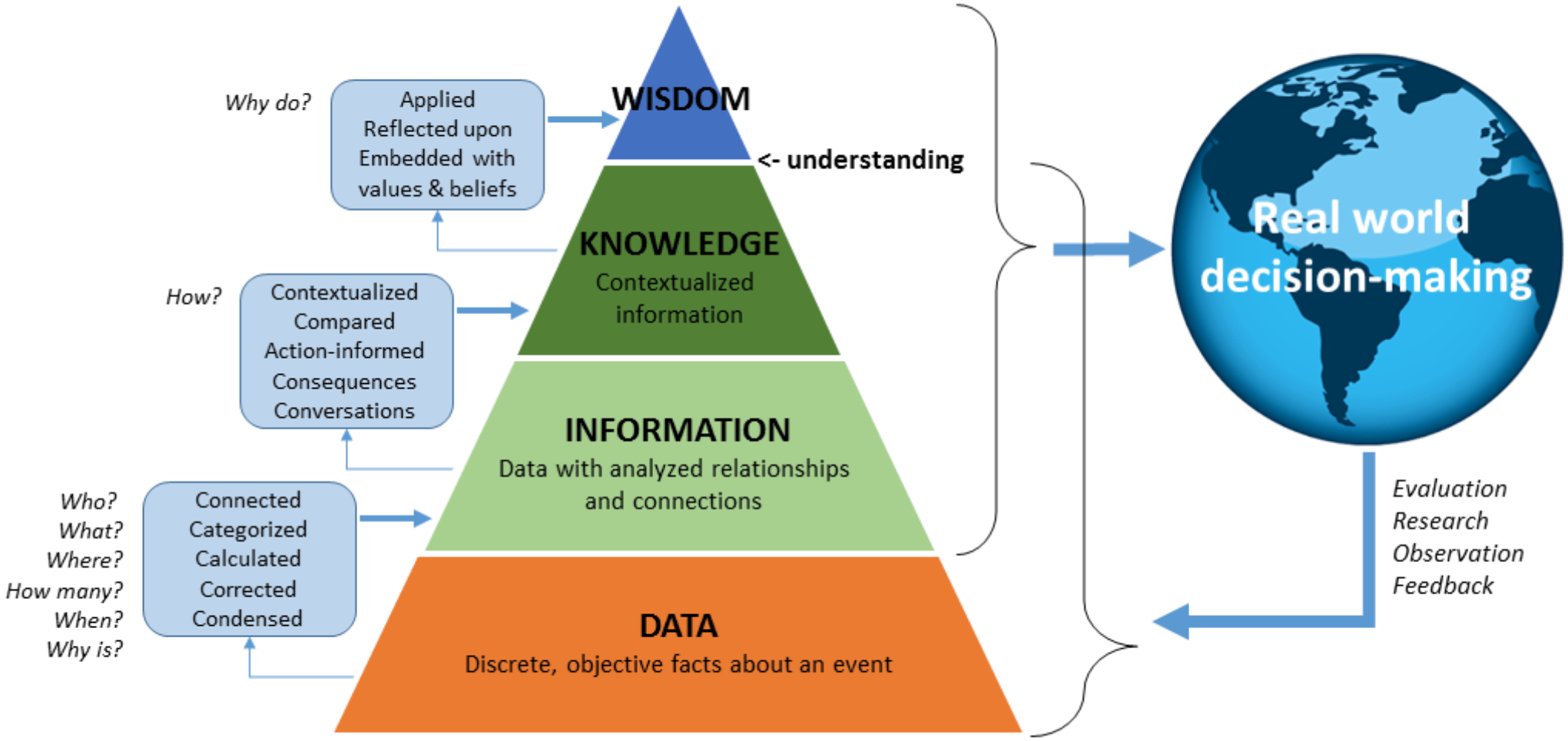
1. **Data science** is an inter-disciplinary field that uses scientific methods, processes, algorithms, Information Technology, Information Systems to extract knowledge and insights from many structured, unstructured and various forms of **data**

2. **Data science** is related with Data Mining, Statistical and Operational Research Data Analysis, Machine Learning and Big Data Technology and Analytics etc. analytics techniques for extracting meaningful insights from **data**

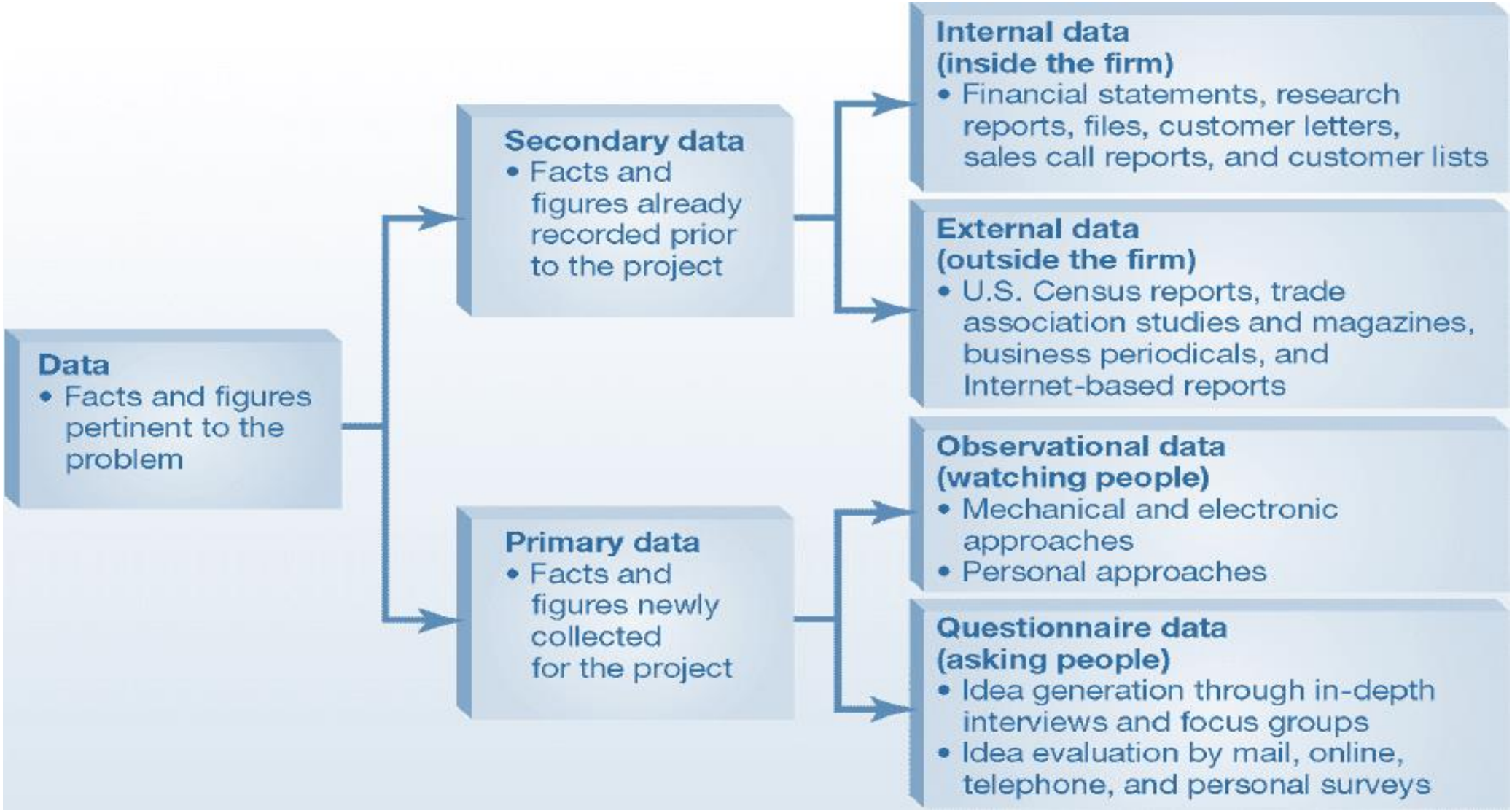
3. **Data science** is the field of study that combines **domain expertise, programming skills**, and **knowledge of mathematics and statistics** to extract meaningful insights from **data**.



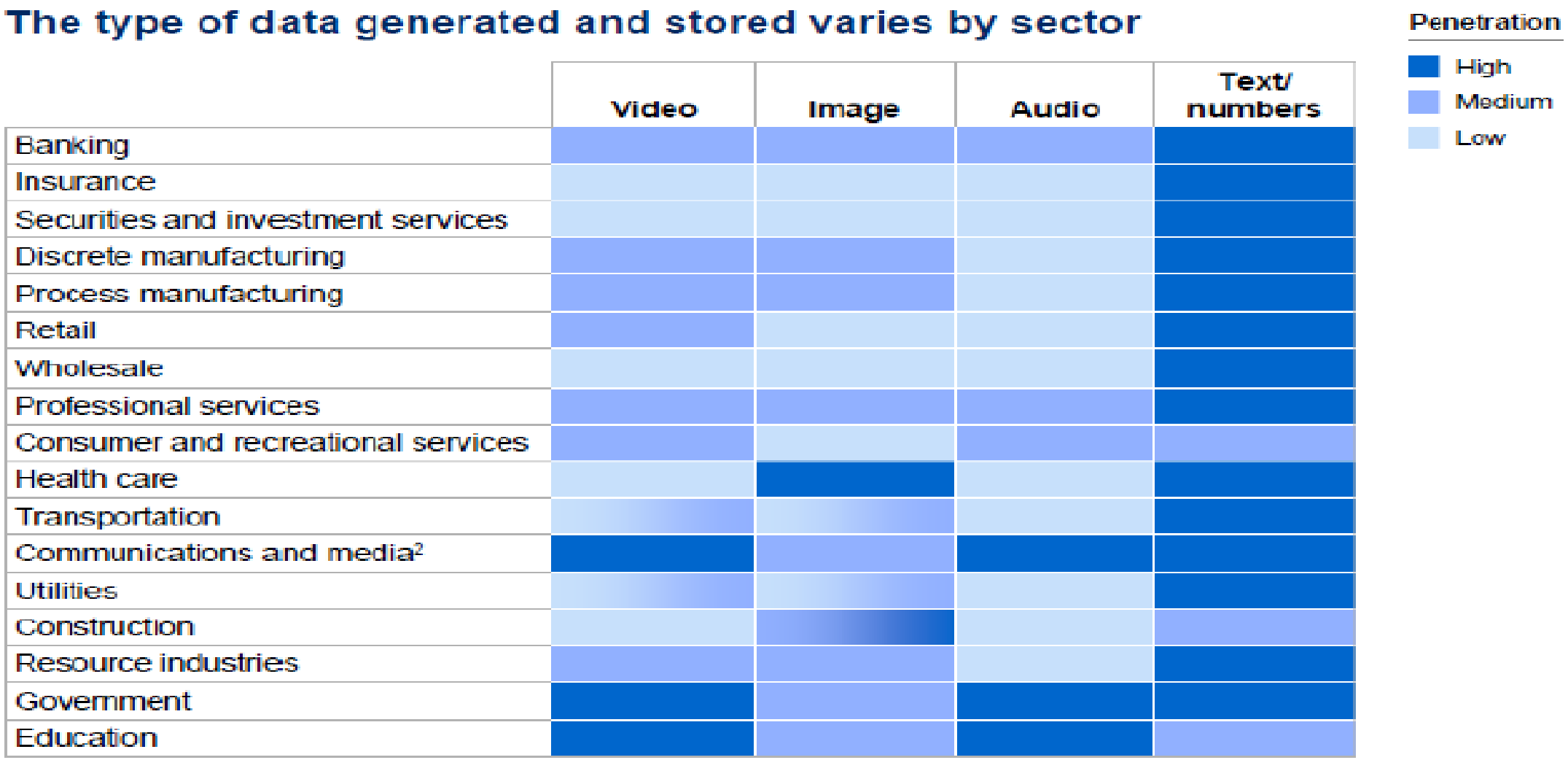
Stages of Data – DIKW Pyramid



Sources of Data



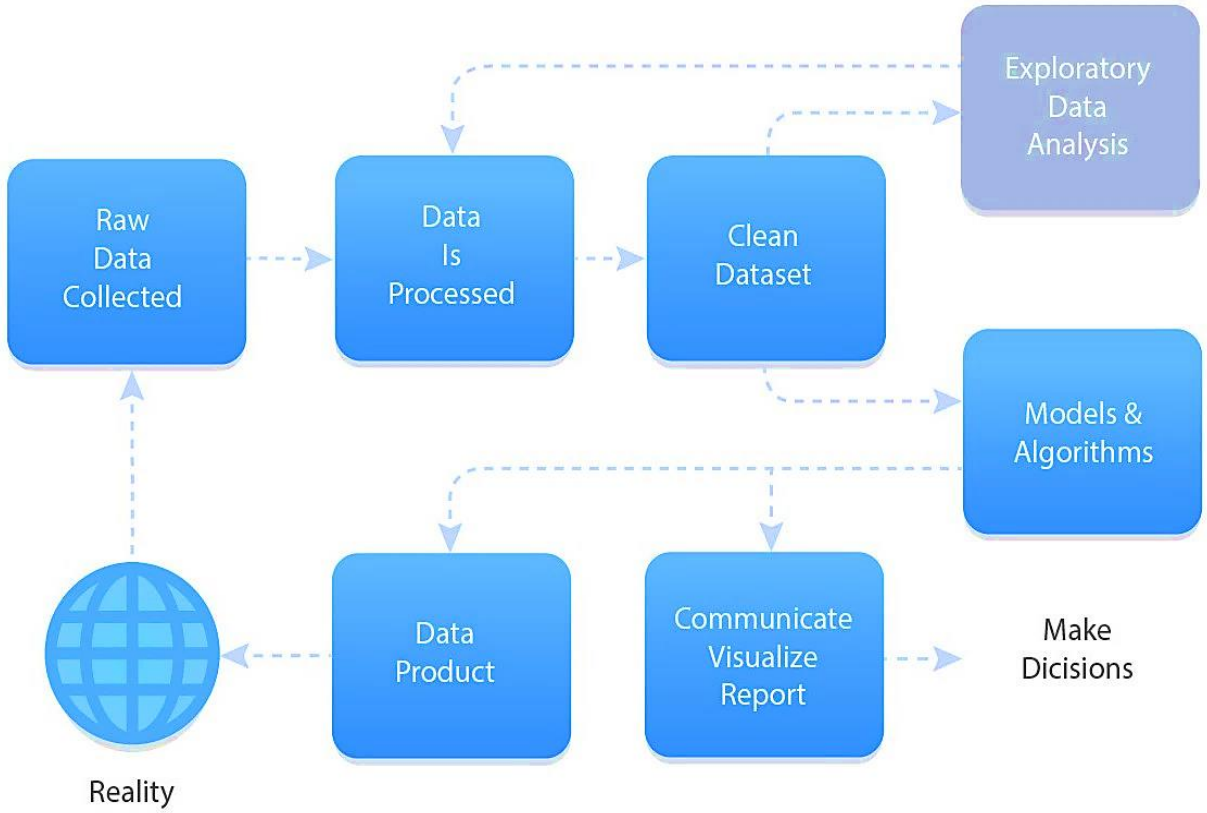
Sources of Data



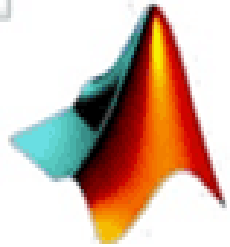
Understanding
Data Analysis vs Data Analytics

Data Analysis

- Data analysis is defined as a **process of cleaning, transforming, analysing and modeling data** to discover useful information for business decision-making.
- The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.



Data Analysis: Tools



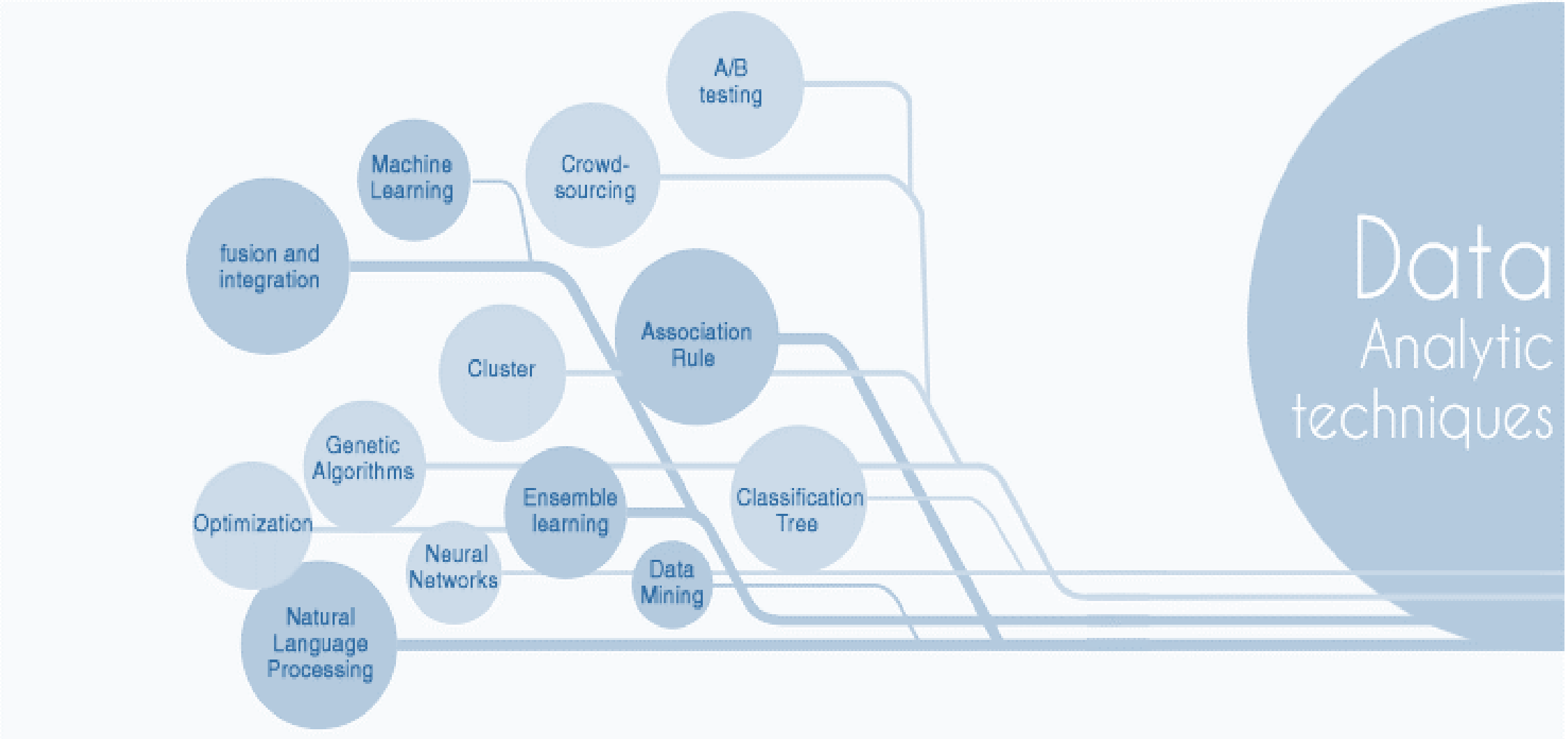
Data Analytics

- Data analytics is the **science of analyzing raw data** in order to extract hidden insights and relationship patterns from the data sets using various analytics techniques to take decisions.
- Many of the techniques and processes of data analytics are automated by using technology based tools and algorithms that work over raw data for human consumption.
- Data analytics techniques can **reveal trends and metrics** that can be used to optimize processes to increase the overall efficiency of a business performances and overall system.

Data Analytics: Tools

Popular Analytics Tools	Top Companies Using Them
Open Source	
R	Accenture, Cognizant, Google, Facebook, Citibank, Genpact, MuSigma, Fractal Analytics
Python	Alibaba, Google, Cognizant, TCS, Genpact, Gramener
Apache Spark	Uber, Pinterest, Ola, Facebook, Infosys, Wipro, Netflix
Apache Storm	Groupon, Twitter, Yahoo, Alibaba, Spotify, Flipboard
PIG & HIVE	Yahoo, Facebook, Twitter, Baidu, Uber, Flipkart
Commercial	
SAS	HSBC, Citibank, Google, Netflix, WNS, Genpact, Accenture, HDFC
Tableau	Barclays, Citibank, Gallup, Ogilvy, LA Times, Toyota, AOL, Dell, HP, Marico, Ashok Leyland
Excel	Almost every company known to mankind
Qlikview	TCS, Capgemini, Accenture, Cisco, Deloitte, Citibank
Splunk	Adobe, Nasdaq, Coca-Cola, Cognizant, Groupon, First Data, GoodData, ING, Intuit

Data Analytics: Techniques



Data Analysis vs Data Analytics

Analysis	Analytics
← Past	Future →
Explains : How? Why?	Explore potential future events
Quantitative = Data + How the scale decreased	Quantitative = Formula + Algorithm
Qualitative = Explains how & why story ends the way it did?	Qualitative = Intuition + Analysis

Data Analysis

≠

Data Analytics

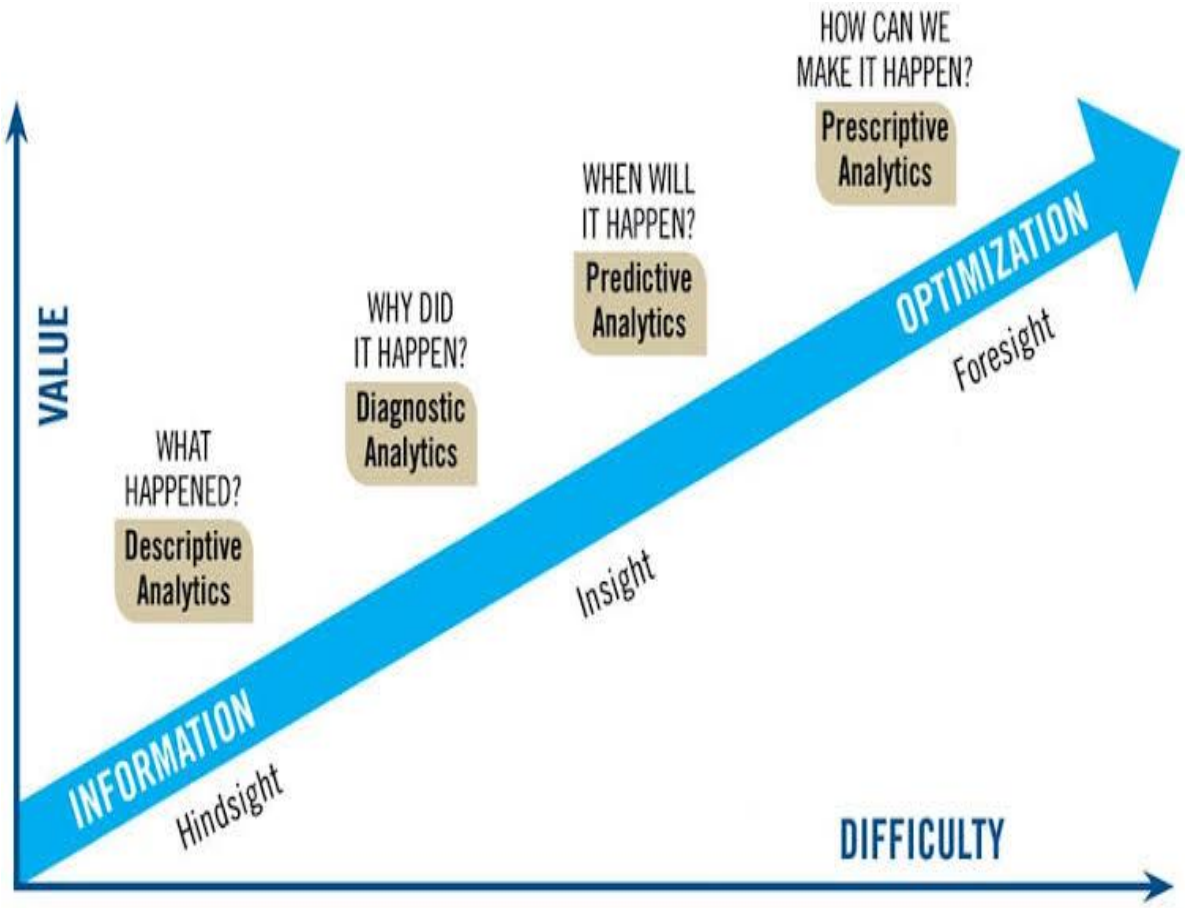
Data Analytics: Types

Descriptive: A set of techniques for reviewing and examining the data set(s) to understand the data and analyze business performance.

Diagnostic: A set of techniques for determine what has happened and why.

Predictive: A set of techniques that analyse current and historical data to determine what is most likely to (not) happen.

Prescriptive: A set of techniques for computationally developing and analyzing alternatives that can become courses of action – either tactical or strategic – that may discover the unexpected.



Business Intelligence

- **Business intelligence** (BI) refers to collection of applications, infrastructure, tools, models, and best practices that enable access to and analysis of information to improve and optimise business decisions and performance.

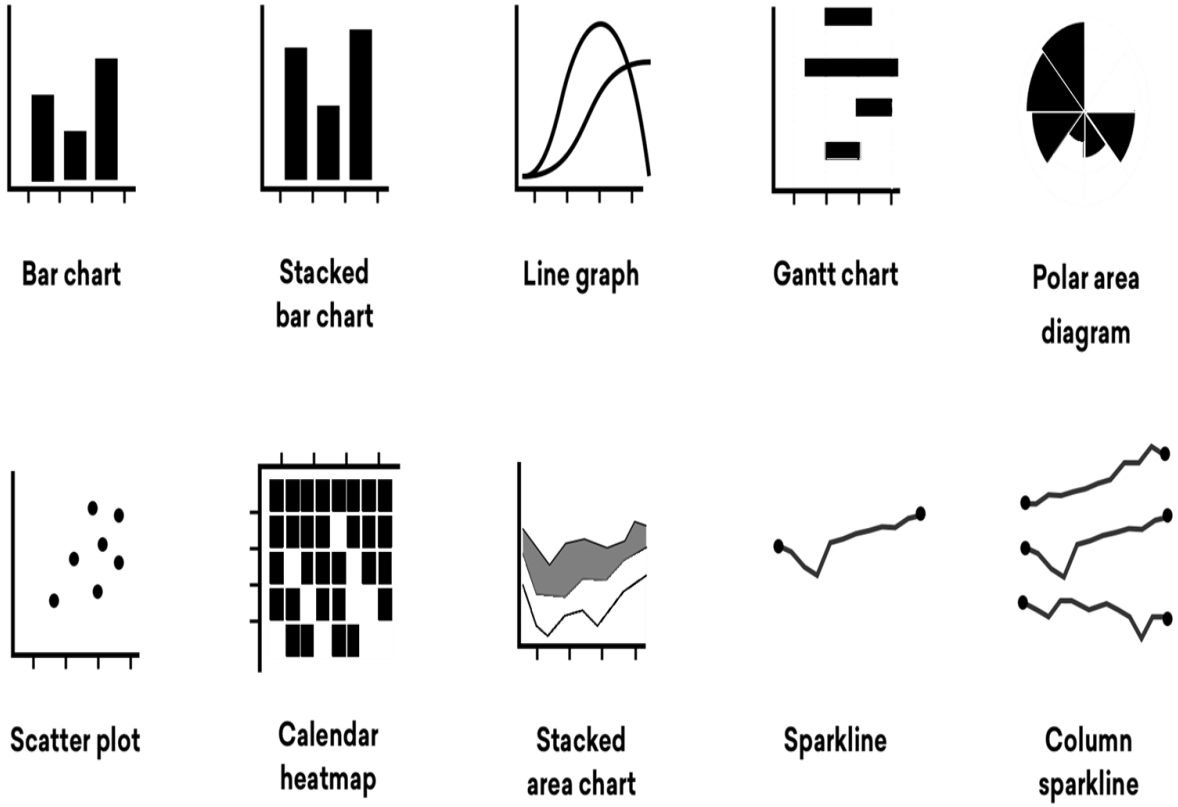
- Data analytics and BI assist both private and public sector in making all types of business decisions, including tactical, operational, and strategic.

- BI solutions are often related to **dashboards, online analytical processing, and data visualisation**, focusing on presenting the information and results of data analysis in management-friendly form.

	Answers to questions	Focus
Business intelligence	What happened? When? Who? How much/many? How long?	Reporting (KPIs, metrics) *Ad hoc* querying OLAP (cubes, slice & dice, drilling) Dashboards/scorecards Operational/real-time BI Automated monitoring/alerting
Advanced analytics	Why did it happen? Will it happen again? What will happen if we change *x*? What else does the data tell (what we never thought to ask)?	Statistical/quantitative analysis Data mining Predictive modelling/analytics Big data analytics Text analytics Multivariate testing

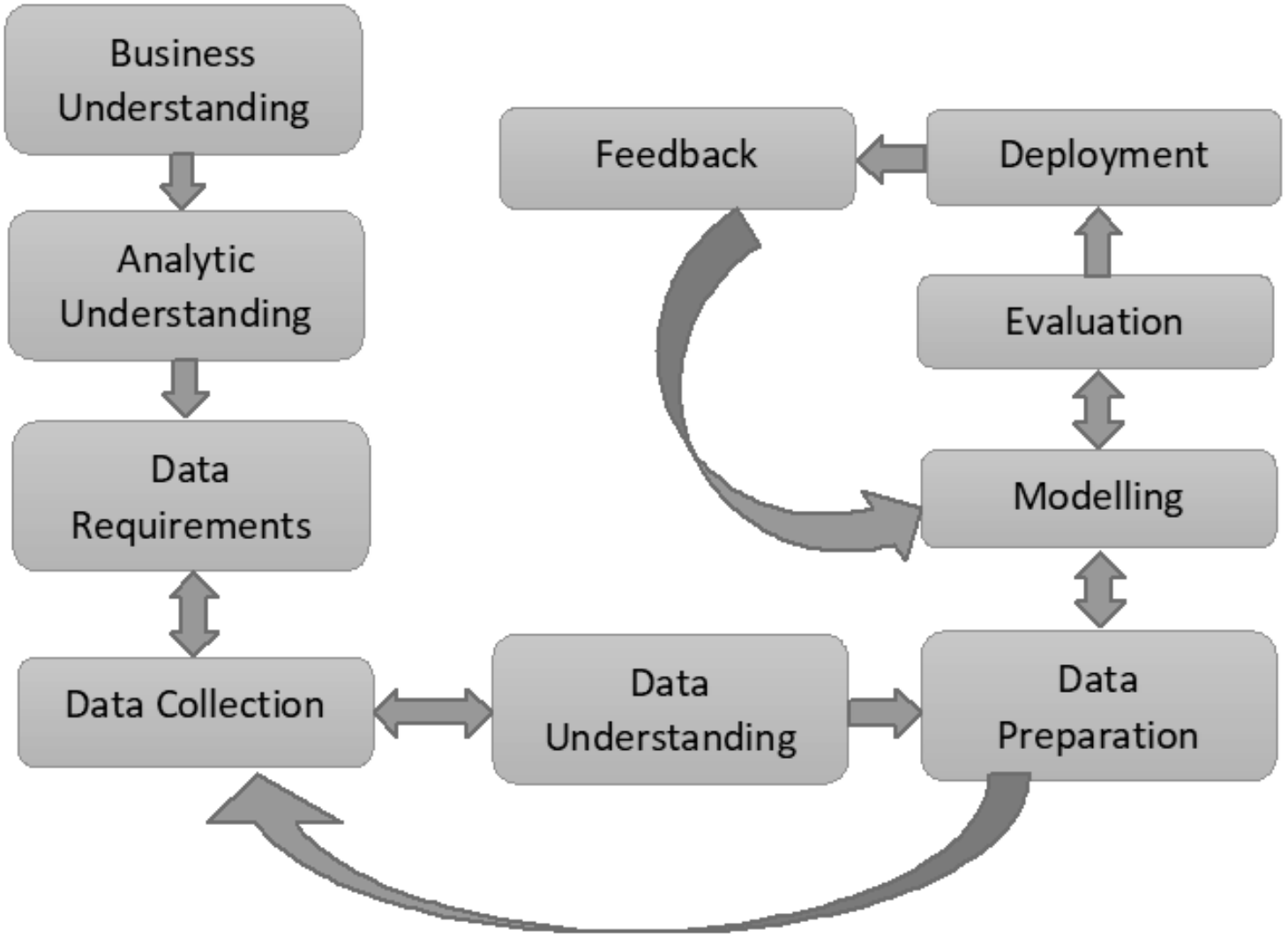
Data Visualization

- Data visualization is the **graphical representation of data**. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- Data visualization refers to techniques used to **communicate insights** from data through visual representation.
- Its main goal is to distil large datasets into visual graphics to allow for **easy understanding** of complex relationships within the data.
- It is often used interchangeably with terms such as information graphics, statistical graphics, and information visualization.



Data Visualization: Tools

Data Science: Life Cycle & Methodology



Data Mining

- Data mining is a **process of discovering patterns** in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

- Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use.

- Data mining allows us to:
  - Sift through all the chaotic and repetitive noise in the data.
  - Understand what is relevant and then make good use of that information to assess likely outcomes.
  - Accelerate the pace of making informed decisions.

Data Modeling

- Data modelling is a conceptual representation of Data objects, the associations between different data objects and the rules.

- Data modelling helps in the visual representation of data and enforces business rules, regulatory compliances, and government policies on the data.

- Data modelling occurs at three levels—physical, logical, and conceptual.

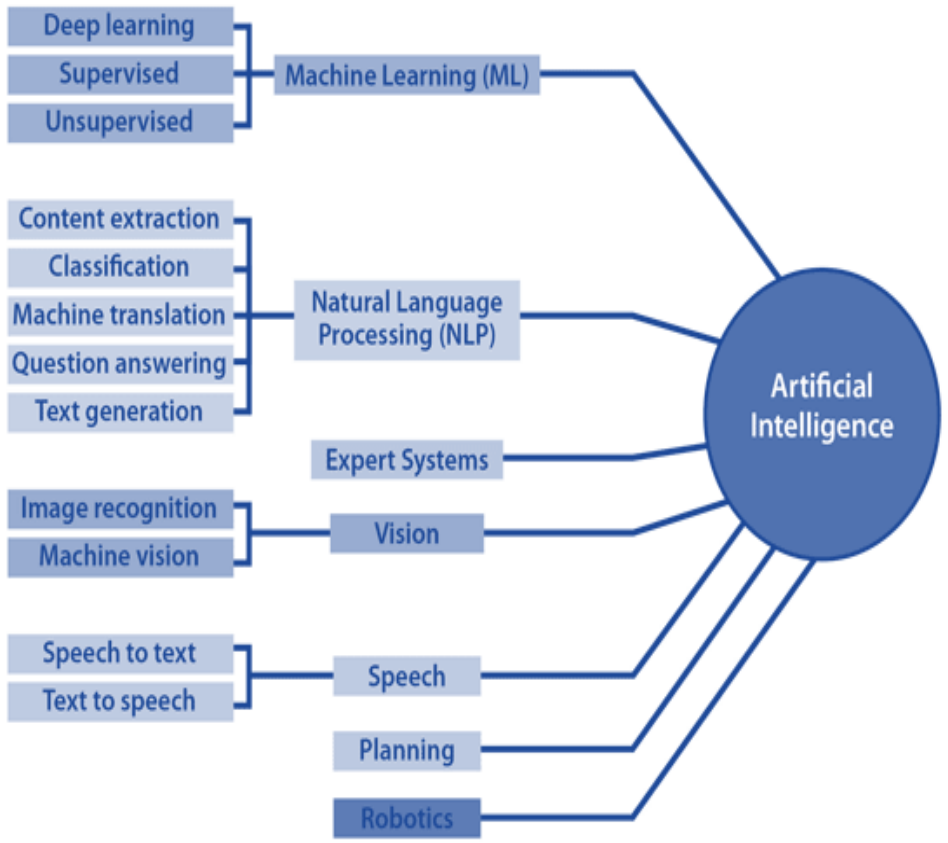
    a) A physical model is a schema or framework for how data is physically stored in a database.
    b) A conceptual model identifies the high-level, user view of data.
    c) A logical data model sits between the physical and conceptual levels and allows for the logical representation of data to be separate from its physical storage.

Artificial Intelligence

- Artificial intelligence (AI) refers to the **simulation of human intelligence in machines** that are programmed to think like humans and mimic their actions.

- The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

- Why is artificial intelligence important?
  - AI automates repetitive learning and discovery through data
  - AI analyzes more and deeper data using neural networks that have many hidden layers
  - AI achieves incredible accuracy through deep neural networks
  - AI adds intelligence to existing products
  - AI adapts through progressive learning algorithms to let the data do the programming



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.

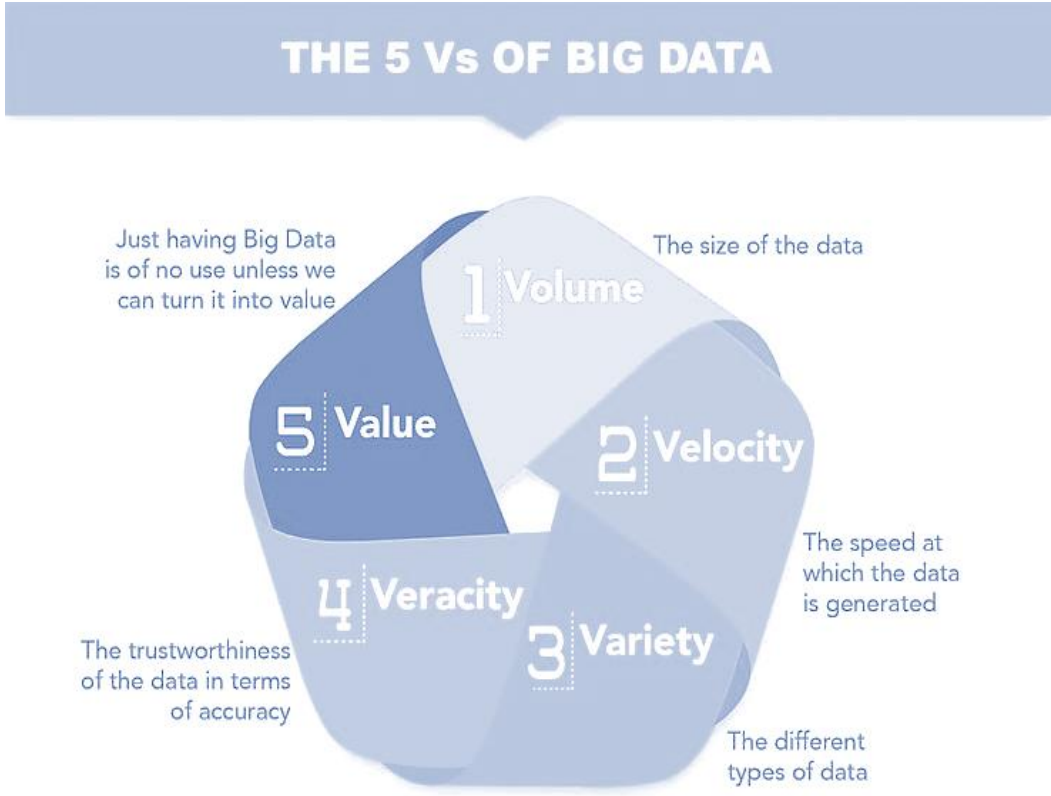
- Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

- While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development.

- Here are a few widely publicized examples of machine learning applications you may be familiar with:
  - The self-driving Tesla car - The essence of machine learning.
  - Online recommendation offers such as those from Amazon and Netflix - Machine learning applications for everyday life.
  - Knowing what customers are saying about you on Twitter - Machine learning combined with linguistic rule creation.
  - Fraud detection - One of the more obvious, important uses in our world today.

Big Data

- Big data is a field that treats ways to **analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex** to be dealt with by traditional data-processing application software.
- Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis.
- 90% of today's data has been created in just the last 2 years.
- Every day we create 2.5 quintillion bytes of data or enough to fill 10 million Blu-ray discs.
- 40 zettabytes (40 trillion gigabytes) of data will be created by 2020.
- Most companies in the US have over 100 terabytes (100,000 gigabytes) of data stored.



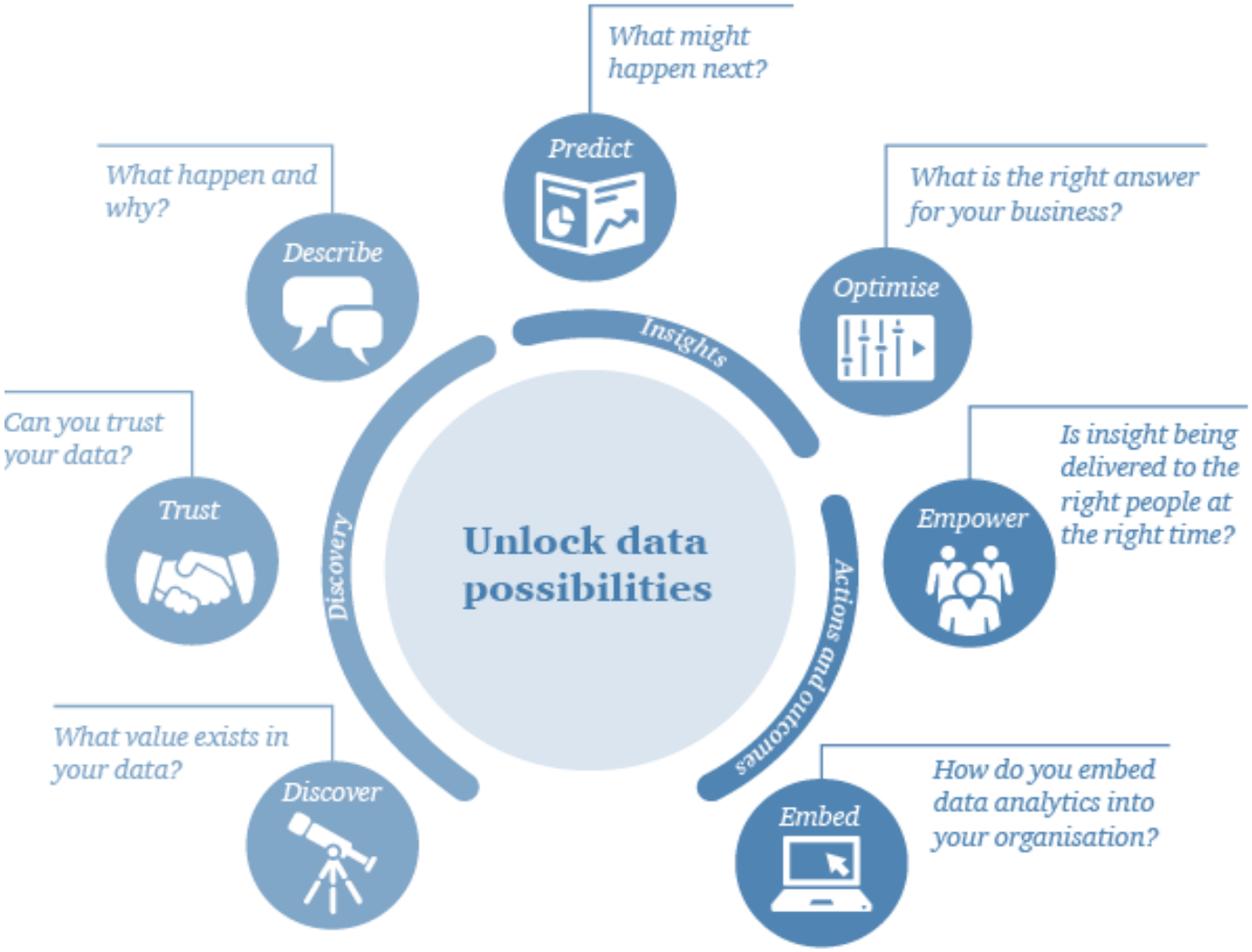
Big Data: Technologies

C4i4 LAB CENTRE FOR INDUSTRY 4.0

SECTION 2

Data Science Applications in Business

Data Science: Applications in Business



Business Intelligence for Making Smarter Decisions

- Traditional Business Intelligence was more descriptive and static in nature. However, with the addition of data science, it has transformed itself to become a more dynamic field.

- Data Science has rendered Business Intelligence to incorporate a wide range of business operations.

- With the massive increase in the volume of data, businesses need data scientists to analyze and derive meaningful insights from the data.
- The meaningful insights will help the data science companies to **analyze information at a large scale** and gain necessary decision-making strategies.

Making Better Products/Services

- Companies should be able to attract their customers towards products.
- They need to develop products that suit the requirements of customers and provide them with guaranteed satisfaction. Therefore, industries require data to develop their product in the best possible way.
- The process involves the **analysis of customer reviews** to find the best fit for the products. This analysis is carried out with the advanced analytical tools of Data Science.
- Businesses evolve with **innovation**. With the growth in data, industries are able to implement not only newer products but also various innovative strategies.
- For example – Airbnb uses data science to improve its services The data generated by the customers, is processed and analyzed. It is then used by Airbnb to address the requirements and offer premier facilities to its customers.

https://www.youtube.com/watch?v=S4RL6prqtGQ&ab_channel=Simplilearn

Leveraging Data to Make Business Better - Walmart

- Walmart handles a plethora of customer data. A staggering amount of about 2.5 petabytes of data is collected from the customers every hour. This data is unstructured that is utilized through Hadoop and NoSQL.

- Walmart is using data science to make **store checkouts more efficient**. With the help of **predictive analytics**, Walmart can analyze data and determine the best form of checkout for each store, that is, self-checkout and facilitated checkout.

- Walmart is using real-time analytics to **analyze the purchasing patterns of the customers**. This allows them to stock up on products that are in demand and also the products which will be in future demand based on several factors.

- Walmart is managing **supply chain and logistics** with the help of data science. It manages its inventory and analyzes the rate of its depletion, thereby taking the necessary steps to mitigate it through efficient logistics.

https://www.youtube.com/watch?v=42xErufN1e8&ab_channel=Walmart%26CoCo

Optimizing Package Routing - UPS

- UPS uses data science to optimize package transport from drop-off to delivery.
- Its latest platform for doing so, Network Planning Tools (NPT), incorporates machine-learning and AI to crack challenging logistics puzzles, such as **how packages should be rerouted around bad weather or service bottlenecks**.
- NPT lets engineers simulate a variety of workarounds and pick the best ones; AI also suggests routes on its own.
- According to a company forecast, the platform could save UPS $100 to $200 million by 2021.

https://www.youtube.com/watch?v=PSNIr1EMu_I&ab_channel=CNBC

SECTION 3

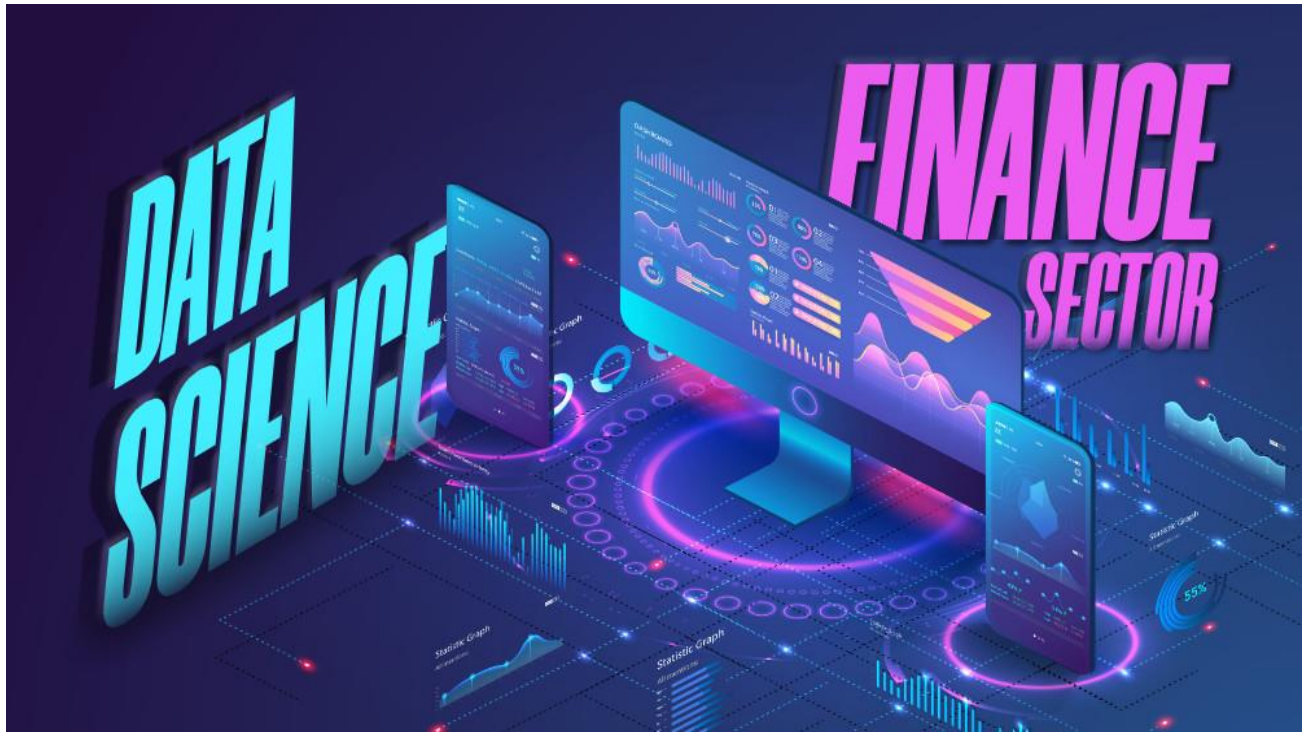
Data Science Applications in Financial Functions

Financial Functions – Data Science Applications in Finance Activities

1. Financial planning
2. Forecasting cash inflows and outflows
3. Raising funds
4. Allocation of funds
5. Effective use of funds
6. Financial control (budgetary and non-budgetary)

Financial Functions – Data Science Applications Finance Industries

1. Financial Risks Analytics
2. Consumer Analytics – Consumer Sentiments Analytics
3. Fraud Detection Analytics
4. Anomaly Detection Analytics
5. Personalized Self-service Analytics
6. Algorithmic Trading



Finding out the Financial Transactional Anomalies

- Transactions can be flagged due to **extreme values**, either low or high. These transactions can result from unintentional error or possible fraud.

- **Cluster analysis** may flag transactions not identified via other methodologies. While universal detection is never guaranteed, flagged transactions demonstrate suspicious characteristics worth investigating.

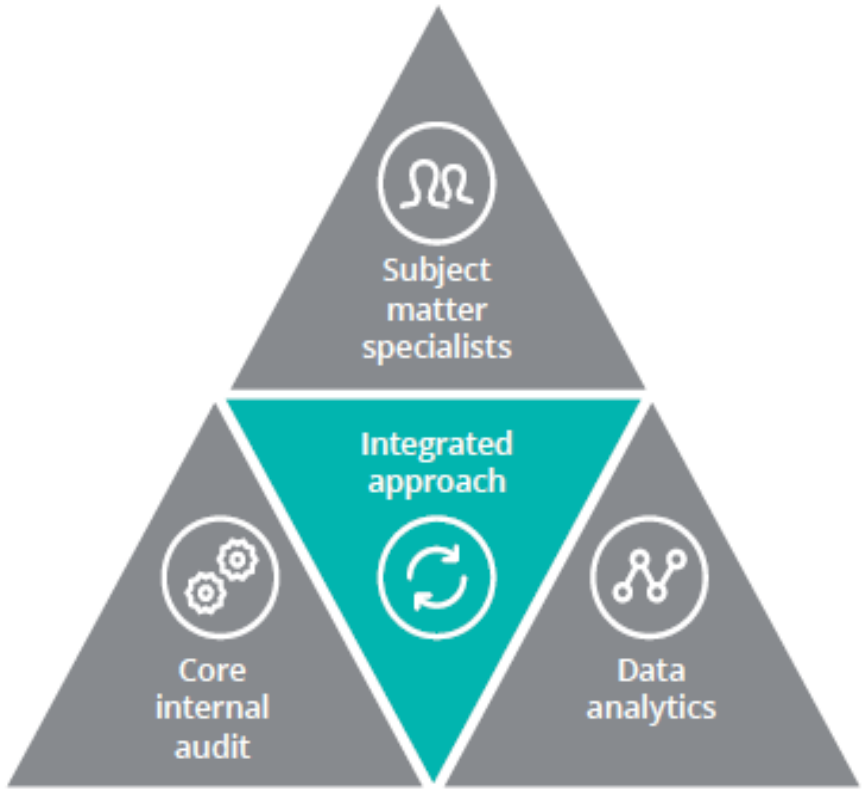
- Feedback from internal auditors can be very useful in improving the model. When flagged transactions end up being errors, fraud, or normal transactions, the validation of the results will provide inside knowledge which may be useful for the improvement of the model.

Financial Audit

- **Increased business understanding** through a more thorough analysis of a client's data and the use of visual output such as dashboard displays rather than text or numerical information allows auditors to and makes it easier to identify anomalies or outliers **better understand the trends and patterns of the business**

- **Increased efficiency** through the use of computer programmes to perform very fast processing of large volumes of data and provide analysis to auditors on which to base their conclusion, **saving time** within the audit and allowing **better focus on judgemental and risk areas**.

- Information obtained through data analytics can be shared with the client, **adding value to the audit** and providing a real benefit to management



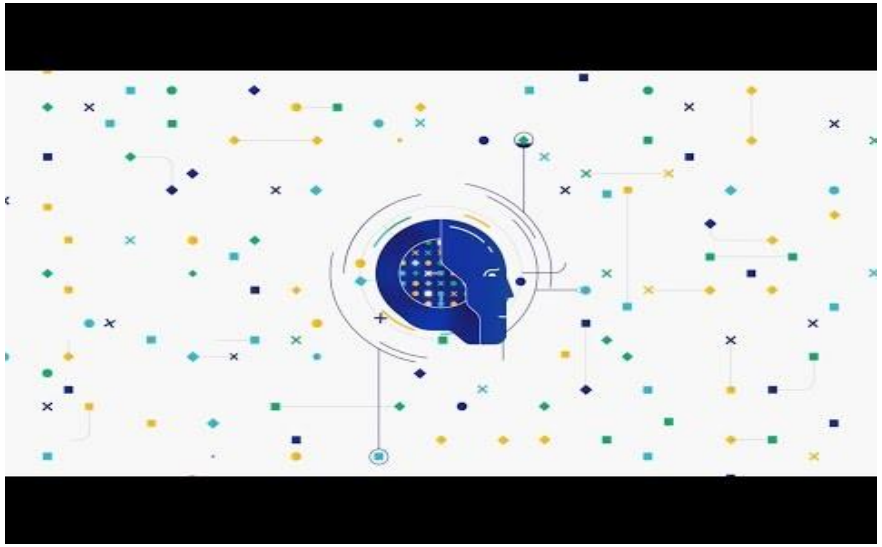
Process Automation

- Process automation is one of the most common applications of machine learning in finance.
- The technology allows to **replace manual work, automate repetitive tasks, and increase productivity**.
- As a result, machine learning enables companies to **optimize costs, improve customer experiences, and scale up services**.
- JPMorgan Chase launched a Contract Intelligence (COiN) platform that leverages **Natural Language Processing**.
- Manual review of **12,000 annual commercial credit agreements would typically take up around 360,000 labor hours**. Whereas, machine learning allows to review the same number of contracts in a **just a few hours.**

https://www.youtube.com/watch?v=dTbVpvfZYHg&ab_channel=jpmorgan

Financial Fraud Detection

- Security threats in finance are increasing along with the growing number of transaction, users, and third-party integrations. And machine learning algorithms are excellent at detecting frauds.
- Banks can use this technology to monitor thousands of transaction parameters for every account in **real time**.
- The algorithm examines each action a cardholder takes and assesses if an attempted activity is characteristic of that particular user. Such model spots **fraudulent behavior with high precision**.
- Financial monitoring is another security use case for machine learning in finance. Data scientists can train the system to **detect a large number of micropayments and flag such money laundering techniques as smurfing**.



https://www.youtube.com/watch?v=96k0sncyoXA&ab_channel=VisaCommunication

Data Science Applications Banking Industry

C4i4 LAB CENTRE FOR INDUSTRY 4.0

Kirloskar Enriching Lives

Better customer targeting and ensuring growth

By understanding clients more fully, and by using analytics of their transactions and trading activities, banks can be sure that they are delivering the best services for what their customers need, resulting in higher levels of retention and acquisition.

Enhancing risk assessment

As banks will be able to assess the risk profiles of their credit applicants in much greater detail, they will also be able to improve their credit assessments. Data analytics will advance the early-warning systems and data collection as well. All of these features will help banks to lower their risk costs, and to become aware of fraud more quickly.

Improving productivity and decision-making

With the advantage of advanced analytics, banks will be able to provide faster and more accurate responses to regulatory requests. Data will also enable better decisions for everyday activities: for example, better placement of ATMs and counters, and how much cash is required at each ATM.

More business opportunities

By collecting data from customers, data analytics will enable banks to develop new business models and new sources of income: for example, by sharing data with other companies, when the customer has agreed to this beforehand.

Digital banks – internet-based banks

In today's society, most people conduct their transactions online, through their smartphones or their computers. By analyzing real-time data, we can advance the customer experience and understand our customers much better.

Data Science Applications in Insurance Industry

Better product design and marketing	*More accurate risk assessment, underwriting, and pricing processes*	*Stronger commitment to helping customers*	*Better claims management*
Insurers can take advantage of new sources of data to better target intended customers with specific – and potentially more suitable – products, making it possible to design offers based on what people need in the future, and to combine these with improvements in technology and regulation.	Data analytics allow insurers to assess the risk profiles of their applicants in much greater detail, which should mean better-informed underwriting decisions as well as premium calculations that will be more accurate in their alignment with the corresponding levels of risk.	There is potential to reward policyholders with lower premiums, if their risk profile improves: this can be indicated by the number of claims, by smartphone apps that can monitor lifestyles, or by telematics devices. The reward of a lower premium could also encourage policyholders to improve their lifestyle.	Data analytics can be used to prioritize claims, and to set straightforward claims apart from complex cases. This can result in faster settlements for the straightforward claims, and more attention for the complex cases.

Conclusions

- The rapid development of scientific and Data Science Technologies over the past decade has brought not only new and sophisticated analytical tools into financial and other industries, but also introduced the power of data science application in everyday strategic and operational management.

- Implementation of Data Science Analytics Applications capabilities will help the Finance Organizations for business growth, finding out new business opportunities, increasing the customer centric business efficiency and also for a risk free/minimized sustainable business continuity.

- Data analytics and science developments have been particularly valuable to financial organizations that heavily depend on financial information in their decision making processes.

- Technology and data science are both forcing and enabling the financial industry to respond to transformative demands and adapt to rapidly changing market conditions in order to survive and thrive in highly competitive global environment.

- The challenges previously associated with data management and analytics should be turned into opportunities, with the aid of higher awareness, best practices, and technology made available in the data space.

Q & A

Thank you

For any queries, contact:

Prof Pankaj Roy Gupta: email-id : prgupta@kiams.ac.in

Mr Sanket Waykar email-id: sanket.waykar@c4i4.org

C4i4 LAB CENTRE FOR INDUSTRY 4.0®

www.c4i4.org