**Analyse the nature, contents and complexity of the knowledge base in engineering**

Data science is the field of applying advanced analytics techniques and scientific principles to extract valuable information from data for business decision-making, strategic planning and other uses. It's increasingly critical to businesses: The insights that data science generates help organizations increase operational efficiency, identify new business opportunities and improve marketing and sales programs, among other benefits. Ultimately, they can lead to competitive advantages over business rivals.

Data science incorporates various disciplines -- for example, data engineering, data preparation, data mining, predictive analytics, machine learning and data visualization, as well as statistics, mathematics and software programming. It's primarily done by skilled data scientists, although lower-level data analysts may also be involved. In addition, many organizations now rely partly on citizen data scientists, a group that can include business intelligence (BI) professionals, business analysts, data-savvy business users, data engineers and other workers who don't have a formal data science background.

Data science plays an important role in virtually all aspects of business operations and strategies. For example, it provides information about customers that helps companies create stronger marketing campaigns and targeted advertising to increase product sales. It aids in managing financial risks, detecting fraudulent transactions and preventing equipment breakdowns in manufacturing plants and other industrial settings. It helps block cyber attacks and other security threats in IT systems.

Data science is also vital in areas beyond regular business operations. In healthcare, its uses include diagnosis of medical conditions, image analysis, treatment planning and medical research. Academic institutions use data science to monitor student performance and improve their marketing to prospective students. Sports teams analyze player performance and plan game strategies via data science. Government agencies and public policy organizations are also big users.

To facilitate evidence-based decision-making, organizations need efficient methods to process large volumes of assorted data into meaningful comprehensions (Gandomi & Haider, 2015). The potentials of using BD are endless but restricted by the availability of technologies, tools and skills available for BDA. According to Labrinidis and Jagadish (2012), BDA refers to methods used to examine and attain intellect from the large datasets. Thus, BDA can be regarded as a sub-process in the whole process of insight extraction from BD. It is certain that for BD to realise its objectives and progress services in business environment, it requires the correct tools and approaches to be analyzed and classified effectively and proficiently. The potential value of BD is solved simply when leveraged to the drive decision-making process. Extant research studies have demonstrated that substantial value and competitive advantage can be attained by businesses from taking effective decisions based on data But, BDA is more perplexing than merely tracing, classifying, comprehending, and quoting data. Davenport and Dyché (2013) emphasize that large organizations regularly gather BD and exploit analytics for support in decision-making as part of their usual procedures, and SMEs are the ones presently struggling to enhance top management decisions while adding more data for the analysis process. Aligning the people, technology, and organizational resources to become a data-driven company is problematic (Weill & Ross, 2009). Given BD can enhance the decision-making and increase organizational output; this is possible when a selection of analytical methods is used to extract sense from the data, such as:

1. descriptive analytics scrutinizes data and information to define the current state of a business situation in a way that developments, patterns and exceptions become evident, in the form of producing standard reports, ad hoc reports, and alerts.
2. inquisitive analytics is about probing data to certify/reject business propositions, for example, analytical drill downs into data, statistical analysis, factor analysis.
3. predictive analytics is concerned with forecasting and statistical modelling to determine the future possibilities.
4. prescriptive analytics is about optimization and randomized testing to assess how businesses enhance their service levels while decreasing the expenses.
5. pre-emptive analytics is about having the capacity to take precautionary actions on events that may undesirably influence the organizational performance, for example, identifying the possible perils and recommending mitigating strategies far ahead in time.

Advocates assert that these types of analytical methods support in improved decision-making and organizational performance by making everything more translucent and quantifiable, while further uncovering inconsistencies as well as potential concerns and opportunities.

Data science applications have varying levels of risk. For example, recommender systems that suggest purchases within an online shopping platform or select advertisements for website visitors are relatively low risk. Although provider sales may be affected if undesirable products are recommended and users may be dissatisfied with their purchases, the overall impact of poor retail recommender systems to individuals and society is generally low. Still, the recommendations can influence the behaviour of large segments of a population and are often coupled with a just-in-time supply chain, which aims to forecast consumer demand given available data and optimize production and shipping of goods. In this case, the systems can have substantial impact, especially if they result in a shortage of necessary items, such as food and medicine, owing to natural disaster or unanticipated interactions with other external factors. But increasingly, as similar data-driven algorithms are used to recommend sentencing or release of criminals, guide testing or treatment of patients, plan urban development, draw political boundaries, allocate funds, and inform other critical public policy decisions, impacts on individuals and society can be profound.

While new volumes and types of information can make analyses more accurate than past methods that relied on sparse surveys with lower than desired survey frequency, response rates, and sample sizes, they still have limitations. Weaknesses in data quality and data analysis might have a wide range of negative policy effects: problems might be misunderstood in their causes and scale; a program that a family depends on might get insufficient funding; or a policy might be enacted that has unintended consequences for large segments of the population.

Thus, it will be important that data are collected and analysed appropriately and that there are clear principles guiding the use of data for human good. Furthermore, the complexity of the analyses and the increasing dependency on data across all fields of human endeavour will drive demand for “smarter” tools and best practices for data science that minimize mistakes in interpretation.

An effective data science workflow involves formulating good questions, considering whether available data are appropriate for addressing a problem, choosing from a set of different tools, undertaking analyses in a reproducible manner, assessing analytic methods, drawing appropriate conclusions, and communicating results. Students need practice applying a unified approach to problem solving with data. Such an integrated approach needs to be introduced in their first courses and remain a consistent theme in subsequent courses. Students need to see that data science is not simply a collection of varied tools (or methods), but rather a general approach to problem solving. Many of the emergent data science programs at every academic level encourage students to assume that they will benefit from continuing professional education throughout their careers. All require that graduates have the capability to identify problems to be solved with data, determine and implement solutions, assess results, and communicate results and findings.

There are many other types of data scientists today, and their roles will continue to change and expand in the future. Beyond the differences among them, there is considerable variance in the lower-order and higher-order knowledge and skills that some data science jobs require. There are also many commonalities among the varied types of data scientists. All data scientists need to learn how to tackle questions with real data. It is insufficient for them to be handed a “canned” data set and be told to analyze it using the methods that they are studying. Such an approach will not necessarily prepare them to solve more realistic and complex problems taken out of context, especially those involving large, unstructured data. Instead, they need repeated practice with the entire cycle beginning with ill-posed questions and “messy” data.