NEW HORIZON COLLEGE OF ENGINEERING

BUSINESS AND INFORMATION TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
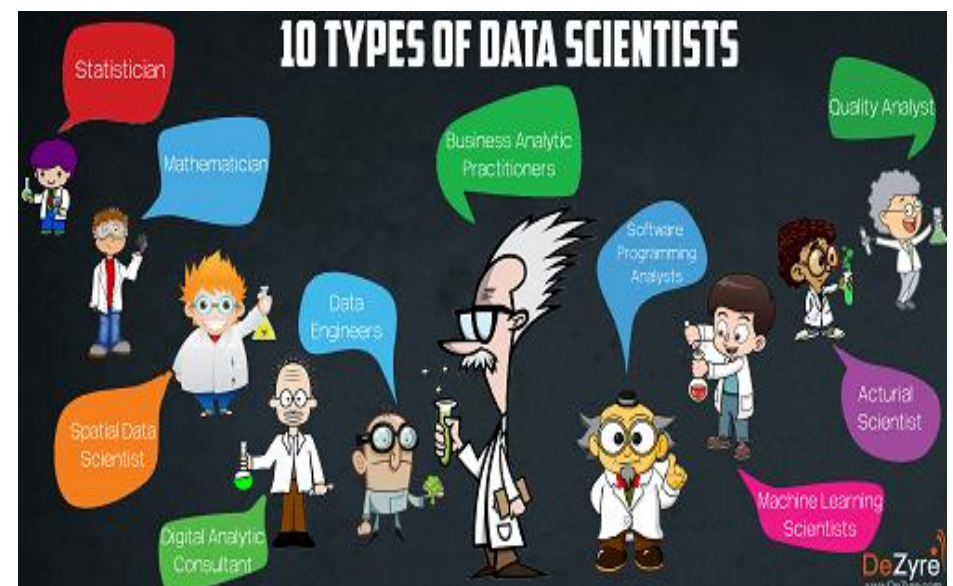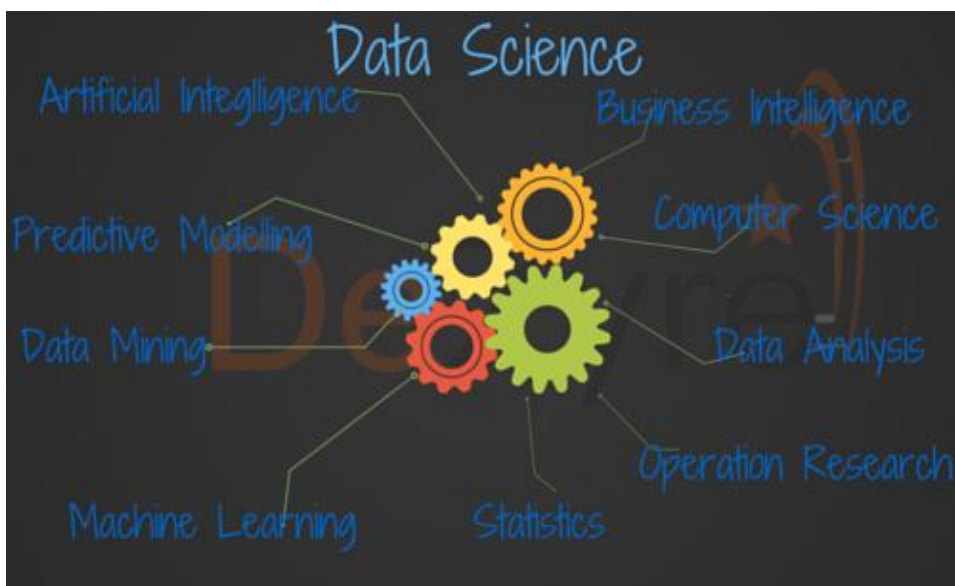
DATA SCIENCE

Data science, also known as data-driven science, is an interdisciplinary field about scientific methods,processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to Knowledge Discovery in Databases (KDD).Data science is a "concept to unify statistics, data analysis and their related methods" in order to "understand and analyze actual phenomena" with data.[j] It employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, information science, and computer science, in particular from the subdomains of machine learning, classification, cluster analysis, data mining, databases, and visualization


Data scientists use their data and analytical ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of Datasets create datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate the data insights/findings.

"Data Scientist" has become a popular occupation with Harvard Business Review dubbing it "The Sexiest Job of the 21st Century" and McKinsey & Company projecting a global excess demand of 1.5 million new data scientists. Universities are offering masters courses in data science. Shorter private bootcamps are also offering data science certificates including student-paid programs like General Assembly to employer-paid programs like The Data Incubator.

MACHINE LEARNING
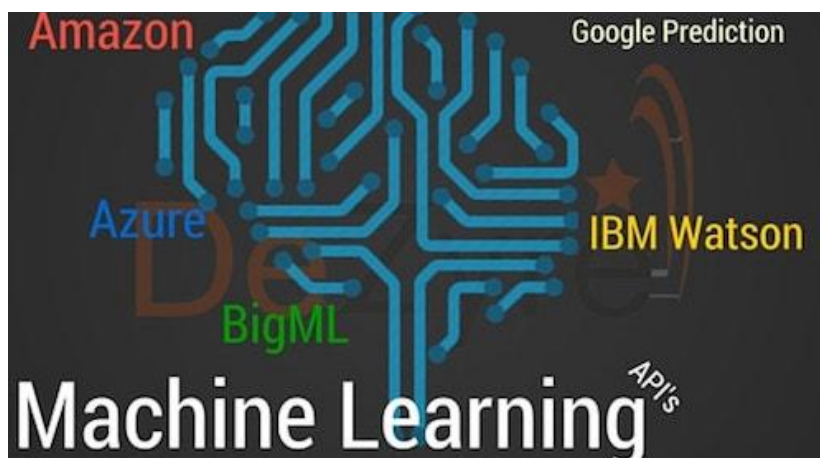
Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data. Both systems search through data to look for patterns. Machine learning uses the data to detect patterns in data and adjust program actions accordingly.

Facebook's News Feed uses machine learning to personalize each member's feed. If a member frequently stops scrolling in order to read or "like" a particular friend's posts, the News Feed will start to show more of that friend's activity earlier in the feed. Behind the scenes, the software is simply using statistical analysis to identify patterns in the user's data and use to patterns to populate the News Feed .

Machine learning is closely related to computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioural profiles for various entities and then used to find meaningful anomalies.

DEEP LEARNING

Deep learning is a class of machine learning algorithms that

- use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised).
- are based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation.
- are part of the broader machine learning field of learning representations of data.
- learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concept

Deep learning algorithms are based on distributed representations. The underlying assumption behind distributed representations is that observed data are generated by the interactions of factors organized in layers. Deep learning adds the assumption that these layers of factors correspond to levels of abstraction or composition. Varying numbers of layers and layer sizes can be used to provide different amounts of abstraction.

Many deep learning algorithms are applied to unsupervised learning tasks. This is an important benefit because unlabeled data are usually more abundant than labeled data. Examples of deep structures that can be trained in an unsupervised manner are neural history compressors and deep belief networks.
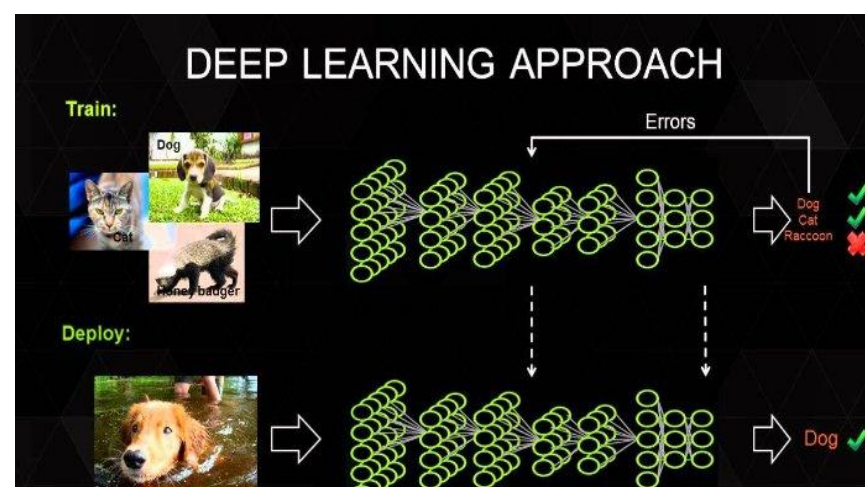
ADVISORY COMMITTEE	STUDENT EDITORS
Dr. Prashanth C S R	Abhishek YSS
Prof & Head, Dept. of CSE	Amar Bahadur Thapa
Prof Beena B M	
Dept of CSE	