



CONL  
722

Big Data Challenges and Opportunities

### 1.2.2: Machine Learning

Machine learning is a subject in itself, however, here we will have a very brief overview of machine learning in the context of big data analytics.

### Why Big Data ?

Provides the ability to access large volume of data to gain critical and useful insights

Learning process is machine managed with minimum human intervention, hence the analysis is simpler and error free .

I hope you remember that the first session explained that big data technologies are widely used because it provides the ability to access a large volume of data to gain critical and useful insights.

The learning process is machine managed with minimum human intervention, hence the analysis is simpler and error-free.

# Machine Learning

---

Machines learning from experience and act accordingly rather than mechanically responding to a predefined program.

Programs will be collect and use the data

Automatic process of pattern recognition

Systems that can outperform human being in handling complex tasks.

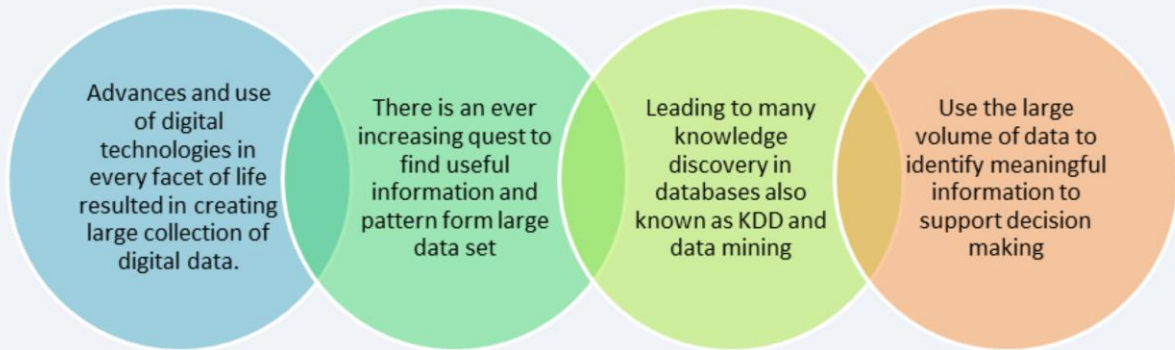
Machine learning can be defined as providing the ability to machines (computers) to learn from experience and act accordingly rather than mechanically responding to a predefined program.

Programs will be written to collect and use the data supporting an automatic process of pattern recognition.

Machine learning is aimed to build systems that can outperform human being in handling complex tasks.

# Learning from Data

---



As you have already seen, the advances in the digital technologies and use of digital technologies in every facet of life resulted in creating a large collection of digital data.

There is an ever-increasing quest to find useful information and pattern form large data set leading to many knowledge discoveries in databases also known as KDD and data mining.

KDD use the large volume of data to identify meaningful information to support decision making.

# Learning from Data

---

## Pattern identification

- Appropriate use of statistical tools from data sampling, modelling, analysis etc.

## Statistics theory in machine learning

- Shifted from rule-based expert systems to data driven methodology

## Train the machine to learn automatically

- Improve the results as it gets more data

## Machines knowledge

- Directly from data
- Solve the problems

## Turning the data into information

Pattern identification relies on the appropriate use of statistical tools from data sampling, modelling, analysis etc.

Statistics theory in machine learning has shifted from rule-based expert systems to a data-driven methodology.

Train the machine to learn automatically to improve the results as it gets more data.

Machines acquiring knowledge directly from data to solve the problems  
And turning the data into information

# Machine learning

---

## Machine learning

Essential component of Big Data analytics

Convert the large data sets into information

## Big Data framework support machine learning

Parallel processing

In-memory database systems

Map-reduce processing platforms like Hadoop and GFS

Bulk Synchronous parallel systems such as Apache HAMA and Giraph.

Machine learning an essential component of Big Data analytics.

Without machine learning, it is impossible to convert the large data sets into meaningful information.

Various components in the Big Data framework are designed to support machine learning.

Big data technologies use Parallel processing, In-memory database systems, Map-reduce processing platforms like Hadoop and GFS, Bulk Synchronous parallel systems such as Apache HAMA and Giraph.

They all support Big Data analysis by processing a large volume of data with minimum processing time.

But without the appropriate machine learning process in place, this would be incomplete.

## Next

---

### Big Data framework



Hadoop, HDFS, Mapreduce, Spark etc.

The next video will provide an overview of the data storage using distributed file systems and big data platform using Hadoop.