Big data – is a movement to find a new better way to manage a large amount of web-generated data and derive insight from it. It is often too complex to process using regular database management tools.

Main characteristics of Big Data

- Volume - refers to the massive scale of big data, often measured in exabytes.
- Velocity – refers to the fast pace at which big data is generated and processed.
- Variety – describes the different types and formats of data, such as structured, semi-structured, and unstructured.
- Veracity – refers to the accuracy and trustworthiness of big data.
- Value – represents the potential of economic or business value that can be extracted from big data.

Some common examples of large data sets

- Web logs.
- Call records.
- Medical records.
- Military surveillance.
- Photography archives.
- Video archives.
- Large-scale e-commerce.

Big Data Analysis Techniques

Associative rule – Is an analysis technique adopted to find patterns in data through correlation between variables in large databases.

Classification tree analysis – this is a machine learning algorithm that adopts a structured mapping of binary decisions which lead to a decision about a class of an object.

Some common applications of classification tree analysis.

- Automatically assign documents to categories.
- Categorize organisms into groupings.
- Develop profile of students who take online courses.

Genetic algorithms – is inspired by mutation, inheritance, and natural selection. Which are used to “evolve” useful solutions to problems that require optimization.

Common applications of genetic algorithms.

- Scheduling of doctors for hospital emergency room.
- Developing combinations of optimal materials and engineering practices required to develop fuel-efficient cars.

Machine Learning – refers to the ability of giving computers the ability to learn from training labelled dataset to predict the future without explicitly being programmed.

- ‘distinguishing between spam and non-spam email messages
- learning user preferences and make recommendations based on this information
- determining the best content for engaging prospective customers
- determining the probability of winning a case, and setting legal billing rates’

Regression analysis is a powerful statistical method that investigates the relationship between two or more variables. Typically, it examines the influence of one or more independent variables on a dependent variable, like weight, speed or age.

- levels of customer satisfaction affect customer loyalty
- number of supports calls received may be influenced by the weather forecast given the previous day
- neighborhood and size affect the listing price of houses

Sentiment analysis

Sentiment Analysis is a type of Natural Language Processing (NLP) technique that automates the process of understanding an opinion about a given subject from written or spoken language. It helps researchers determine the sentiments of speakers or writers.

Sentiment analysis is being used to help:

- improve service at a hotel chain by analyzing guest comments
- customize incentives and services to address what customers are really asking for
- determine what consumers really think based on opinions from social media.

Social network analysis

Social network analysis maps and measures the relationships and flows between people, groups, organizations, computers, URLs, and other connected information or knowledge entities. The nodes in the network represent the people and groups while the links identify the relationships or flows between the nodes.

Some examples of the application of social network analysis include:

- understanding how people from different ethnic groups form ties with outsiders
- finding the importance of a particular individual within a group
- determining the social structure of a customer base.

Data Mining

Data mining extracts patterns from large data sets by combining methods from statistics and machine learning, within database management. It is also referred to as the process of finding anomalies, patterns and correlations within large data sets to predict outcomes.

Natural Language Processing (NLP)

NLP is as a subspecialty of computer science, artificial intelligence, and linguistics, which uses algorithms to analyze human (natural) language.

IoT Architectures

Object is equipped with sensors that gather data on the action to be taken which is transferred over a network and actuators that allow things to act.

Gateway provides connectivity between the object and the **cloud** part of the IoT solution, to enables data pre-processing and filtering before moving it to the cloud and transmits control commands going from the cloud to things. The advantage of adopting a cloud gateway is that it ensures compatibility with various protocols and communicates with field gateways using different protocols depending on what protocol is supported by the relevant gateway.

Data lake is used for storing the data generated by connected devices in its original format. When specific data is needed for analysis, it is extracted from the data lake and loaded to a **big data warehouse**, where it is filtered, cleaned, structured and matched.

Data analysts use data from the big data warehouse, diagrams, or infographics, to find trends or understate the correlations and patterns to create more suitable algorithms for control applications.

Control applications are responsible for sending automatic commands and alerts to actuators. E.g., When sensors show that the soil is dry, watering systems get an automatic command to water plants.

User applications are the software component of an IoT system which connects IoT users to the devices and gives them the option to monitor and control their smart object through a mobile phone or web application.

3 Layer Architecture of IoTs

The **perception** layer is the physical layer, which has sensors for sensing and gathering parameters about the environment or identifies other smart objects in the environment.

The **network** layer is responsible for connecting to other smart things, network devices, and servers, and is also used for transmitting and processing sensor data.

The **application** layer is responsible for delivering application specific services to the user.

5 Layer Architecture of IoTs

The **Business layer** enables the organization to leverage data generated by IoT devices to make informed decisions and improve business processes.

The **application** layer is responsible for delivering application specific services to the user.

The **processing** layer stores, analyses, and processes huge amounts of data that come from the transport layer through various technologies such as databases, cloud computing, and big data processing.

The **transport** layer transfers the sensor data from the perception layer to the processing layer and vice versa through networks such as wireless, 3G, LAN, Bluetooth, RFID, and NFC.

The **perception** layer is the physical layer, which has sensors for sensing and gathering parameters about the environment or identifies other smart objects in the environment.