



Lebanese University - Faculty of Information

## Data Science Overview

### **Specialized foreign course lectures 2023**

#### **Data science class**

#### **Dr. Wafaa Hoteit**

This educational course provides introductory knowledge about data science through a set of foreign language texts to familiarize you with the terms and vocabulary used in this field.

This is required in the final exam of the course

In this course, we will introduce you to world of Data Science :

### **Data Science Definition**

Data Science concept has been further developed and has been increasingly linked to data analytics and big data. Companies have realized the value of Data Science and its usefulness in discovering knowledge and helping with decision-making. It is difficult to define data science in a formal way because it is expanding rapidly and revolutionizing many industries.

### **The following are prominent Data Science definitions we found in the literature:**

1- Data science is a field that involves extracting insights from data using statistical analysis and machine learning. It encompasses data acquisition, cleaning, analysis, and visualization. Its applications range from finance to healthcare, enabling organizations to make informed decisions based on data-driven insights.

2- Data science is the use of computers, equipment, programming systems, and algorithms to solve problems and explain actual phenomena, using theories derived from other sciences, including mathematics, information systems, statistics, and computer science.

3- Data science is the study of where information comes from, what it represents, and how it can be turned into a valuable resource in the creation of business and IT strategies.

**The main purpose** of a data science is to analyze data in ways that provide useful information to a company. That can include a mix of structured, unstructured and semistructured data, typically in large amounts that make it difficult to find meaning in the data without using advanced analytics approaches. Some common data science applications in businesses include anomaly detection to aid in fraud detection and cybersecurity efforts; pattern recognition for analyzing customer purchases, stock trading and other use cases; and predictive modeling of customer behavior, market trends and financial risks.

### **Data Science Importance**

In “Why Data Science Matters And How It Powers Business Value,” the company details eight ways that data scientists can add value to business.

1. Empowering management to make better decisions
2. Directing actions and defining goals based on trends
3. Challenging staff to adopt best practices and focus on issues that matter
4. Identifying business opportunities
5. Decision making with quantifiable, data-driven evidence

6. Testing these decisions
7. Identifying and refining of target audiences
8. Recruiting the right talent

**To complete this section, we will also provide a simple definition of the concepts of data mining, artificial intelligence, machine learning, as these are related to data science and to each other.**

### **What is Concept Mining**

Concept mining is a process that focuses on extracting ideas and concepts found in documents. Concept mining is helpful in a number of applications, including the evaluation of software programs and in various applications that are generally classified as part of the artificial intelligence discipline within computer science. The use of concept mining can take place with any type of documents, regardless of how they are formatted.

### **What is data mining?**

Data mining aims to understand and discover new, previously unseen knowledge in the data. Data mining applies algorithms to the complex data set to reveal patterns that are then used to extract useful and relevant data from the set. Statistical measures or predictive analytics use this extracted data to gauge events that are likely to happen in the future based on what the data shows happened in the past.

Data mining, also known as knowledge discovery in data (KDD), is a branch of data science that brings together computer software, machine learning (i.e., the process of teaching machines how to learn from data without human intervention), and statistics to extract or mine useful information from massive data sets.

Data mining is defined as the process of filtering, sorting, and classifying data from larger datasets to reveal subtle patterns and relationships, which helps enterprises identify and solve complex business problems through data analysis.

**Artificial intelligence (AI):** is concerned with making machines smart aiming to create a system which behaves like a human. Artificial Intelligence, conversely, refers to the simulation of human intelligence in machines. AI systems can learn from experience, adapt to new inputs, and perform tasks that typically require human intelligence, like speech recognition, natural language processing, and decision-making. AI can be categorized into two types: Narrow AI (Weak AI) and General AI (Strong AI).

**Machine learning :** Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning aims to develop algorithms that can learn from historical data and improve the system with experience. Machine learning is an artificial intelligence tool that processes mass quantities of data that a human would be unable to process in a lifetime. Machine learning perfects the decision model presented under predictive analytics by matching the likelihood of an event happening to what actually happened at a predicted time.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

### **Data Science and Machine Learning**

Data Science is a combination of algorithms, tools, and machine learning technique which helps you to find common hidden patterns from the given raw data. Whereas Machine learning is a branch of computer science, that deals with system programming to automatically learn and improve with experience.

### **Artificial Intelligence and Data Science**

In the fast-paced world of technology, two terms have gained immense popularity and intrigue: data science and artificial Intelligence (AI). Both these fields have been transforming various industries and shaping the future of how we interact with machines and data., While data science and AI are closely related and often used together, they are not the same. Data science is a broader field that involves extracting insights from data using various techniques, including AI. On the other hand, Artificial Intelligence (AI) is a specialized field within computer science dedicated to developing machines with the capability to perform tasks that would typically necessitate human intelligence.

### **Difference Between Data Science and Artificial Intelligence**

In simple terms, data science is the process of extracting useful insights from unstructured data. It's an interdisciplinary approach that merges various fields of computer science, statistics, and scientific processes and methods in order to draw conclusions from raw data points. Data science is believed to have brought about a fourth industrial revolution, and it's now at the heart of business decision-making. Companies have realized the enormous value in data processing and analysis.

Artificial intelligence, often referred to as AI, is a collection of complex computer algorithms that mimic human intelligence. Computers that are programmed with AI can "learn" as they go, becoming better at solving specific types of problems as they take in more data. It also involves

translations, understanding human speech, image recognition, speech recognition, and the process of decision-making.

According to the father of Artificial Intelligence John McCarthy, it is “The science and engineering of making intelligent machines, especially intelligent computer programs”.

Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks, and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.

Artificial intelligence is a product of human creation, developed to enable computers to read, understand, and learn from data, which helps in the decision-making process. These decisions are based on inferences that are otherwise difficult for humans to catch.

Now that you understand how the two relate to each other, let’s take a closer look at how they differ.

- 1- The significant difference is that data science involves pre-processing analysis, prediction, and visualization. AI is the implementation of a predictive model to foresee events.
- 2- Data science is an umbrella term for statistical techniques, design techniques, and development methods. Artificial intelligence has to do with algorithm design, development, efficiency, conversions, and the deployment of these designs and products.
- 3- Python and R are the tools used in data science, whereas TensorFlow, Kaffee, and scikit-learn are tools used in AI. Data science is primarily concerned with making use of data analysis and data analytics (where it uses past and present data to predict future data). Artificial Intelligence is concerned with machine learning.
- 4- Data science was developed to find hidden patterns and trends in data. The discipline aims to extract useful data, process it, make sense of it, and ultimately use it to make important decisions. On the other hand, artificial intelligence is used to handle data autonomously, removing the human from the entire task to work on its own.
- 5- By using data science, complex models can be built for extracting various facts, statistical techniques, and insights. On the other hand, artificial intelligence is meant for building models that emulate cognition and human understanding to a certain level. By emulating cognition, the aim is to create self-sufficiency, meaning the machine would no longer require any human input.

In simple terms, data science focuses on the process of extracting, cleaning, and analyzing data to gain valuable insights, in contrast, AI involves creating intelligent systems that can carry out

tasks autonomously, learn from their experiences, and enhance their performance as they gain more knowledge over time.

## **Data vs Big Data**

Data is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. In other words, it is a set of facts and figures which are useful in a particular purpose like a survey or an analysis. Moreover, when you arrange them in an organized form, they refer to as information. .

While **Big Data** is defined as a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size. Big data refers to data that is so large, fast or complex that it's difficult or impossible to process using traditional methods. The act of accessing and storing large amounts of information for analytics has been around for a long time. But the concept of big data gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three V's:

**Volume.** Organizations collect data from a variety of sources, including transactions, smart (IoT) devices, industrial equipment, videos, images, audio, social media and more. In the past, storing all that data would have been too costly – but cheaper storage using data lakes, Hadoop and the cloud have eased the burden.

**Velocity.** With the growth in the Internet of Things, data streams into businesses at an unprecedented speed and must be handled in a timely manner. RFID tags, sensors and smart meters are driving the need to deal with these torrents of data in near-real time.

**Variety.** Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data and financial transactions.

## **Types Of Big Data**

Big data types in Big Data are used to categorize the numerous kinds of data generated daily. Primarily there are 3 types of data in analytics:

1. **Structured Data:** or Organized data is easy to evaluate and sort since it has predetermined organizational characteristics. Each field is independent and accessible individually or with information from other fields since it is specified. Since of this, structured data is very significant because it enables the rapid data collection from numerous database locations. It is any data that can be processed, is easily accessible, and can be stored in a fixed format is called structured data.

2. Unstructured Data: in Big Data is where the data format constitutes multitudes of unstructured files (images, audio, log, and video). This form of data is classified as intricate data because of its unfamiliar structure and relatively huge size. A stark example of unstructured data is an output returned by 'Google Search' or 'Yahoo Search.'

3. Semi-structured Data: is a combination of both unstructured and structured types of data. This form of data constitutes the features of structured data but has unstructured information that does not adhere to any formal structure of data models or any relational database. Semi-structured data is a type of data that is not purely structured, but also not completely unstructured. It contains some level of organization or structure, but does not conform to a rigid schema or data model, and may contain elements that are not easily categorized or classified.

### **Data vs. Information**

Below, we'll take a deeper dive into data vs information and how these elements can work together in business decision-making. We'll also take a look at how a knowledge management platform can help you organize information and establish a data-driven culture.

#### **What Is Data?**

Data is a Latin word currently used in English. Data is the plural of Latin datum, "something given." In scientific use, the word data is often considered to be a plural noun meaning "pieces of information": The data are reliable. But most people think of data as a mass noun meaning "information" and use the word with a singular verb and pronoun: The data is reliable because it was tested by experts.

It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. In other words, it is a set of facts and figures which are useful in a particular purpose like a survey or an analysis. Moreover, when you arrange them in an organized form, they refer to as information.

#### **What Is Information?**

Information is defined as knowledge gained through study, communication, research, or instruction. Essentially, information is the result of analyzing and interpreting pieces of data. Whereas data is the individual figures, numbers, or graphs, information is the perception of those pieces of knowledge. For example, a set of data could include temperature readings in a location over several years. Without any additional context, those temperatures have no meaning. However, when you analyze and organize that information, you could determine seasonal temperature patterns or even broader climate trends. Only when the data is organized and compiled in a useful way can it provide information that is beneficial to others.

### **The Key Differences Between Data vs Information**

- Data is a collection of facts, while information puts those facts into context.
- While data is raw and unorganized, information is organized.
- Data points are individual and sometimes unrelated. Information maps out that data to provide a big-picture view of how it all fits together.
- Data, on its own, is meaningless. When it's analyzed and interpreted, it becomes meaningful information.
- Data does not depend on information; however, information depends on data.
- Data typically comes in the form of graphs, numbers, figures, or statistics. Information is typically presented through words, language, thoughts, and ideas.
- Data isn't sufficient for decision-making, but you can make decisions based on information.

### **Data Science vs Big Data**

Data Science utilizes algorithms and tools to draw meaningful and commercially useful insights from raw data. It involves tasks like data modelling, data cleansing, analysis, pre-processing etc. Big Data is the enormous set of structured, semi-structured, and unstructured data in its raw form generated through various channels.

Essentially, big data is the process of handling large volumes of data. It includes standard practices for data management and processing at a high speed maintaining the consistency of data.

### **Basic Components of Data Science**

Data science consists of many algorithms, theories, components etc. Before detail study of data science, we need to understand them. Now we will give you some insight into data science and its various components.

1. Data : The primary key component is the data itself. Data is a collection of factual information based on numbers, words, observations, measurements which can be utilized for calculation, discussion and reasoning. we must understand its types:

2: Statistics: Statistics is a branch of mathematics which deals with the collection, interpretation, analysis, presentation, and organization of data. It uses programming to analyze data. There are two main categories of Statistics:

Descriptive Statistics: Descriptive Statistics helps to organize data and only focuses on the characteristics of data providing parameters. For example, you want to find the average height of students in a classroom, in descriptive statistics, you will record the heights of all students in the class, and then you would find the maximum, minimum and average height of the class.

Inferential Statistics: Inferential statistics generalizes a large data set and applies probability before concluding. It also allows you to infer parameters of the population based on sample stats and build models on it.

3: Machine Learning: Machine learning is the field of computer science stemming from AI. It uses statistical techniques to give computers the ability to learn without being programmed. The machine progressively improves its performance on a specific task by changing the structure or program. There are three main goals of machine learning.

4: Big Data: Big Data is the name given to data is such a large quantity that storing or processing this data requires a large number of computers. It is characterized by three Vs:

Volume: Data in large volumes ranging from terabytes to zettabytes.

Variety: Data can show a lot of variety and diversity. It can be a mixture of two or more types of data, for example, structured and unstructured both.

Velocity: Data is being generated at a constantly growing rate. Essentially it is the speed of data.

In data science, data is grouped into many forms and types. Big data can be referred to humongous volumes of data which cannot be processed by using traditional applications. Data scientists use different tools to study and process big data, for example, Hadoop, Spark, R, and Java etc.

## 5. Programming languages

Data Science is all about data. To organize and analyze this data we use programming. Programming languages are of many types. The two most widespread being Python and R. Generally, data organization and investigation is finished by computer programming, therefore, Generally, data organization and investigation are finished by computer programming.

6- Data Analysis : before moving on to the modeling part, we first perform an exploratory data analysis to get a basic idea of the data and patterns which are available in it this gives us a direction to work on if we want to apply some complex analysis methods on our data.

Talking about statistics, machine learning, and programming, all of these constitute an integral part of the big numbers game called Data Science. Delving into each of them separately:

1- Data Visualization- It absorbs information quickly, improves insights, helps make faster decisions, improves the ability to maintain audiences' interest, eliminates the role of data scientists & eases out the distribution of information hence collected. Visualization means representing the data in visuals such as maps, graphs, etc. so that people can understand it easily. It makes it easy to access a vast amount of data. The main goal of data visualization is to make it easier to identify patterns, trends, and outliers in large data sets.

2- Deep Learning- These algorithms and multi-layered ANN require very powerful machines & are very useful in detecting patterns from input data.

3- Domain Expertise- A high level of expertise in the area can vastly improve the accuracy of the model you want to build.

4- Data Strategy: Developing a data strategy requires choosing what data to collect and why. Making the connection between the data you intend to collect and your business goals is necessary for choosing a data strategy. Data are not all made equal. The time and effort you invest into collecting data, presenting it correctly, and eliminating 'junk' data that doesn't advance your company objectives will ultimately reflect how challenging it is to do so and how valuable the data may be.

5- Data Engineering: The use of technology and systems to access, organize, and utilize data is known as data engineering. It typically entails the development of software to address data-related issues. These solutions often start with creating a data system, followed by the design of data pipelines and endpoints within it. This may require integrating a large number of different technologies. Data scientists often have assistance from data engineers, who help streamline analytics projects by handling much of the upfront work required to get data into the hands of data scientists.

### **Data Science life Cycle**

Data Life Cycle and Data Phases are two concepts related to the management of data. A data cycle is a process of acquiring data from an external source, manipulating it to suit the needs of the user, and then outputting the results while Data Phases refer to the distinct stages of a project where data is collected, processed and analyzed. Understanding these concepts can help organizations better manage their data and make informed decisions.

A data lifecycle also called the information life cycle, refers to the different stages a unit of data undergoes, from initial collection to when it's no longer considered useful and deleted. It's a continuous, policy-based process where each phase informs the next . these steps provide a general framework for conducting data science work. are :

1. Discovery: The life cycle of Data Science begins with the discovery of data. This stage involves collecting data from discrete sources like various social media platforms, online sources, logs, etc. This stage can be carried out by any individual who might not be a Data Scientist but has a piece of proper understanding of how to ask the right questions to examine different factors affecting a Data Science project. These factors include determining the various requirements of the project, estimating the overall expenses, time, technologies, etc.

2. Data Preparation: The effective analysis of data requires massive datasets but without compromising the quality of data. So after collecting the data and other resources, this stage is

carried out to select a quality data subset from a very large amount of raw data. Thus this process is also referred to as Data Cleaning. This stage is performed to locate and remove the outliers and several anomalies from the data as well as to discover various trends and relationships in the data.

3. Modeling, now this is the stage in which the data is actually analyzed to extract meaningful insights from it. This requires the knowledge of various areas of math like statistics, probability, linear regression, logistic regression, etc. The main aim of this phase is to identify suitable machine learning algorithms and to find as well as build the model which best fits the business needs.

4. Evaluation: Evaluating the performance of a knowledge mining technique may be a fundamental aspect of machine learning. Evaluation measures can differ from model to model, but the foremost widely used data processing techniques are classification, clustering, and regression. After building the models, these models are applied to the datasets to get the desired results. The results are then evaluated to test the performance of the models. These observations are used to improve the model to get the best results.

5. Communication: Communication is a very important phase in the Data Science life-cycle. In this phase, the Data Scientist presents all the key findings of the project to the stakeholders and other members of the organization. This is to evaluate the accuracy of the results and to ensure that the project satisfies all the user requirements. Depending on the feedback received from the stakeholders, the Data Scientist will make the required changes.

### **Data Science languages**

The language of data science is a set of tools that used in the field of data science to analyze and extract knowledge from data. Language in this context refers to the programming language and tools used to manipulate data and perform mathematical and statistical operations. such as cleaning and analyzing data, building machine learning models, and visualizing data.

### **What is Programming?**

It is the practice of structuring or arranging a set of instructions that instruct or direct a computer system to carry out certain tasks. These languages are used to design websites, create apps, develop operating systems, control spacecraft, and analyze data.

### **What is a Programming Language?**

A programming language is a language that allows people to write specific commands to be executed on a computer. It is a set of instructions that can be used to interact with and control a computer.

A programming language is a way for the Developers or Programmer to communicate with computers and other devices(IOT) using machine understandable code.

Programming languages consist of a set of rules that allows string values to be converted into various ways of generating machine code, or, in the case of visual programming languages, graphical elements.

Generally speaking, a program is a set of instructions written in a particular language (C, C++, Java, Python) to achieve a particular task.

A programming language is a set of instructions that can be used to interact with and control a computer.

Computer programming is an approach to instructing computers to perform various tasks. It is a way of communicating with machines or computer systems. Programmers write a set of instructions or commands using a language understandable by machines or computer systems, called programming language. On receiving instructions, machines or computer systems perform the specified tasks.

### **Basic Terminologies in Programming Languages**

Here is some of top programming terms that can help you get started in programming:

1- Algorithm: A set of guidelines for solving a problem. Algorithms are written using a language that computers can understand to tell them how to calculate and process data. An algorithm is a set of instructions or rules designed to solve a definite problem. The problem can be simple like adding two numbers or a complex one, such as converting a video file from one format to another.

2- Computer Program : A computer program is a set of instructions fed to a computer to perform a specific task. It enables us to communicate with computers and instruct them how to behave. Every action taken by a computer is controlled by a computer program. We can store a computer program as a file on a computer's hard drive. When we execute that file, the central processing unit (CPU) reads the instructions stored within the file, executes it, and the computer will perform accordingly. To understand better what a computer program is, here are some examples:

Microsoft Word is a computer program that enables us to create and edit documents.

A web browser, like Google Chrome or Mozilla Firefox, helps us to browse and retrieve the requested information from the web.

Operating system.

3. Programming Language: A programming language is a language that programmers use to communicate with computers. We can think of a programming language as a language used to

write instructions and create programs. Some popularly used programming languages are C, C++, Python, Java, Ruby, Perl, and JavaScript.

4. Application Programming Interface (API) is a set of rules, routines, and protocols to build software applications. APIs help in communication with third party programs or services, which can be used to build different software. Companies such as Facebook and Twitter actively use APIs to help developers gain easier access to their services.

4. Bug: A bug is a general term used to denote an unexpected error or defect in hardware or software, which causes it to malfunction. Even though bugs are often considered to be insignificant computer glitches, there have been instances where bugs have caused life-threatening conditions and led to major financial losses. This makes it imperative to invest in the process of finding bugs before programs are rolled out for their application. This process is known as testing.

6. Object-Oriented Programming: Object-oriented programming (OOP) is a model defined by programmers that revolve around objects and data rather than 'actions' and 'logic'. In OOP, not only the data type of a data structure is defined, but also the types of functions that can be applied to it. Through this, the data structure becomes an object that consists of both data and functions. Languages that use OOP concepts are Java, Python, C++, and Ruby.

7. Code: or source code is a term used to describe a written set of instructions, written using the protocols of a particular language, such as Java, C or Python. The code can also be used informally to describe text written in a specific language. There are instances where references to the code are made for different languages, such as 'PHP Code', 'HTML Code', 'Java Code' or 'CSS Code'.

8. Command-line interface: The command-line interface is a user interface based on the text. The UI is used to view and manage computer files. Command-line interfaces are also called command-line user interfaces, console user interfaces and character user interfaces. During the early 1960s and through the 1970s and 1980s, the command line interface was the primary means of interaction with most computers on terminals.

9. High-level language: A high-level language (HLL) is a programming language that lets the developer write programs irrespective of the nature or type of computer. But if a computer has to understand a high-level language, it should be compiled into a machine language. HLLs are considered high-level because they are in close proximity to human languages and further from machine languages. High-level languages include BASIC, C, C++, Pascal, Prolog, and FORTRAN.

10. Low-level language: A low-level language is a language that is very close to machine language and provides a little abstraction of programming concepts. Low-level languages are

closer to the hardware than human languages. The most common examples of low-level languages are assembly and machine code.

11. **Source data:** Source data or data source is the key location from which data is used in the program. The source data can come from a database, spreadsheet or even a hard-coded data location. When a program is executed to display data in a table, the program retrieves the data from its source and then presents it in the arrangement as defined in the code.

12. **Syntax:** Similar to human languages, programming languages have their own set of rules on how statements can be conveyed. The set of these rules is known as syntax. While a number of programming languages share many features, functions, and capabilities, they differ in syntax. Without the proper use of the syntax, one cannot write an executable program, and a wrong syntax will lead to a plethora of errors.

### **key data science techniques**

Techniques are several methods that a data scientist uses to perform different tasks. Such as collecting, storing, filtering, classifying, validating, analyzing and processing for final outcome. Data scientists and analysts generally work on the below techniques:

**1. Classification Analysis:** This type of analysis demands mathematical approaches. Likewise decision trees, linear programming, statistics, and neural networks. There is a need to identify and assign categories to the gathered data. For this purpose, we use classification analysis to analyze the data for a higher degree of precision. In contrast, classification algorithms are derived in the form of classes to attain target variables

**2. Regression Analysis:** it is used to determine how closely interrelated independent data variables depend on a dependent data variable. Similarly, it is a machine learning algorithm that helps to note down the changes of one of the values of a dependent variable. This technique aims to build models on datasets for estimating the value of the dependent variables.

**3. Linear Regression:** used to estimate a linear relationship between input variables and output variables.

**4. Anomaly Detection:** Anomaly detection is also called outlier detection. It is a stage in data mining where identifying data points, observations. And events derived from a dataset's apparent behavior occur. In addition, it is helpful to avoid hacking, intrusion detection, monitoring, fraud detection in credit card transactions. And operating environment to detect the fault.

Various classification, regression and clustering methods are key data science techniques used in analytics applications to identify relationships between different data elements. Examples include decision trees and naive Bayes classifiers for classifying data into categories; linear regression and multivariate regression; and k-means clustering and hierarchical clustering. Association

analysis is another technique, akin to clustering, that's done to find association rules between related data points.

## Major Types of Programming Languages

While you'll find dozens of ways to classify various programming languages, they generally fall into five major categories. Keep in mind that some languages may fall under more than one type:

1. Procedural programming languages: A procedural language follows a sequence of statements or commands in order to achieve a desired output. Each series of steps is called a procedure, and a program written in one of these languages will have one or more procedures within it.

Common examples of procedural languages include:

- C and C++
- Java
- Pascal
- BASIC

2. Functional programming languages: Rather than focusing on the execution of statements, functional languages focus on the output of mathematical functions and evaluations. Each function—a reusable module of code—performs a specific task and returns a result. The result will vary depending on what data you input into the function. Some popular functional programming languages include:

- Scala
- Erlang
- Haskell
- Elixir
- F#

3. Object-oriented programming languages (OOP): This type of language treats a program as a group of objects composed of data and program elements, known as attributes and methods. Objects can be reused within a program or in other programs. This makes it a popular language type for complex programs, as code is easier to reuse and scale. Some common object-oriented languages include:

- Java
- Python

- PHP
- C++
- Ruby

4. Scripting languages: Programmers use scripting languages to automate repetitive tasks, manage dynamic web content, or support processes in larger applications. Some common scripting languages include:

- PHP
- Ruby
- Python
- bash
- Perl
- Node.js

#### 5. Logic programming languages

Instead of telling a computer what to do, a logic programming language expresses a series of facts and rules to instruct the computer on how to make decisions. Some examples of logic languages include:

- Prolog
- Absys
- Datalog
- Alma-0

#### Other ways to classify programming languages

You'll find many more ways to categorize languages beyond the five listed above. Let's take a closer look at there other ways you can think about programming languages:

Machine and assembly languages: A machine language consists of the numeric codes for the operations that a particular computer can execute directly. The codes are strings of 0s and 1s, or binary digits (“bits”), Machine language instructions typically use some bits to represent operations, such as addition, and some to represent operands, or perhaps the location of the next instruction. Machine language is difficult to read and write, since it does not resemble

conventional mathematical notation or human language, and its codes vary from computer to computer.

**Algorithmic languages:** Algorithmic languages are designed to express mathematical or symbolic computations. They can express algebraic operations in notation similar to mathematics and allow the use of subprograms that package commonly used operations for reuse. They were the first high-level languages.

**High-level vs. low-level languages:** The biggest factor that differentiates high- and low-level programming languages is whether the language is meant to be easily understood by a human programmer or a computer. Low-level languages are machine-friendly, which makes them highly efficient in terms of memory usage but difficult to understand without the help of an assembler. Since they're not very people-friendly because they don't use human language, they're also not widely used to code. Examples of these machine languages include machine code, binary code, and assembly languages.

## **Best Programming Languages**

There are many programming language available across the world, each language is created for specific features and when the language is written it is required to be translated into bytecode and machine code which the computer can then perform. As per a data there are over 700 different kinds of programming languages which are available that can be utilized to talk to machines, here are the 10 best programming languages in 2024.

1- Python: Python is perhaps the most user-friendly programming language of any. It's often said that Python's syntax is clear, intuitive, and almost English-like, which makes it a popular choice for beginners. Python has a variety of applications that make it a versatile, powerful option when choosing the best programming language for your use case. If you're interested in back-end web development, for example, then the open-source Django framework, written in Python, is popular, easy to learn, and feature-rich. Django has been used in the development of some popular sites like Mozilla, Instagram, and Spotify.

Python can be used effectively to create any kind of program that does not need direct access to the computer hardware.. It is high-level programming with an open-source that was formulated in 1991 by Guido van Rossum. This language is presented by (PSF) Python Software Foundation).

2- Golang : Go language was developed in 2007 but launched in 2009 at Google by Ken Thompson, Rob Pike and Robert Griesemer, Go is the little language that could. As a low-level language, Go is ideal for engineers who want to enter the field of systems programming. It encompasses much of the same functionality of C and C++ without the difficult syntax and steep learning curve. It's the perfect language for building web servers, data pipelines, and even machine-learning packages. As a compiled language, Go runs "close to the metal," allowing for a

blazing-fast runtime. It's an open-source language, and ambitious developers can see their personal contributions adopted and enjoyed by programmers worldwide.

3- Swift: This open-source programming language was first announced by Apple Inc. Swift is a relatively new programming language used to develop iOS and macOS applications. Swift has been optimized for performance and built from the ground up to match the realities of modern iOS development. Not only does iOS run on every iPhone and iPad, but it's also the basis for other operating systems such as watchOS (for Apple Watches) and tvOS (for Apple TVs). In addition, Apple isn't going anywhere as a tech industry leader, and iOS apps continue to be the most profitable in the mobile app marketplace.

Any software that operates code and writes, it's the characteristic of Swift. Some of the software are desktops, phones etc. The data of Swift is organized automatically. Among all the programming languages of 2018 Swift rated 14th and it is in a good demand and it is more faster than the other programming language such as Python and Objective -C. Swift the most widely used programming languages for mobile app development.

4- JavaScript: A programming language that is used for many things like web development, application, web applications, web browsers etc. With the comparison of another language as common as HTML and CSS, Javascript has more characteristics on web pages. It's a scripting language that helps functional programming styles and a useful programming language for a web page. In this language, a programmer can find all calculations related to math and permits to add HTML data to Document Object Model(DOM) which is the data representation of the objects that comprise the structure and content of a document on the web. . It manages prototypes that emulate the most important characteristic.

5- C++ One of the well know programming language which is used for game development, system programming, high performance application and so on.. You can find this language in graphical user interfaces, embedded system and in operating system. This language is easy to learn and quite similar to java, C & C#.

6- Java: is one of the most popular programming languages used today. Java is widely used in web and application development as well as big data. Java is also used on the backend of several popular websites, including Google, Amazon, Twitter, and YouTube. It is also extensively used in hundreds of applications. New Java frameworks like Spring, Struts, and Hibernate are also very popular. With millions of Java developers worldwide, there are hundreds of ways to learn Java.

7- R: is the popular programming language among the Data Scientists, which can be used on Windows, UNIX, and Mac Operating System. The best feature of R language is a data visualization that would be tougher in Python, but it is less beginner-friendly than Python. . Applications built in R are used for processing statistics, including linear and nonlinear modeling, calculation, testing, visualization, and analysis. Applications coded using R can interface with a number of databases and process both structured and unstructured data.

8- C and C++: C is probably the oldest and popular programming language and is the root of other programming languages such as C#, Java, and JavaScript. C++ is an enhanced version of C. Many developers today skip learning C on its own, while others think learning C first provides a valuable foundation for C++ development. Both languages are widely used in computer science and programming.

Both C and C++ are considered high-performance languages. As such, they are widely used in developing applications where performance is a critical issue, such as client/server applications, commercial products like Firefox and Adobe, and video games.

9- SQL is a standard database query language. It is used to access and manipulate data in databases. SQL is a declarative language that specifies the desired results, but not the steps to achieve those results. SQL is a powerful tool for accessing and manipulating data, and it is the world's most widely used database query language. SQL is a standard database query language that enables users to manipulate and query data in a database quickly. SQL is widely used in many applications and environments, such as web applications, data warehouses, and e-commerce applications.

Along with Python, SQL programming is essential to efficiently using data in web applications. SQL specifically enables the storage, extraction, and manipulation of data within an existing database.

10- HTML : (HyperText Markup Language) is the standard language for creating web pages and applications. HTML is used to create web pages. this language is used for: Web documents , Website development , and Website maintenance> Technically, HTML is a markup language, which means that it is responsible for formatting the appearance of information on a website. Essentially, HTML is used to describe web pages with ordinary text. It doesn't have the same functionality as other programming languages in this list and is limited to creating and structuring text on a site. Sections, headings, links and paragraphs are all part of the HTML domain.

11- CSS : this language is used for: Web documents , Website development , and Website design, CSS, or cascading style sheets, is usually applied in conjunction with HTML and governs the site's appearance. While HTML organizes site text into chunks, CSS is responsible for determining the size, color and position of all page elements. CSS is convenient, too; the cascading part of the name means that an applied style will cascade down from parent elements to all children elements across the site. This feature means that once users determine aesthetics for the main parent, they won't have to manually repeat their code across a website. Moreover, the delegation of site organization to HTML and aesthetics to CSS means that users don't have to completely rewrite a web page just to change a color. CSS is an approachable language that allows beginning programmers to dip their toes in the metaphorical coding pool. If you're new to coding, there's no reason not to learn CSS before tackling more complex languages!

12- JAVA: used for: E-commerce , Finance, and App development, Java is a general use and object-oriented programming language. In object-oriented programming, developers create objects that encompass functions and data, which can then be used to provide structure for programs and applications. Java is an object-oriented programming language providing a large number of tools and libraries. It is simple, portable, secure, platform-independent, object-oriented, and multi-threaded; that is why it is suitable for data science and machine learning.

### **Data Analysis vs. Data Analytics : Definition ,Types and difference**

When comparing data analysis vs. data analytics, it's important to note that while these two concepts are similar, they're not identical

#### **What is data analysis?**

Data analysis is an activity that many organizations perform to evaluate data they collected from various sources, such as their financial records, surveys, and social media posts from competitors and their target market. Its goal is to find meaning in the data to help decision-makers better understand their organization and the market they serve. It refers to the process of using data and analytical tools and techniques to find new insights and make predictions, often for the benefit of an organization.

Data analysts are the people who perform data analysis. They often use a variety of methods to do this, including using software to manipulate data into an organized format and employing different techniques, such as data visualization, to identify trends.

#### **Types of data analysis**

Data analysts use a variety of techniques to find meaning in their data. The method they employ often depends on the format of the data, their goals, and the available resources. For example, when reviewing financial records, they may choose to run a basic analysis on the data using software to explore and make comparisons. In general, data analysis techniques can either be quantitative or qualitative.

Quantitative data analysis often involves using software to generate statistics and other numerical data from records and surveys. This method can help a company answer questions about the state of its finances, such as how much money it spent on a particular product or the daily number of customers who visit a retail store

#### **What is data analytics?**

Data analytics is a broad field that involves the collection, analysis, and interpretation of data to help an organization make better decisions and achieve its goals. It's useful in a wide variety of industries, including business, health care, education, and government. The practice of data analytics encompasses a number of processes that make up a data science workflow or pipeline.

These often include gathering, preparing, exploring, managing, storing, and evaluating data. The last stage is usually sharing or delivering the results of data analysis to an organization's decision-makers or business users.

## **Types of data analytics**

Data analytics enables data analysts and other professionals working with data to explore, interpret, and share information about a wide variety of topics, such as human behaviour, organizational trends, and consumer preferences. There are four basic types or categories of data analytics. The four types include:

**Descriptive analytics:** is the simplest of the four types of data analytics. It involves historical data gathering and visualization to gain insight that organizations may use to aid them in their decision-making. It aims to inform them of what happened in the past. For example, a retail chain may use descriptive analytics to analyze its sales data over the past year to determine what products are most popular with its customers. This is a useful way for the company to understand its customers better and inform its product development process.

**Diagnostic analytics:** While descriptive analytics is about describing historical data, diagnostic analytics focuses on determining the cause of a particular outcome or event in the past. It uses data to identify trends that may be useful for an organization to know. For example, a real estate company may use diagnostic analytics to evaluate the sales data of its recent buyers and determine what price point is most attractive to them. This can help the company improve its sales strategy.

**Predictive analytics:** This type differs from descriptive and diagnostic analytics because it doesn't simply examine historical data. Instead, it uses data to foretell future outcomes or events to help companies attain their goals. To do this, data analysts often create optimized simulation models to help them make accurate predictions. For example, a retailer may use predictive analytics to analyze its data and predict the best time of year to offer customer discounts. This allows the retailer to make the most of its resources and potentially increase its profits.

**Prescriptive analytics:** Prescriptive analytics is the most advanced type of data analytics. It's similar to predictive analytics in that it also offers valuable insight to organizations. It's different because it focuses on recommending the best course of action or solution based on the results of predictive analyses. Prescriptive analytics usually requires the use of machine learning algorithms and other advanced tools

## **Data Analysis vs. Data Analytics : differences**

Data analysis involves extracting meaning from data in a way that's useful to a decision-maker. Data analytics is broader in scope. While data analysis focuses on exploring data in its raw form, data analytics uses various processes to convert data into actionable information.

Data analytics is a general term for a field that includes the entire management of data and the methods and instruments employed, While data analysis is the process of dissecting a given data set into its component pieces and analyzing each one separately, as well as how the parts relate to one another.

Data analytics is the broad field of using data and tools to make business decisions While Data analysis is a process involving the collection, manipulation, and examination of data for getting a deep insight. While Data analysis, a subset of data analytics, refers to specific actions. It is taking the analyzed data and working on it in a meaningful and useful way to make well-versed business decisions.