Study of Various Data science and Data Analytics Approaches

Prof.Kamaljeet Singh Kalsi

Department of Computer Science & Engineering

Gyan Ganga College of Technology

Jabalpur (MP), India

kamaljeetsingh@ggct.co.in

Prof. Om Singh Parihar

Department of Computer Science & Engineering

Gyan Ganga College of Technology

Jabalpur (MP), India

omsinghparihar@ggct.co.in

https://mail.google.com/mail/u/0/images/cleardot.gif ABSTRACT

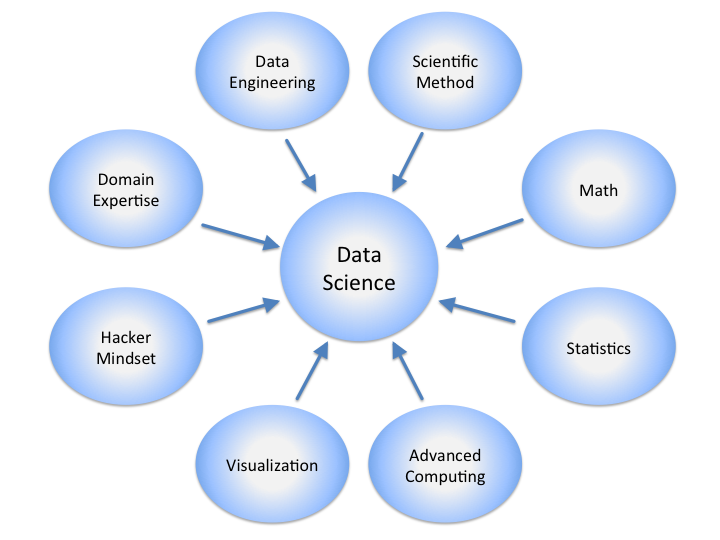
Data Science is the study of extracting, collecting, aggregating, representing, and safeguarding data for use in commercial or technological difficulties whereas Data analytics is the science of studying raw data in order to draw conclusions about it. Many methods and techniques used in data analytics have been transformed into mechanical operations and algorithms that process raw data for human consumption. In this paper, we explain the concept, characteristics & need of Data science and Data Analytics.

Keywords— Information, Data Science, Data Analytics.

# INTRODUCTION

**Science** is the extraction of learning from large amounts of disorganized or unstructured information, which is a continuation of the discipline of information mining and perceptive research, also known as information disclosure and information mining. "Unstructured data" might include messages, features, images, online networking, and other client-generated content. Information science typically necessitates dealing with massive amounts of data and building computations to extract knowledge from this data.

Data science is the study of data to extract meaningful business insights. It is a multidisciplinary approach that combines principles and practises from mathematics, statistics, artificial intelligence, and computer engineering to analyse large amounts of data. This analysis enables data scientists to ask and answer questions such as what happened, why it happened, what will happen, and what can be done with the results. Data science is important because it combines tools, methods, and technology to generate meaning from data. Modern organisations are inundated with data, and there is an explosion of devices that can collect and store data automatically. Online systems and payment portals collect more data in e-commerce, medicine, finance, and every other aspect of human life. We have massive amounts of text, audio, video, and image data. Data processing has become faster and more efficient as a result of advances in artificial intelligence and machine learning. Industry demand has spawned an ecosystem of data science courses, degrees, and job opportunities. Data science is expected to grow rapidly in the coming decades due to the cross-functional skillset and expertise required.Data Science is a multidisciplinary field that extracts knowledge and insights from data using scientifically based processes, algorithms, and systems. Data science is a synthesis of statistics, machine learning, and data analysis, each of which is used to understand and conceptualise real-world phenomena, as previously stated. in figure1.



**Figure 1:** Features of Data Science

Data science has numerous applications. Data science is more important than ever before, thanks to the exponential increase in computational power and the proliferation of a more connected world. Data science can be used in almost any discipline that involves analysing large amounts of data. Data science is an invaluable tool for the creation and operation of much of our modern world, from warehouse delivery logistics to the cutting edge of artificial intelligence. Shipping companies all over the world use data science to improve shipping routes, delivery times, and even mode of transportation. Shipping companies can improve their efficiency and cut costs by processing large amounts of tracking data. UPS, for example, analysed all of their trucks' delivery routes and compared them to fuel consumption. Data from hospitals and healthcare facilities can be analysed to improve patient care and better understand the nature of disease and illness. Data science gives doctors the tools they need to provide faster turnaround times for lab results, which leads to increased patient satisfaction and, in some cases, life-saving preventative treatment. To detect fraud, banks use data science to comb through their financial data. Customers are now profiled based on their spending and saving data in order to reduce the bank's losses. Using data science, banks are able to better understand their customer's past expenditures, creating [probabilities](https://deepai.org/machine-learning-glossary-and-terms/probability) for potential risk and default. Similarly, with the use of data science tools to understand a customer's spending account, a bank's customers are now alerted more promptly to potential compromises to their own account.

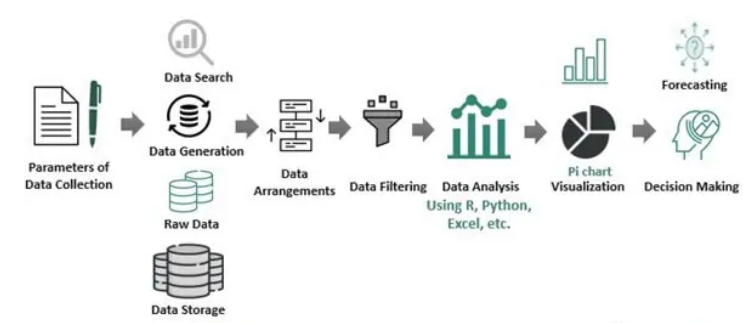
**Data analytics is** a broad term that encompasses a wide range of data analysis techniques. Data analytics techniques can be applied to any type of information to gain insight that can be used to improve things.. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be refered to optimize processes to inhance the overall efficiency of a business or system.

With the rise in global data volume, big data technology and related analytical procedures are increasingly being employed to describe huge datasets. Big data, in contrast to other typical datasets and procedures, contains semi structured and unstructured data that requires greater real-time analysis. Big data also provides information about new prospects for deciding new values, assists us in improving an in-depth grasp of the hidden values, and introduces new obstacles, such as how to excellently arrange and manage such large datasets. The volume of information from numerous sources is increasing; it also gives knowledge on certain difficult topics that require quick resolution. Another key technique that plays a major role in big data analytics difficulties is large data visualization.

This makes the data generation process easier because information technology (IT) has lately made significant advancements. For instance, users upload almost 72 hours' worth of video clips to YouTube every minute. This data growth challenges the stream along with the main issues of clubbing and integrating huge volume of data from widely distributed data sources such as social media applications.

Additionally, the unanticipated expansion of cloud computing and the Internet of Things (IoT) encourages data growth. The industry standard for storing and retrieving business data for big data assets is provided by cloud computing. On the Internet of Things, sensors are used to collect and send data that will be analysed and stored in the cloud. Such data sizes and kinds surpass the capabilities of the IT systems and setups of current businesses, as well as their need for real-time processing and computing power. The enormous heterogeneous datasets must be stored and retrieved using specialised hardware and software infrastructure, which is complicated by the growth in data volume.

Data analytics tends to examining unprocessed databases to conclude meaningful and actionable inferences about the content, which is being contain by it. It helps financial researchers and analyst view trends in the unprocessed data and withdrawal significant knowledge from it. Various paradigm of data analytics is mentioned in figure 2.



F

Figure 2: Various Paradigm of Data Analytics model

# RELATED WORK

"The extraction of a reasonable answer from a given body of data does not assure that a reasonable answer can be derived from a given body of data," the author writes. [1] [2]. Data increasingly benefits both research and industrial areas such as health care, finance service and commercial recommendation [3]. Big Data architecture has to obtain high speed data from a different kind of data sources and it has to deal with different access control protocols. It is where a filter could be recognized to store only data which could be helpful or underdone data with a lesser degree of uncertainty [4]. In some applications, the conditions of generation of data are important, thus it could be interesting for further analysis to capture these metadata and store them with the corresponding data. In 1974, Peter Naur published a “Concise Survey of Computer Methods”[5] which was a survey of contemporary data processing models that are used in a wide range of applications.Naur offered the following definition of data science: “The science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences.”

Arranging, packaging, and communicating information are the three components of data science (the ABC of information).Bundling, though, is an essential component of data wrangling, which also comprises data collecting and sorting. But what sets data science apart from other disciplines is that it also requires constant awareness of the What, How, and Who.also, why. A data science researcher needs to be aware of and have a clear vision for the results of the data science transformation. A data science researcher needs to have a well defined plan for how this output will be achieved within the constraints of available resources and time.finally, results are made public through textual reports and/or executable code[6]. In conclusion, managing big data necessitates having an infrastructure that is linearly scalable, capable of processing high throughput, multi-formatted data, automatically recoverable, fault tolerant, has a higher level of parallelism, and utilises distributed data processing [7]. Fraudulent claims or transaction processing can take many different forms in business operations. Therefore, the most significant application of big data is fraud recognition and control. [8].

Mobile devices, social networks, the Internet of Things, multimedia, and numerous more new applications can all produce big data [9]. The efficient management of urban areas, particularly real-time management, can be facilitated by advanced analytics of these data utilising machine learning techniques [10].

# CASE STUDY OF DATA SCIENCE

In the era of "data science and advanced analytics," nearly every aspect of daily life is captured digitally and stored as data. Therefore, the modern electronic world is a treasure trove of different types of data, including business, financial, healthcare, multimedia, internet of things (IoT), cybersecurity, social media, etc.

The amount of structured, semi-structured, and unstructured data grows daily. To understand and study the actual phenomena with data, data science is often referred to as a "concept to unify statistics, data analysis, and their related approaches”,a recommendation, an idea, a model, a paradigm, a tool, or a system.

In order to refer data, data science is used in four main ways:

**1. Descriptive analysis**

Descriptive analysis examines data to gain insights into what happened or what is happening in the data environment. Data visualisations like pie charts, bar charts, line graphs, tables, or created narratives are what define it. For instance, an airline booking service might keep track of information like how many tickets are purchased each day. For this service, descriptive analysis will provide peak booking periods, peak booking periods, and high-performing months.

**2. Diagnostic analysis**

A diagnostic analysis is a deep examination of data to determine why things occurred. Drill-down, data discovery, data mining, and correlations are some of the techniques used. On a given data set, multiple data operations and transformations can be performed to discover unique patterns in each of these techniques. For example, the flight service may focus on a particularly strong month to better understand the booking surge. This could lead to the discovery that a large number of customers visit a specific city to attend a monthly sporting event.

**3. Predictive analysis**

Predictive analysis makes accurate forecasts about data patterns that may occur in the future using historical data. Machine learning, forecasting, pattern matching, and predictive modelling are some of the techniques used. In each of these techniques, computers are trained to reverse engineer causality connections in the data.For example, the flight service team might use data science to predict flight booking patterns for the coming year at the start of each year. The computer program along with algorithm may look at previous data and predict booking spikes for certain destinations. Having anticipated their customer’s future travel requirements, the company could start targeted advertising for those cities from February.

**4. Prescriptive analysis**

It not only forecasts what is likely to happen, but it also suggests the best way to respond to that outcome. It can assess the potential consequences of various options and recommend the best course of action. It employs graph analysis, simulation, complex event processing, neural networks, and machine learning recommendation engines. To return to the flight booking example, prescriptive analysis could look at previous marketing campaigns to maximise the benefit of the upcoming booking spike. A data scientist could forecast booking outcomes for various levels of marketing spend across multiple marketing channels. These data projections would increase the flight booking company's confidence in their marketing decisions..

Data scientists work with sophisticated technologies such as:

1.**Machine learning models** and related software are used for predictive and prescriptive analysis using artificial intelligence.

2.**Cloud computing:** Cloud technologies have provided data scientists with the flexibility and processing power needed for advanced data analytics.

3.**Internet of things (IoT**) refers to various devices that can automatically connect to the internet. These devices collect data for data science initiatives. They generate massive amounts of data, which can be used for data mining and extraction.

4.**Quantum computing:** Quantum computers can perform complex calculations at high speeds. They are used by skilled data scientists to construct complex quantitative algorithms.

For example, Amazon Web Services (AWS) offers a variety of tools to help data scientists all over the world:

**1.Storage of data**

Amazon Redshift can run complex queries against structured or unstructured data for data warehousing. AWS Glue can be used by analysts and data scientists to manage and search for data. AWS Glue generates a unified catalogue of all data in the data lake, complete with metadata to make it discoverable.

**2.Learning by machine**

Amazon SageMaker is a machine learning service that is fully managed and runs on the Amazon Elastic Compute Cloud (EC2). Users can use it to organise data, create, train, and deploy machine learning models, and scale operations.

**3.Analytics**

a.**Amazon Athena** is a query service that allows you to easily analyse data stored in Amazon S3 or Glacier. It is quick, serverless, and uses standard SQL queries.

b.**Amazon Elastic MapReduce (EMR)** uses servers such as Spark and Hadoop to process large amounts of data.

c.**Amazon Kinesis** enables the real-time aggregation and processing of streaming data. It makes use of website clickstreams, application logs, and IoT device telemetry data.

d.**Petabytes of data** can be searched, analysed, and visualised using Amazon OpenSearch.

Data science is frequently defined as the application of sophisticated analytical techniques and scientific concepts to extract useful business intelligence from data. Advanced analytics focuses on anticipating future events by using data to identify trends and predict what is likely to occur.While basic analytics provide a general description of data, advanced analytics go a step further by providing a more in-depth understanding of data and assisting in the analysis of granular data, which is what we are interested in. In the field of data science, popular types of analytics include "descriptive analytics," which explains what happened, "diagnostic analytics," which explains why it happened, "predictive analytics," which predicts what will happen in the future, and "prescriptive analytics," which prescribes what should happen, which specifies what should be done. These types of analytics are briefly covered in "Advanced Analytics Methods and Smart Computing."Due to its learning power for smart computing as well as automation, such sophisticated analytics and decision-making based on machine learning techniques, a major part of artificial intelligence (AI), can also play a important role in the Fourth Industrial Revolution (Industry 4.0).

**Data Science Methodology:**

Every Data Scientist needs a methodology to solve data science’s problems. For example, let’s suppose that you are a Data Scientist and your first job is to increase sales for a company, they want to know what product they should sell on what period. You will need the relevant methodology to organize your work as systematic, analyze different types of data, and solve their problem. Your customer doesn’t care about how you do your job; they only care if you will manage to do it in time.

Methodology in Data Science is the key way to organize your relavent work, doing it better, and without losing time. Data Science Methodology is composed of 10 parts in figure 3:

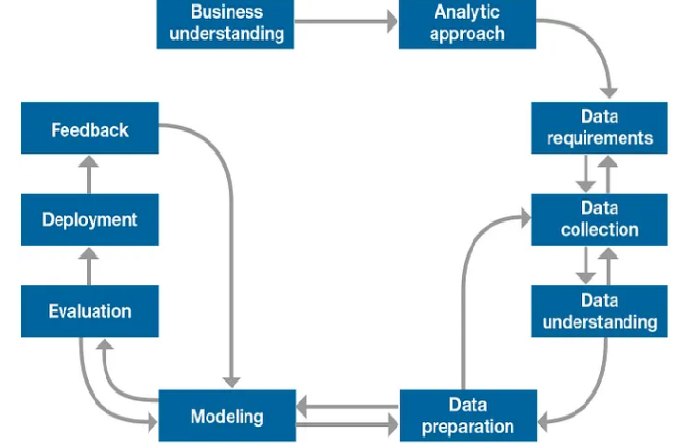


Figure 3: Steps of Data Science Methodology

1. **From Problem to Approach**

Every customer request begins with a problem, and Data Scientists' job is to first understand the problem and then approach it using statistical and machine learning techniques.

The Business Understanding stage is critical because it clarifies the customer's goal. At this stage, we must ask the customer numerous questions about every aspect of the problem; this ensures that we will study data-related issues, and at the end of this stage, we will have a list of business requirements.

The Analytic Approach is the next step, where the data scientist can define the analytic approach to solve the problem after the business problem has been clearly stated.

This step entails expressing the problem in the context of statistical and machine-learning techniques, and it is critical because it helps identify what types of patterns will be required to effectively address the question. A predictive model may be used to determine the probabilities of something; a descriptive approach may be required to show relationships; and statistical analysis is the best way to solve our problem if counts are required. We can use different algorithms for each approach.

1. **From Requirements to Collection**

Once we have found a way to solve our problem, we will need to discover the correct data for our model.Data Requirements is the stage where we identify the necessary data content, formats, and sources for initial data collection, and we use this data inside the algorithm of the approach we chose.

In the Data Collection Stage, data scientists identify the available data resources relevant to the problem domain. To retrieve data, we can do web scraping on a related website, or we can use repository with premade datasets ready to use. Usually, premade datasets are CSV files or Excel; anyway, if we want to collect data from any website or repository, we should use Pandas, a useful tool to download, convert, and modify datasets. Here is an example of the data collection stage.

1. **From Understanding to Preparation**

Data scientists use descriptive statistics and visualisation techniques to better understand data now that the data collection stage is complete. Data scientists examine the dataset to determine its content, determine if additional data is required to fill any gaps, and verify the data's quality. In the Data Understanding stage, data scientists attempt to learn more about the previously collected data.We have to check the type of each data and to learn more about the attributes and their names.Data scientists prepare data for modelling in the Data Preparation stage, which is one of the most important steps because the model must be clean and error-free. At this stage, we must ensure that the data is in the proper format for the machine learning algorithm that we selected in the analytic approach stage. The dataframe must have appropriate column names as well as a unified boolean value (yes, no or 1, 0). We must pay attention to the names of each data because they may be written in different characters but mean the same thing; for example (WaTeR, water), we can fix this by making all column values lowercase.

1. **From Modeling to Evaluation**

We can begin modelling once the data has been prepared for the chosen machine learning algorithm.

The data scientist has the opportunity to determine whether his work is ready to go during the Modeling stage. Modeling is concerned with creating models that are either descriptive or predictive, and these models are based on statistical or machine learning analytic approaches. Descriptive modelling is a mathematical process that describes real-world events and the relationships between the factors that cause them. For example, a descriptive model might investigate whether a person would prefer this if they did this. Predictive modelling is a process that forecasts outcomes using data mining and probability. This step can be repeated more times until the model understands the question and answer to it.

In the Model Evaluation stage, data scientists can evaluate the model in two ways: Hold-Out and Cross-Validation. In the Hold-Out method, the dataset is divided into three subsets: a training set as we said in the modeling stage; a validation set that is a subset used to assess the performance of the model built in the training phase; a test set is a subset to evaluate the likely future performance of a model.

1. **From Deployment to Feedback**

Data scientists must familiarise stakeholders with the tool produced in various scenarios, so that once the model has been evaluated and the data scientist is confident it will work, it can be deployed and put to the ultimate test.

The Deployment stage is determined by the model's purpose and may be rolled out to a small group of users or in a test environment. A real case study example could be for a model destined for the healthcare system; the model could be deployed for some low-risk patients first, and then for high-risk patients.

The feedback stage is usually where the customer contributes the most. Customers can decide whether or not the model is suitable for their needs after the deployment stage.

# DATA ANALYTICS APPROACHES

The definition of data analytics is an advanced scientific field in which financial analysts collect raw data from the past and draw meaningful inferences for proper action based on the information contained. They employ a variety of statistical, machine learning, and other technical tools. Companies then use the inferences to make sound business decisions.

Corporations use data analytics to spot trends and provide information by examining various data types (past, real-time, raw, qualitative, and organised). They include automating judgement, insight, and activity in certain situations. In a nutshell, data analytics skills entail extracting raw data, organising it, and then converting it into homogeneous, cognitive, and visual information to assist businesses and organisations. The simple results then allow businesses to develop strategies for future actions to improve their operations. Business analytics can also help in identifying useful patterns in consumer and employee behaviour when interacting with customers to solve their problems. It is also useful in predicting future performance based on past data in a logical and data-backed manner. As a result, businesses are better equipped to deal with unforeseen mishaps, make informed decisions, and thus plan accordingly to sustain the business.

Corporations such as Google have created data analytics certification in big data analytics for the same purpose. These courses teach data analytics with Microsoft Excel to both employees and individuals. Furthermore, they aid in the advancement of innovation and development in modern businesses.

**Types of Data Analytical Approaches:**

The field of data analytics has always been vast, with four major categories.

1.**Descriptive Analytics**

It helps to clarify what happened. Such methods distil large datasets into concise summaries that stakeholders can comprehend. These strategies allow for the creation of key performance indicators (KPIs), which aid in the monitoring of success or failure. It is beneficial to analyse various industries and employ metrics such as return on investment (ROI). Technical measures aid in the tracking of productivity in specific industries.

**2.Predictive Analytics**

It aids in addressing concerns about what will happen soon. These methods rely on historical data to identify trends and assess their likelihood of repetition. Furthermore, predictive analysis techniques that employ a variety of artificial intelligence and statistical approaches, such as regression and decision trees,

**3.Prescriptive Analytics**

It aids in answering the question of what companies must do. As a result, businesses can use predictive analytics insights to make data-driven decisions. In the event of ambiguity, this enables businesses to make sound decisions. The foundation of prescriptive analytics tools is machine learning algorithms that can identify trends in massive datasets.

**4.Diagnostic Analytics**

It contributes to explaining why certain events occurred. These methods help with more basic descriptive analytics. They dig deeper into the descriptive analytics results to find the source of the problem. Finally, research analysts conduct additional investigations to determine why performance metrics improved or declined. This is usually accomplished in three steps:

**Tools of Data Analytics:**

Various methods for extracting insightful information from the provided data may be available. Regrettably, some of them rely on coding, while others do not. Among the most popular analytics data tools appear to be:

**1. SAS**

SAS is a copyrighted piece of C-based software comprised of over 200 distinct components. Learning it is simple because its programming code is considered top-level. Nonetheless, it simply publishes the results via an Excel worksheet. As a result, many businesses, including Twitter, Netflix, Facebook, and Google, use it. Furthermore, SAS is improving to demonstrate that it is a significant player in the data analytics business, despite challenges from innovative coding languages such as R and Python.

**2.Microsoft Excel**

Businesses also use it to make real-time changes to data gathered from various other sources, such as stock market reports. It is especially important when performing rather complex data analytics when compared to other programmes such as R or Python. It ensures a clear picture of the data. Furthermore, financial analysts and sales managers frequently use it to address difficult company issues.

3**.R** is one of the key programming languages for generating detailed statistical visualisations. It is open and free software that can be used with Windows, MacOS, and a variety of UNIX operating systems. It also includes an easy-to-use command-line interface. Nonetheless, learning it can be difficult, especially for those with no prior coding experience. It is also extremely beneficial.

**4. Python**

Python is one of several effective technologies at the user's disposal for data analytics. It contains a large number of libraries and packages. A freeware, open-source application called Python has modules like Matplotlib and Seaborn that can be used for complex visualisation. Pandas is the name of the widely used Python data analytics package. Analysts usually use Python as a beginner's coding language due to its effectiveness and versatility. Python is utilised on many different platforms and has many different applications.

**Techniques used in Data Analytical approaches:**

Analysts can categorise data analytics methods into those that rely on machine learning, artificial intelligence, graphs and visualisation, statistics, and mathematics.

**1.Techniques in Statistics and Mathematics**

It includes four different types: time series analysis, classification analysis, regression analysis, and descriptive analysis.

**a.Descriptive analysis** is used to describe performance in relation to a given benchmark of analytical approaches by taking into account historical data and key performance indicators.

**b.Analysis of Regression -** It is a technique that explains how one or more independent or dependent variables are related to one another. It could use a variety of models, including nonlinear or multiple life data models.

**c.Analysis of Classifications** - It is also the most popular and commonly applied technique for data analytics. It is used by analysts to predict which group fresh observations will belong to. information about an established

**d.Time Series Analysis** is a popular format that is frequently used when carefully analysing a particular time series. When time is a factor in the outcome, it analyses changes that have happened over time. It facilitates the identification of cyclical patterns, seasonal fluctuations, and systemic patterns in observational data.

**2. Graphs and Visualization Methods**

The Word Cloud Chart, Line Chart, Gantt Chart, Bar Chart, Column Chart, Area Chart, Pie Chart, and Scatter Plot are a few of the tools used.

**3.Techniques for machine learning and artificial intelligence**

Artificial neural networks, decision trees, evolutionary programming, and fuzzy logic are a few examples of approaches.

**Process Of Data Analytics Approaches:**

The process of data analytics follows the following steps-

1.Establishing the parameters in order to categories data.

2.Gathering data from various authenticated sources.

3.Further rearranging the data.

4.Filtering the data and make sure no errors found in it. Afterward, one analysis the data to ensure it isn’t lacking information.

5.Further, one utilizes data which not contain error to analyze and identify with the help of the tools like Excel, R, or Python etc.

6.When the concepts are known, one varies the raw data into graphics for the management and employees to understand it in better way.

7.In the last step; the management goes through the data analytics and decides whether to act upon them or not.

These technologies are necessary for data scientists to enhance the efficiency of the process. The main features of data analytics are as:

**1. Data preparation and wrangling**

Before employing any iterative models, Data Preparation processes should be carried out once during the project. On the other hand, Data Wrangling is carried out during iterative analysis and model building. This concept originated during the feature engineering era.

**2. Data investigation**

Data exploration is the first stage of data analysis, and it entails looking at and visualising data to discover insights immediately or identify areas or patterns that require more research. In order to better understand the big picture, users may more quickly get insights by employing interactive dashboards and point-and-click data exploration.

**3.Scalability**

A quicker server with more potent processors and memory is required to scale up, or vertically scale, a system. This method requires less energy and network equipment, but it might only be a temporary fix for a number of big data analytics platform flaws, especially if more expansion is predicted.

**4. Assistance with various forms of analytics**

The big data revolution has led to the development of new stages, types, and forms of data analysis. Data analytics is taking off in boardrooms throughout the globe, giving tools for enterprise-wide commercial success. However, what do these signify for businesses? Developing the necessary knowledge leads to information that gives organisations a competitive edge, which is essential for businesses to successfully use big data.

**5. Version management**

The act of tracking and regulating changes to software code is known as version control, also referred to as source control. Software development teams can maintain track of changes to source code over time with the aid of electronic technologies called version control systems.

**6. Data administration**

Data management is the process of acquiring, storing, and utilising data in a safe, efficient, and economical manner. Data management enables decision-making and actions that will benefit the business as much as is practical by optimising the use of data within the constraints of policy and regulation. An effective data management strategy is more crucial than ever as firms depend more and more on intangible assets to generate value.

**7. Integration of Data**

The practise of merging data from several sources to provide individuals a unified viewpoint is known as data integration. Data integration's main goal is to make data more accessible and easier for users and systems to obtain, use, and process. When carried out properly, data integration can improve data quality, release resources, reduce IT expenses, and foster creativity without materially altering existing applications or data structures. Apart from the fact that IT companies have always required to integrate, the advantages of doing so might never have been this great.

**8.Data Governance**

The process of ensuring that data is reliable, accurate, accessible, and useable is known as data governance.

**9. Data protection**

Data security is the practise of guarding against unauthorised access to, corruption of, or theft of digital data at any stage of its lifetime. All aspects of data security, including administrative and access controls, logical programme security, and physical hardware and storage device security, are included in this idea. Another important aspect of data analytics is data security. Another important aspect of data analytics is data security. The organization's policies and procedures are also covered.

**10. Displaying data**

In our increasingly data-driven world, having simple ways to see and understand data is more important than ever. After all, employers are increasingly looking for workers with data capabilities.

**Characteristics of Big Data Analytics:**

The following are some traits of big data analytics:

**Volume-**Large data dimensions and volumes handled and analysed by businesses.

**Value:** From a business perspective, value is the most important "V," and big data often has value in the insight and pattern recognition that lead to more effective operations, greater customer relationships, and other concrete and quantifiable corporate advantages.

**Variety:** There are many different types of data, including unstructured, semi-structured, and raw data.

The quantity of social media posts or search queries made in a day, hour, or other time frame are examples of data that businesses acquire, retain, and manage at a high velocity.

**Veracity:**The "truth" or accuracy of data and information assets, which usually affects the level of strength at the executive level.

**Importance of Data Analytics:**

Today, across all industries, big data analytics is the driving force behind everything we do online.

Take the music streaming service Spotify as an example. The 96 million customers of the company produce vast volumes of data every day. The cloud-based technology uses this data to automatically create new music using a clever recommendation engine that takes likes, shares, search history, and other factors into account. The methods, instruments, and frameworks developed as a result of big data analytics make this possible.If you use Spotify, you've definitely noticed the top recommendations section, which is determined by your preferences, past usage, and other variables. Utilizing a recommendation engine that utilises algorithms and data-gathering technologies for data filtering is beneficial.

**Applications of Big Data:**

Today, there is a tonne of data available worldwide. Let's examine big data applications after learning about the main characteristics of big data analytics. Large corporations use this information to grow their operations. Examining this data enables one to draw meaningful conclusions in the following instances:

**1. Keeping track of consumer spending and shopping patterns**

Large retailers' management teams must keep records of client spending trends, shopping patterns, and preferred products. Examples of these stores are Amazon, Walmart, Big Bazaar, and others (so that they can keep those products in the store). based on information about the product being searched.

**2.Recommendation**

Large retail outlets provide recommendations to customers by examining their shopping habits and buying trends. On e-commerce platforms like Amazon, Walmart, and Flipkart, product recommendations are provided. They monitor the products that clients are interested in and, using this data, recommend those things to them.

**3. Intelligent Traffic System**

Cameras positioned along the side of the road, at points of entry and exit to the city, and a GPS device put in the automobile are used to collect information about the movement of traffic on various roadways (Ola, Uber cab, etc.). The least time-consuming, jam-free methods are recommended once all of this data has been analysed. Smart city traffic systems can be developed via big data analysis.

Virtual personal assistant systems (like Siri on Apple devices, Cortana on Windows devices, and Google Assistant on Android) can answer to a range of client questions thanks to big data research. The user's location, local time, season, and other details relevant to the question asked, etc. are all recorded by this software. It analyses all of this information and provides a solution.

**4. IoT**

Manufacturing organisations incorporate IOT sensors inside their machines to collect operational data. By examining this data, it is possible to predict how long a machine will operate without the need for maintenance, enabling a business to take the necessary steps before the machine experiences a number of issues or stops working completely. It might not be essential to replace the entire machine as a result.

**5. Energy Sector**

Intelligent electric metres send data about power usage to a server every 15 minutes, where it is evaluated and used to determine when the city's power load is lowest. In order to reduce their electricity costs, it is advised for a family or an industrial facility to operate their heavy machinery at night, when power demand is lower.

# COMPARISON OF DATA SCIENCE AND DATA ANALYTICS APPROACHES

Although the two terms are sometimes used synonymously, data analytics is a division of data science. The term "data science" serves as a catch-all for all facets of data processing, including data gathering, modelling, and insights. Data analytics, on the other hand, focuses mostly on statistics, arithmetic, and statistical analysis. While data science is related to the broader picture around organisational data, it only focuses on data analysis. Most often, data scientists and data analysts collaborate to achieve shared business objectives. A data analyst might devote more time to routine analysis while generating consistent reports. The methods used to alter, store, and analyse data may be created by a data scientist. To put it simply, a data analyst interprets current data, whereas a data scientist develops new techniques.

* **Difference between Data Science and Data Analytics**

| **Feature** | **Data Science** | **Data Analytics** |
| --- | --- | --- |
| Coding Language | Python is the most commonly used language for data science along with the use of other languages such as C++, Java, Perl, etc. | The Knowledge of Python and R Language is essential for Data Analytics. |
| Programming Skills | In- depth knowledge of programming is required for data science. | Basic Programming skills is necessary for data analytics. |
| Use of Machine Learning | Data Science makes use of machine learning algorithms to get insights. | Data Analytics doesn’t makes use of machine learning. |
| Other Skills | Data Science makes use of Data mining activities for getting meaningful insights. | Hadoop Based analysis is used for getting conclusions from raw data. |
| Scope | The scope of data science is large. | The Scope of data analysis is micro i.e., small. |
| Goals | Data science deals with explorations and new innovations. | Data Analysis makes use of existing resources. |
| Data Type | Data Science mostly deals with unstructured data. | Data Analytics deals with structured data. |
| Statistical Skills | The statistical skills are necessary in the field of Data Science.. | The statistical skills are of minimal or no use in data analytics. |

* **Similarities between Data Science and Data Analytics**

Both use big data and data science to improve company or societal outcomes.

Both call for a foundation in mathematics, statistics, and programming (Hadoop, R, SAS, SQL, and Python). A data scientist should be knowledgeable about business.

# CONCLUSION

Science and business are both being revolutionized by the rising subject of data science. Data-driven work is growing more prevalent across almost all industries, which has an impact on both the employment that are accessible and the skills that are needed. More facets of the economy, society, and everyday life will depend on data as more data and analytical methods become available. Teachers, administrators, and students must immediately start thinking about the best ways to get ready for and keep up with this data-driven era of the future. Data analysis is the process of examining data in order to find relevant information that is congruent with the study's goal. Inspection, modification, modelling, and transformation of data are all included in data analysis depending on the research topic. The final inference from the data analysis, literature review, and findings is the conclusion.

##### REFERENCES

1. Dhar, V. (2013). "Data science and prediction". Communications of the ACM 56.
2. Jeff Leek (2013-12-12). "The key word in 'Data Science' is not Data, it is Science". Simply Statistics.
3. Shuhui Jiang, Xueming Qian, Tao Mei, Yun Fu, Personalized Travel Sequence recommendation on Multisource Big Social Media, 2016, IEEE Transactions on Big Data,Vol.2, Issue:1
4. Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, Understandable Big Data: A Survey , Computer Science Review, 2015, Vol: 17, pp: 71-80
5. Peter Naur: Concise Survey of Computer Methods, 397 p. Studentlitteratur, Lund, Sweden, ISBN 91-44-07881-1, 1974
6. Challenges in Data Science: A Comprehensive Study on Application and Future Trends, International Journal of Advance Research in Computer Science and Management Studies, Vo0lume 3, Issue 8, August 2015, ISSN: 232 7782.

[7] H.V. Jagadish, D. Agarwal, P.Bernstein, Challenges and Opportunities in Big Data, The Community Research Association, 2015.

[8]http://www.techrepublic.com/blog/big-data-analytics/10-emerging-technologies-for-big-data.

[9] Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. J Big Data. 2015;2(1):1–32.

[10] Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci. 2021;2(3):1–2