**ASSIGNMENT**

**Q:-1 Define Data Science and Big Data**.

**ANS:-** Data science is about using data to answer questions and solve problems. It helps people find useful information in large amounts of data, to make better decisions. Big data refers to large and complex data sets that are difficult to process and analyze using traditional data processing tools.

**Q:-2** **Elaborate the points based on Data science and Big data respectively**.

1. Use of technology
2. Benefits, feature and scope

**ANS:-** **DATA SCIENCE**

1. **Use of technology**.

Data science heavily relies on tech to collect, process, analyze, and interrupt large amounts of data tools such as:-

* Programing Languages: Python, R, and SQL are commonly used to clean, process and analyze data.
* Data Analysis Tools: Machine learning libraries and statistical software (like SAS) are used to build models and predict trends.
* Data Visualization Tools: Tools like Tableau and Power BI are used to create charts, graphs, and dashboards that help communicate findings clearly.
* Cloud Computing: Platforms like AWS, Google cloud, and Azure allow data scientists to store and process vast amounts of data quickly.

1. **Benefits**

Data science allows organizations to make data driven decision, leading to more accurate predictions, improved customer experiences, and cost savings. It helps in automating tasks, optimizing processes, and uncovering hidden insights.

1. **Features**

It involves data cleaning, statistical analysis, machine learning, and data visualization. Data science is adaptable across various industries, from healthcare and finance to retail and entertainment.

1. **Scope**

The scope of data science is broad and constantly expanding. It includes areas like predictive analysis, natural language processing, AI and deep learning. Data science is becoming crucial for industries like finance (for fraud detection), healthcare (for personalized medicine), marketing (for customer targeting), and more.

**BIG DATA**

1. **Use of technology.**

Big Data requires advanced technologies to store, process and analyze massive datasets that traditional systems can’t handle. Technologies used in big data include:-

* No SQL Databases: Databases like Cassandra and Mongo DB are designed to handle unstructured and semi-structured data.
* Cloud Storage: Amazon S3, Google Cloud Storage and other cloud platforms are crucial for storing the massive amounts of data that big data involves.

1. **Benefits:**

Big Data allows businesses to analyze huge volumes of information quickly, leading to better decision-making, more personalized services, and competitive advantages. It can help identify trends, improve operational efficiency, and predict future outcomes.

1. **Features:**

Big Data is characterized by the “3Vs”- Volume, Velocity and Variety. It deals with large-scale, rapidly generated data that comes from diverse sources, like social media, sensors, and mobile devices.

1. **Scope:**

Big data is transforming industries like retail (through customer insights), healthcare (for medical research and patient care), finance (risk management and fraud detection), and government (policy making and resources allocation),

**Q:-3 Define types of data:-**

1. **Structured data. 7. Audio data.**
2. **Semi structured data. 8. Video data.**
3. **Unstructured data. 9. Image data.**
4. **Graph based data. 10. Streaming data.**
5. **Machine generated data. 11. Natural language.**
6. **Network data.**

**ANS:-**

1. **Structured data:** Data organized in a predefined format like rows and columns (e.g., databases, spreadsheets).
2. **Semi- Structured data:** Data that doesn’t follow a strict structure but contains tags or markers to separate elements (e.g. JSON, XML files).
3. **Unstructured data:** Data without any predefined structure, making it harder to organize (e.g., text documents, images, videos).
4. **Graph based data:** Data represented as nodes and edges, used to model relationships between entities(e.g., social networks, knowledge graphs).
5. **Machine based data:** Data generated by machines or sensors often used in IOT applications (e.g. Temperature, readings, machine logs).
6. **Network data:** Data that describes the structures and connections within a network. (e.g. computer networks, social networks).
7. **Audio data:** Sound-based data, usually represented as waveformsor digital audio files (e.g., ,usic, podcasts).
8. **Video data:** Moving image data, often stored in formats like MP4 or AVI (e.g., movies, video clips).
9. **Image data:** Static visual data, stored as pixels in formats like JPEG or PNG (e.g. photos, illustrations).
10. **Streaming data:** Continuous flow of data generated in real time (e.g. live video feeds, financial tickers).
11. **Natural language:** Human language data, typically in the form of text or speech, used in natural language processing (NLP) e.g. spoken conversations, text documents).

**Q:-3 Discuss data science process:-**

**ANS:-** The data science process follows a systematic approach to solving problems using data:

1. **Data collection:** The first stepsinvolves gathering revelant data from various sources like databases, web scraping, API’s, sensors or public datasets.
2. **Store data:** After collecting data, it needs to be stored in a way that is both secure and easily accessible for further processing.
3. **Data pre-processing:** This step involves cleaning and preparing raw data for analysis. It’s essential to ensure that the data is in the right form at and is free of errors.
4. **Machine learning algorithms:** Machine learning algorithms are used to create predictive modles based on the prepared data. The model learns patterns in the data and can make predictions on new, unseen data,

* **Train data:** A subset of the data used to train the machine learning model, teaching it the relationships within the data.
* **Test data:** A separate subset of data used to evaluate how well the model performs on new, unseen data.

1. **Data analysis:** Data analysis involves interpreting the results from the machine learning model and extracting meaningful insights. This stage often involves statistical analysis, hypothesis testing, and data visualization.
2. **Data employment:** Once a model is built and evaluated, it is deployed into a production environment to make real-time predictions or decisions. This mean integrating the model into an application or system.

**Q:-3 Introduce:-**

* 1. **AI**
  2. **ML**
  3. **DS**
  4. **Applications of DS**
  5. **History of AI**

**ANS:-** **1.** **AI:** AI refers to the simulation of human intelligence in machines that are programmed to think, learn and solve problems. AI systems can perform tasks that typically require human intelligence, such as visual perception, decision making and language translation.

Example: - Virtual assistants like Siri or Alexa, self-driving cars.

1. **ML (MACHINE LEARNING):** A subset of AI, machine learning is the study of computer algorithms that improve automatically through experience. In ML, systems use data to learn patterns and make decisions with minimal human intervention.

Example: - Spam filtering in email, predictive text suggestions, image recognition software.

1. **DS (DATA SCIENCE):** Data science is a field that uses scientific methods, processes, algorithms, and systems to extract insights and knowledge from data. It combines elements of statistics, mathematics, programming and domain expertise to solve complex data-related problems. Example: - Predictive analytics, customer segmentation or detecting fraud using large datasets.
2. **Applications of DS (DATA SCIENCE):** Applications include predictive modeling, recommendation engines, healthcare diagnostics, and financial analysis. Example: -

* **Finance**: Fraud detection, credit risk modeling.
* **Retail:** Personalized recommendations, demand forecasting.
* **Healthcare:** Predictive models for disease, image analysis in medical diagnostics.
* **Marketing:** Customers segmentation, churn predication.

1. **History of AI:**

* **1950’s:** Alan Turning’s concept of a “Turning test” and the development of the first AI programs.
* **1960’s – 70’s:** Early AI research focused on symbolic AI and rule-based systems.
* **1980’s:** Introduction of neural networks and the back propagation algorithm.
* **2000’s:** Present: growth of machine learning, deep learning, and AI applications like voice recognition, natural language processing, and robotics.