**Assignment 01**

1. What is Data Science?

* Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate insights which analysts and business users can translate into tangible business value.
* Data science incorporates various disciplines -- for example, data engineering, data preparation, data mining, predictive analytics, machine learning and data visualization, as well as statistics, mathematics and software programming.

1. What is Artificial Intelligence?

* Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.
* **ARTIFICIAL INTELLIGENCE DEFINED: FOUR TYPES OF APPROACHES:**

1. Thinking humanly: mimicking thought based on the human mind.
2. Thinking rationally: mimicking thought based on logical reasoning.
3. Acting humanly: acting in a manner that mimics human behavior.
4. Acting rationally: acting in a manner that is meant to achieve a particular goal.
5. What is Machine Learning?

* Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data. Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications. More than half of the patients with cancer receive ionizing radiation (radiotherapy) as part of their treatment, and it is the main treatment modality at advanced stages of local disease. Radiotherapy involves a large set of processes that not only span the period from consultation to treatment but also extend beyond that to ensure that the patients have received the prescribed radiation dose and are responding well. The degrees of the complexity of these processes can vary and may involve several stages of sophisticated human-machine interactions and decision making, which would naturally invite the use of machine learning algorithms into optimizing and automating these processes including but not limited to radiation physics quality assurance, contouring and treatment planning, image-guided radiotherapy, respiratory motion management, treatment response modeling, and outcomes prediction. The ability of machine learning algorithms to learn from current context and generalize into unseen tasks would allow improvements in both the safety and efficacy of radiotherapy practice leading to better outcomes.
* **3 types of machine learning:**

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Future of Data Science and Real life applications.

* Data Science encompasses many breakthrough tech concepts like Artificial Intelligence, Internet of Things, Deep Learning to name a few. With its progress and technological developments, data science’s impact has increased drastically.The importance of gathering and collecting data is crucial as it enables retailers to determine and thus influence our purchasing habits. Hence, it exercises major control through its purchasing power.
* **Future Scope of Data Science**

1. Companies’ Inability to handle data
2. Companies’ Inability to handle data
3. Data Science is constantly evolving
4. An astonishing incline in data growth
5. Virtual Reality will be friendlier
6. Blockchain updating with Data science

* **Real Life Application:**

1. Healthcare
2. Targeted Advertising
3. Website Recommendations
4. E-Commerce
5. Transport
6. Text and Advanced Image Recognization
7. Gaming
8. Security
9. Customer Insights
10. Augmented Reality
11. Why Python so popular? Future of Python.

* Python is one of the most dynamic and adaptable programming languages in the industry today. Since its inception in the 1990s, Python has become the most popular programming language and even today thousands are learning this Object-Oriented Programming language. If you are new to programming, you have probably heard about the current hype Python has generated due to its features, and you are probably wondering what makes Python the leading back-end language in 2022.
* **Python Popularity:**

1. Python is easy to learn
2. Python has an active, supportive community
3. Python is flexible
4. Python offers versatile web-development solutions
5. Python is well suited to data science and analytics
6. Python is efficient, fast, and reliable
7. Python is widely used with IoT Technology
8. Python empowers custom automation
9. Python is the academic language

* **Future of Python:**

1. Python is being used worldwide as a wide range of application development and system development programming language. Huge brands and search engine giants are using python to make their task easier. Google, Yahoo, Quora, Facebook are using python programming to solve their complex programming problems. Therefore, you should start preparing soon and build your concepts in Python. You can take up the Vskills Certified Python Developer that will help you build your basics in Python. Moreover, this Vskills Certification in Python Programming develops your skills on various areas like Data Structures, Object Oriented Python, Working with XML, Working with Files & Working with Modules etc.
2. Why Python for data analytics?

* Python is a popular multi-purpose programming language widely used for its flexibility, as well as its extensive collection of libraries, which are valuable for analytics and complex calculations.
* Python’s extensibility means that it has thousands of libraries dedicated to analytics, including the widely used Python Data Analysis Library (also known as Pandas).
* For the most part, data analytics libraries in Python are at least somewhat derived from the NumPy library, which includes hundreds of mathematical calculations, operations, and functions.

1. Importance of Python in machine learning.

* Python offers all the skillsets that are required for a machine learning or AI project – stability, flexibility and a large number of tools. Python helps developers to be productive and confident about the product that they are manufacturing, from the stages of development to deployment and till the maintenance stage.
* Simple and consistent.
* Libraries and Frameworks.
* Platform Independence.
* Great Community Base.

1. Importance of Python in deep learning & NLP.

* Python offers concise and readable code. While complex algorithms and versatile workflows stand behind machine learning and AI, Python’s simplicity allows developers to write reliable systems. Developers get to put all their effort into solving an ML problem instead of focusing on the technical nuances of the language.
* Additionally, Python is appealing to many developers as it’s easy to learn. Python code is understandable by humans, which makes it easier to build models for machine learning.
* Many programmers say that Python is more intuitive than other programming languages. Others point out the many frameworks, libraries, and extensions that simplify the implementation of different functionalities. It’s generally accepted that Python is suitable for collaborative implementation when multiple developers are involved. Since Python is a general-purpose language, it can do a set of complex machine learning tasks and enable you to build prototypes quickly that allow you to test your product for machine learning purposes.
* **Extensive selection of libraries and frameworks:**

1. Keras, TensorFlow, and Scikit-learn for machine learning
2. NumPy for high-performance scientific computing and data analysis
3. SciPy for advanced computing
4. Pandas for general-purpose data analysis
5. Seaborn for data visualization
6. Supervised Learning with examples.

* Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically supervised learning is when we teach or train the machine using data that is well labelled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labelled data.
* For instance, suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all the different fruits one by one like this:

1. If the shape of the object is rounded and has a depression at the top, is red in color, then it will be labeled as –Apple.
2. If the shape of the object is a long curving cylinder having Green-Yellow color, then it will be labeled as –Banana.

* Now suppose after training the data, you have given a new separate fruit, say Banana from the basket, and asked to identify it.
* Supervised learning deals with or learns with “labeled” data. This implies that some data is already tagged with the correct answer.
* **Types of Supervised Learning:**

1. Regression
2. Logistic Regression
3. Classification
4. Naive Bayes Classifiers
5. K-NN (k nearest neighbors)
6. Decision Trees
7. Support Vector Machine
8. Unsupervised learning with examples.

* Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.
* Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.
* For instance, suppose it is given an image having both dogs and cats which it has never seen. Thus the machine has no idea about the features of dogs and cats so we can’t categorize it as ‘dogs and cats ‘. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having dogs in them and the second part may contain all pics having cats in them. Here you didn’t learn anything before, which means no training data or examples.
* **Types of Unsupervised Learning:**

1. Clustering
2. Exclusive (partitioning)
3. Agglomerative
4. Overlapping
5. Probabilistic
6. Clustering Types:-
7. Hierarchical clustering
8. K-means clustering
9. Principal Component Analysis
10. Singular Value Decomposition
11. Independent Component Analysis

**Jobs:**

1. Role and responsibilities of a Database Engineer.

* Database engineers design and maintain databases, increase data storage capacity, and troubleshoot database functionality issues. They also review database performance and user reports.
* **Database Engineer Responsibilities:**

1. Maintaining and enhancing the performance of existing database programs.
2. Assisting database development teams in designing new database programs that meet the organization's data storage needs.
3. Monitoring databases and related systems to ensure optimized performance.
4. Writing new support programs and scripts to increase data storage capacity.
5. Reviewing database and user reports, as well as system information.
6. Performing debugging procedures on database scripts and programs, as well as resolving conflicts.
7. Mentoring database administrators and providing them with technical support.
8. Adhering to best practices in securely storing, backing up, and archiving data.
9. Documenting processes related to database design, configuration, and performance.
10. Keeping abreast of developments and best practices in database engineering.
11. Role and responsibilities of a Data Analyst.

* Data analysts work with data to help their organizations make better business decisions. Using techniques from a range of disciplines, including computer programming, mathematics, and statistics, data analysts draw conclusions from data to describe, predict, and improve business performance. They form the core of any analytics team and tend to be generalists versed in the methods of mathematical and statistical analysis.
* **Data analyst responsibilities:**

1. Analyzing data using statistical techniques and providing reports
2. Developing and implementing databases and data collection systems
3. Acquiring data from primary and secondary sources and maintain data systems
4. Identifying, analyzing, and interpreting trends or patterns in complex data sets
5. Filtering and cleaning data
6. Working with management to prioritize business and information needs
7. Locating and defining new process improvement opportunities
8. Role and responsibilities of a Data Engineer.

* Data engineers are vital members of any enterprise data analytics team, responsible for managing, optimizing, overseeing, and monitoring data retrieval, storage, and distribution throughout the organization.
* **Data Engineer responsibilities:**

1. Analyze and organize raw data
2. Build data systems and pipelines
3. Evaluate business needs and objectives
4. Interpret trends and patterns
5. Conduct complex data analysis and report on results
6. Prepare data for prescriptive and predictive modeling
7. Build algorithms and prototypes
8. Combine raw information from different sources
9. Explore ways to enhance data quality and reliability
10. Identify opportunities for data acquisition
11. Develop analytical tools and programs
12. Collaborate with data scientists and architects on several project
13. Role and responsibilities of a Data Scientist.

* Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate insights which analysts and business users can translate into tangible business value.
* **Data scientist responsibilities:**

1. Identifying relevant data sources for business needs
2. Collecting structured and unstructured data
3. Sourcing missing data
4. Organising data in to usable formats
5. Building predictive models
6. Building machine learning algorithms
7. Enhancing the data collection process
8. Processing, cleansing & verifying of data
9. Analysing data for trends and patterns and to find answers to specific questions
10. Setting up data infrastructure
11. Develop, implement and maintain databases
12. Assess quality of data and remove or clean data
13. Generating information and insights from data sets and identifying trends and patterns
14. Preparing reports for executive and project teams
15. Create visualisations of data
16. Role and responsibilities of a Machine Learning Engineer.

* Machine learning is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence.
* **Machine Learning responsibilities:**

1. Study and transform data science prototypes
2. Design machine learning systems
3. Research and implement appropriate ML algorithms and tools
4. Develop machine learning applications according to requirements
5. Select appropriate datasets and data representation methods
6. Run machine learning tests and experiments
7. Perform statistical analysis and fine-tuning using test results
8. Train and retrain systems when necessary
9. Extend existing ML libraries and frameworks
10. Keep abreast of developments in the field