

COURSE NAME:

Data Science & Machine Learning with Python

Batch: 13

Data Science Discussion - Assignment 01

Submitted By:

Md. Hasinur Rahman

Submitted To:

Rashedul Alam Shakil

Short Questions

Q1. What is Data Science?

Data Science is a multi-disciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data to extract meaningful insights for business.

Q2. 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.

Q3. What is 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 and gradually improving its accuracy.

Q4. Future of Data Science & Real Life Applications.

The future of data science is bright, and with increased applications across domains, its prospects are immense. Along with artificial intelligence and machine learning, data science will contribute towards a higher level and intelligent decision-making for businesses.

Real Life Application of Data Science:

- a) Video on You Tube: Youtube automatically suggests the related video when we watch any type of video.
- b) Social media: Friend suggestion on Facebook.
- c) Uber employ Data Science to improve price and delivery routes, as well as optimal resource allocation, by combining numerous factors such as consumer profiles, geography, economic indicators, and logistical providers.
- d) Banks, for example, utilize sophisticated machine-learning algorithms to detect fraud based on a user's usual financial activity.
- e) Advertisement on various websites uses data science to show related add to the viewer.

Q5. Why Python is so popular? Future of Python.

The reasons for Python popularity are as follows:

- Python is easy to learn
- Python has an active, supportive community
- Python is flexible
- Python offers versatile web-development solutions
- Python is well suited to data science and analytics
- Python is efficient, fast, and reliable
- Python is widely used with IoT Technology
- Python empowers custom automation
- Python is the academic language

Future of Python:

Python is growing as a rapid pace. As per the recent study and career research, it was mentioned that there are more or less 40,000 jobs that are completely for python experienced software developers in many countries. Besides, search tags are also higher for python on different platforms. It is said that Data science collaborated with python has a great job and upcoming future taking one to a new edge.

Q6. Why python for data analytics?

Python is good for data analytics because Python works well on every stage of data analysis. There are many libraries that make Python more preferred for data analysis. For example, Scrapy and BeautifulSoup libraries are used for data mining, NumPy and Pandas are used for Data Processing & Modelling and Matplotlib & Seaborn are widely used for Python data visualization.

Q7. Importance of Python on Machine Learning?

Python language comes with many libraries and frameworks that make coding easy and saves a significant amount of time. The most popular libraries are NumPy, which is used for scientific calculations; SciPy for more advanced computations; and scikit, for learning data mining and data analysis. These libraries work alongside powerful frameworks like TensorFlow, CNTK, and Apache Spark. These libraries and frameworks are essential when it comes to Machine Learning projects.

Q8. Importance of Python in Deep Learning and NLP?

Python provides huge number of libraries and Framework for Deep learning and NLP.

Deep Learning: For Deep Learning Python comes with TensorFlow, Keras a deep learning API written in Python, PyTorch, fastai and many other libraries for Deep Learning.

Natural Language Processing(NLP): Python Comes with FastText for efficient learning of word representations and sentence classification, spaCy for Industrial-strength Natural Language Processing (NLP), Gensim for topic modelling, document indexing and similarity retrieval with large corpora and many other library for NLP.

Q9. Supervised Learning with Examples?

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output. It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, et

In the real-world example of supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

Q10. Unsupervised Learning with Examples?

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.

Some real world examples of unsupervised learning are:

- a) Customer segmentation or understanding different customer groups around which to build marketing or other business strategies.
- b) Genetics, for example clustering DNA patterns to analyze evolutionary biology.
- c) Recommender systems, which involve grouping together users with similar viewing patterns in order to recommend similar content.
- d) Anomaly detection, including fraud detection or detecting defective mechanical parts (i.e., predictive maintenance).

JOBS

Q11. 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.

Roles & Responsibilities of a Database Engineer:

- Recommend and implement solutions for performance monitoring and tuning.
- Review work of less senior team members, provide guidance from technical perspective.
- Analyze problems, anticipate future problem areas, and implement solutions.
- Perform environment setup and configuration, proactive monitoring and maintenance.
- Lead development of database structures that fit into the overall architecture of the system under development.
- Work with architects and lead developers to come up with optimal database design.
- Engage in capacity planning exercises to ensure adequate system resources.
- Support the development of long and short term requirements for database administration and design.
- Collaborate with the technical teams and the EHR vendor to ensure proactive support.
- Participate in your change control process for all planned application and technical activities.
- Work with report writers to provide data needed for reports.
- Take on full roadmap items, work with other functional teams, and be able to deliver high quality results on time.
- Investigate and find the root cause for software problems reported by clients.
- Direct organization of requirements and data into a usable database schema by directing development of ad hoc queries, scripts, macros, updates to existing queries.
- May perform database administration and maintenance, including database installation and configuration, backups, upgrades, patching.
- Document schema design in E-R diagrams using MS Visio and maintain data dictionaries for all tables.
- Review SQL code written by application developers to ensure compliance to coding standards and best practices as well as maximum performance.

- Create deployment and rollback scripts for all database objects manually or auto generated.
- Analyze access patterns and propose the best combination of indexes, constraints, foreign keys, and queries.

Q12. Roles 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.

Roles and Responsibilities of a Data Analyst:

- Using automated tools to extract data from primary and secondary sources.
- Removing corrupted data and fixing coding errors and related problems.
- Developing and maintaining databases, data systems – reorganizing data in a readable format.
- Performing analysis to assess quality and meaning of data.
- Filter Data by reviewing reports and performance indicators to identify and correct code problems.
- Using statistical tools to identify, analyze, and interpret patterns and trends in complex data sets that could be helpful for the diagnosis and prediction.
- Assigning numerical value to essential business functions so that business performance can be assessed and compared over periods of time.
- Analyzing local, national, and global trends that impact both the organization and the industry
- Preparing reports for the management stating trends, patterns, and predictions using relevant data
- Working with programmers, engineers, and management heads to identify process improvement opportunities, propose system modifications, and devise data governance strategies.
- Preparing final analysis reports for the stakeholders to understand the data-analysis steps, enabling them to take important decisions based on various facts and trends.

Q13. Roles and Responsibilities of a Data Engineer?

A Data Engineer is a data professional who uses their expertise in data engineering and programming to build systems that collect, manage and convert raw data into usable information for business analysts

Roles and Responsibilities of a Data Engineer:

- **Work on Data Architecture:** systematic approach to plan, create, and maintain data architectures while also keeping it aligned with business requirements.
- **Collect Data:** Before initiating any work on the database, they have to obtain data from the right sources. After formulating a set of dataset processes, data engineers store optimized data.
- **Conduct Research:** Data Engineers conduct research in the industry to address any issues that can arise while tackling a business problem.
- **Improve Skills:** Data Engineers don't rely on theoretical database concepts alone. They must have the knowledge and prowess to work in any development environment regardless of their programming language. Similarly, they must keep themselves up-to-date with machine learning and its algorithms like the random forest, decision tree, k-means, and others.
- **Create Models and Identify Patterns:** Data Engineers use a descriptive data model for data aggregation to extract historical insights. They also make predictive models where they apply forecasting techniques to learn about the future with actionable insights. Likewise, they utilize a prescriptive model, allowing users to take advantage of recommendations for different outcomes. A considerable chunk of a data engineer's time is spent on identifying hidden patterns from stored data.
- **Automate Tasks:** Data Engineers dive into data and pinpoint tasks where manual participation can be eliminated with automation.

Q14. Role & Responsibilities of a Data Scientist?

Data scientists utilize their analytical, statistical, and programming skills to collect, analyze, and interpret large data sets and use this information to develop data-driven solutions to difficult business challenges.

Data scientist roles and responsibilities include:

- Data mining or extracting usable data from valuable data sources
- Using machine learning tools to select features, create and optimize classifiers
- Carrying out preprocessing of structured and unstructured data
- Enhancing data collection procedures to include all relevant information for developing analytic systems
- Processing, cleansing, and validating the integrity of data to be used for analysis

- Analyzing large amounts of information to find patterns and solutions
- Developing prediction systems and machine learning algorithms
- Presenting results in a clear manner
- Propose solutions and strategies to tackle business challenges
- Collaborate with Business and IT teams

Q15. Role and Responsibilities of a Machine Learning Engineer?

The job of a Machine Learning Engineer is quite similar to that of a Data Scientist, in the sense that both roles involve working with vast volumes of data.

Roles & Responsibilities of a Machine Learning Engineer:

- To study and convert data science prototypes.
- To design and develop Machine Learning systems and schemes.
- To perform statistical analysis and fine-tune models using test results.
- To find available datasets online for training purposes.
- To train and re-train ML systems and models as and when necessary.
- To extend and enrich existing ML frameworks and libraries.
- To develop Machine Learning apps according to customer/client requirements.
- To research, experiment with, and implement suitable ML algorithms and tools.
- To analyze the problem-solving capabilities and use-cases of ML algorithms and rank them by their success probability.
- To explore and visualize data for better understanding and identify differences in data distribution that could impact model performance when deploying it in real-world scenarios.