PMP Planet Multimedia Publishers
The Ultimate Resource

DATA SCIENCE

8

Adheres to NEP 2020 Guidelines

ANSWER KEY

1 Introduction to Data

Worksheet – 1.1

One Word Answer.

1. Qualitative
2. Quantitative
3. Discrete
4. Continuous

Activity 1.1: Discrete or Continuous

a) Discrete
b) Discrete
c) Continuous
d) Continuous
e) Continuous
f) Continuous
g) Discrete
h) Continuous

Activity 1.2: Qualitative or Quantitative

a) Numerical
b) Categorical
c) Numerical
d) Categorical
e) Numerical
f) Categorical

Activity 1.3: Real World Applications of Data

a) ✔
b) ✘
c) ✔
d) ✔
e) ✔

Worksheet – 1.2

Objective Problems.

1. c) Both a) and b)
2. a) Correctly attended questions
3. d) All of the above
4. b) Qualitative data

Assessment Time

A. Multiple Choice Questions.

1. a) Data are the raw facts
2. b) Categorical data
3. a) Categorical data
4. a) Qualitative data
5. d) Qualitative
6. d) All of these
7. c) 890
8. d) All of these

PMP

B. Fill in the Blanks.

1. Data Analysis
2. Forms
3. Measurement
4. Qualitative
5. Nominal data, Ordinal data
6. Quantitative data
7. Continuous
8. Google Maps

C. State 'T' for True and 'F' for False Statements.

1. F
2. T
3. T
4. F
5. F
6. T
7. T
8. T

D. Answer in One Word.

1. Qualitative data
2. Quantitative data
3. Discrete data
4. Continuous data

E. Answer the Following Questions.

1. Data are the raw facts used to derive useful meaning or information. The real-world applications of data are:

   a) Wearable devices Sensors in the Healthcare Industry: Nowadays, hospitals are fully equipped with wearable devices and sensors which can provide real-time feed to the electronic health record of a patient. The data collected from those sensors help the doctors to treat the patient accordingly.

   b) Barcode Reader at Billing Counter in Supermarkets: Barcode Reader is equipped with a laser beam light that translates the binary code of every Universal Product Code into its associated price. Each UPC price has been stored in the store's database. Here, UPC and the associated price are the forms of data.

2. **Qualitative data:** Qualitative data is a type of data that describes a piece of information. For example, "What a great leader he is?". Sometimes, it is also known as Categorical data as it is based on properties, attributes, labels, and other identifiers.

   Quantitative data: Quantitative data is the type of data whose value is measured in the form of numbers or counts. For example: '1', '5.67', '365. Sometimes, it is also known as numerical data as it is a group of quantifiable information that can be used for mathematical computations and statistical analysis.

PMP

3. **Discrete data:** Discrete data is a count that involves integers only. For example, the Number of Students and teachers in a school etc. Discrete data are used for simple statistical analysis as its computation or calculation is easy.

   Continuous data: Continuous data are not considered clean integers as it is usually collected from very precise measurements. This kind of data is measured using specific tools as it changes very rapidly. For example, the weather temperature, amount of time required to complete a task, Speed of cars etc.

4. Nowadays, data is the utmost and prime necessity of these industries.

   a) **Transportation:** Google maps is a popular mobile application used for navigating from source to destination as it helps us locate the least traffic routes by predicting traffic-affected areas using real-time data processing.
   Ride-Sharing Apps like Uber and Ola generate and use huge amounts of data regarding drivers, their vehicles, locations etc. The collected data will be further analysed to predict supply, demand, location of drivers and fares.

   b) **Entertainment:** In the entertainment industry, data is used to recommend movies, web series or TV shows according to the taste and preferences of the customer. This data might be collected from search history and viewed the history of an individual account.

   c) **E-commerce:** An e-commerce company uses a recommendation engine to advertise its products according to the taste and preferences of the customer. This recommendation engine is completely based on data.

5. Yes, data exists in numerous forms as it is collected from multiple sources in different types. Some of the common forms are:
   a) Text
   b) Images
   c) Videos
   d) Numbers
   e) Spreadsheets
   f) Sound

PMP

F. Application Based Questions.

1. Quantitative data.
2. Data exist in numerous forms like Text, Numbers, Images, Video, etc.
3. The two general types of qualitative data are Nominal and Ordinal Data. The two general types of quantitative data are Discrete and Continuous.

G. High Order Thinking Skills.

- Do it Yourself.

PMP

2 Introduction to Data Science

Activity 2.1: Based on Data Science

a) Yes, speech recognition applications like Siri and Cortana are great applications of data science because these applications are dependent on the data which are being fed by the user.

b) Weather, YouTube, Gaana, etc.

Worksheet 2.1

State 'T' for True and 'F' for False Statements.

1.F 2.T 3.F 4.T 5.T

Activity 2.2: Data Science without Programming

The field of data science is not difficult if someone has no idea about computer programming because data science is a multidisciplinary field that uses scientific methods, processes, algorithms and systems to extract meaningful interpretations from the typically large data.

Activity 2.3: Career opportunities in Data Science

- Do it Yourself.

Worksheet 2.2

Identify the Following.

1. Database Developer
2. Senior Data Scientist
3. Data Architect
4. Data Scientist

PMP

Worksheet 2.3

Algorithms used in Data Science.

C	L	A	S	S	I	F	I	C	A	T	I	O	N	P
A	L	A	H	R	R	E	G	R	E	S	S	I	O	N
S	K	R	R	A	E	C	L	A	S	K	R	R	S	R
F	L	T	E	A	I	A	L	A	F	L	T	E	F	A
G	V	Y	A	R	N	S	K	R	G	V	Y	A	G	A
T	G	O	C	T	F	F	L	T	T	G	O	C	T	R
R	B	F	G	A	O	C	L	A	R	B	F	G	R	T
F	H	I	V	R	R	S	K	R	R	S	K	V	F	A
H	U	C	X	T	C	F	L	T	E	F	L	X	H	R
K	J	R	S	Y	E	G	V	Y	A	G	V	S	K	T
N	K	S	R	L	M	T	G	O	C	T	G	C	L	A
V	L	A	F	M	E	R	B	F	G	R	B	A	L	A
C	T	R	F	N	N	F	H	I	V	F	H	S	K	R
D	C	L	U	S	T	E	R	I	N	G	L	F	L	T

PMP

Assessment Time

A. Multiple Choice Questions

1. b) Data Science
2. d) Biology
3. c) Recommendation engine
4. b) Face recognition
5. c) Both a) and b)
6. b) Anomaly detection
7. d) Artificial intelligence
8. d) Data scientist
9. c) Multiclass Classification
10. a) Regression

B. Fill in the Blanks

1. Data	2. Data Science	3. Algorithms
4. Structured, Unstructured		5. Python
6. Machine Learning	7. Data Scientists	8. Databases
9. Database Designer	10. Regression	

C. State 'T' for True and 'F' for False Statements.

1. F	2. F	3. T	4. T
5. T	6. F	7. T	8. F
9. T	10. T		

D. Answer in One Word.

1. Data Science	2. Business Intelligence Analyst
3. Data Architect	4. Binary Classification
5. Anomaly Detection	

E. Answer the Following Questions.

1. Data science is a technical term used to analyse data for making decisions.

2. **Binary Classification:** Binary classification is the task of classifying the elements of a dataset into two groups based on a classification rule.
   Clustering: Clustering is the task of grouping a set of objects in such a manner that objects in the same group are more similar to each other than to those in other groups.

3. A classification algorithm is used to classify the elements of the data set into two or more than two groups whereas a Regression algorithm is used to predict the numerical value of a continuous variable based on historical data.

4. A data scientist is someone who can collect and analyse large datasets to make better and smarter decisions. He/ she uses the knowledge of every domain like Mathematics, Computer science and Statistics to process and analyse data in such a way that he or can find meaningful insights or interpretations.

5. The various career options available in the field of data science are:
   a) Data Scientist
   b) Business Intelligence Analyst

PMP

c) Data Mining Engineer
d) Data Architect
e) Senior Data Scientist
f) Database Developer

6. The various advantages of data science are:
   a) Data Science helps organizations to enhance profits and improve efficiency by selling their products at the right place and in the right time as per the tastes and preferences of the customer.
   b) Data Scientists help companies make smart business decisions in complex situations.

   In addition to the advantages, some of the disadvantages are:
   a) The field of data science is completely dependent on domain knowledge because a person with considerable background in Mathematics and Computer Science will find difficulty at some level to solve data science problems.
   b) The arbitrary data yields unexpected results which is considered one of the biggest disadvantages of Data Science.

7. E-commerce applications like Myntra, Navigation applications like Google Maps and Media streaming applications like YouTube.

8. a) **Data Scientist:** A data scientist has a prominent role in a company as he/she can explain the importance of data in a simpler way which can be easily understood by others.
   b) **Data Architect:** Data Architect is a person who is responsible to design and manage data systems within an organization. He/ She is also responsible for setting policies regarding storage and accessing of data and integrating new data technologies into existing IT infrastructures.
   c) **Business Intelligence Analyst:** A Business Intelligence Analyst is a person who is responsible for drafting and changing a company's strategy by assessing the market using data and determining the latest trends in the industry. The role of a Business Intelligence Analyst is more technical as it requires more knowledge of popular machines.

F. Application Based Questions.

1. Regression Algorithm 2.Business Intelligence Analyst 3. Clustering

G. High Order Thinking Skills.

- Do it Yourself

PMP

3 Data Visualization

Activity 3.1: Charts in Excel

- Do it Yourself

Worksheet 3.1: Representing Data in Graphical Form

Choose the Correct Option.

1. c) Charts
2. c) Data Visualization
3. c) Both a) and b)

Worksheet 3.2: Based on Charts

- Do it Yourself

Assessment time

A. Multiple choice questions.

1. d) Data Visualization
2. c) Both a) and b)
3. a) Charts
4. d) Line chart
5. d) Both a) and b)
6. d) No Uniqueness
7. c) Both a) and b)
8. c) Regression analysis
9. a) Cohort analysis
10. a) Power BI

B. Fill in the Blanks.

1. Data Visualization
2. comparisons, relationships
3. Qualitative, Quantitative
4. Many
5. Pie Chart
6. Garbage in Garbage Out
7. Actionable insights
8. Questions
9. Predictive analysis
10. Visual

C. State 'T' for True and 'F' for False Statements.

1. T
2. F
3. F
4. T
5. F
6. F
7. T
8. T
9. F
10. T

PMP

D. Answer in One Word.

1. Power BI
2. Pie Chart
3. Regression Analysis
4. Cohort
5. Google Data Studio

E. Answer the Following Questions.

1. The term "Data Visualization" refers to the process of displaying or representing data or information in a graphical or visual format like Charts, Graphs, Histograms, etc.

2. Data Visualization helps people to draw actionable insights from massive amounts of data in a short amount of time. Sometimes, an immense amount of data seems boring and laborious. In such a situation, data visualization is a great technique to maintain people's interest with the information they can easily understand.

3. The various kinds of Statistical techniques used for data visualization are:
   a) **Regression Analysis:** In general, a set of statistical methods used for estimating the relationships between a dependent variable and one or more independent variables is known as Regression analysis. It provides detailed insight that can be applied to further improve products and services.
   b) **Cohort Analysis:** Cohort analysis is a subset of behavioural analytics that breaks down the data in a data set into related groups or cohorts for analysis. These related groups, or cohorts, usually share common attributes or characteristics within a defined time range.
   c) **Predictive Analysis:** Predictive analytics is the process of making predictions about future outcomes using historical data combined with statistical modelling, data mining techniques and machine learning. Nowadays, companies are widely using predictive analytics to find patterns in this data to identify various types of risks and opportunities involved within.

4. Data must be collected from various sources by considering various factors like completeness, accuracy, format, etc. because expected outcomes will be achieved if the data is relevant. If data is of bad quality or not appropriate or incomplete, then the required output can never be achieved.

PMP

5. In banking and financial institutions, predictive analysis is used to analyse individual demographics as it finds patterns in the data set to identify various types of risks and opportunities involved within.

6. Yes, it is important to assess the end-users for visualization because end users are the most vital aspect of the entire data preparation method for analysis.

F. Application Based Questions.

1. Cohort Analysis
2. Power BI, Google Data Studio
3. Scatter Chart

G. High Order thinking Skills.

- Do it Yourself

PMP

4 Data Science and Artificial Intelligence

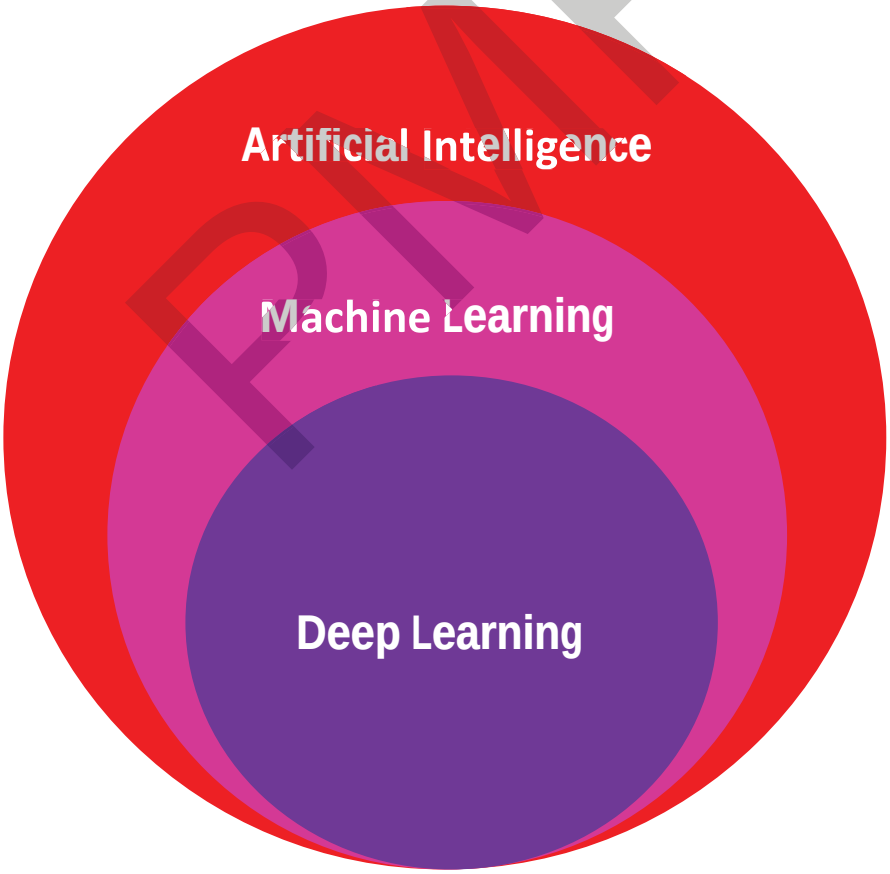
Worksheet 4.1: Applications of Image Recognition Technology

Here, some mobile applications are given. Put a tick mark in the front of the applications which use Image recognition technology.

1. ✔ 2. ✔ 3. ✔ 4. ✔ 5. ✘

Activity 4.1: Based on Relationship between AI, Deep Learning and Machine Learning

Here, four Venn Diagrams are given. You must select the appropriate one that clarifies the relationship between these three.



PMP

Assessment time

A. Multiple Choice Questions.

1. a) Data Science
2. d) All of these
3. c) Text Analytics
4. a) Tagging a friend on social networking site
5. c) Both a) and b)
6. d) All of these
7. a) Machine learning
8. d) All of these
9. c) Natural language processing
10. d) Predictive learning

B. Fill in the Blanks.

1. Data Science, Artificial Intelligence
2. Data Science
3. Price optimization
4. Predictive
5. Text
6. Text analytics
7. Artificial Intelligence
8. Machine Learning
9. Deep Learning
10. Tokenization

C. State 'T' for True and 'F' for False Statements.

1. T	2. F	3. T	4. F
5. T	6. T	7. F	8. T

PMP

D. Answer in One Word.

1. Text Analytics
2. Image Analytics
3. Google Photos
4. Biometrics System
5. Artificial Intelligence

E. Answer the Following Questions.

1. The two important applications of data science in the following sectors are:

   a) **Banking:**

   - ✓ Banks can easily segregate important any useful data using the various tools and techniques of data science.
   - ✓ Banks can easily predict customer behaviour using the previous history and credit reports of the customer at the time of sanctioning loans.

   b) **Transportation:**

   - ✓ Ride Sharing applications use data science for price optimization and providing better experiences to their customers.
   - ✓ Using data science, transport industries can easily spot and foresee the occurrences of traffic, accidents, or vehicle breakdown and suggest efficient responses.

   c) **E-commerce:**

   - ✓ E-commerce companies can recommend products and offer discount prices based on customers' past behaviour using data science.
   - ✓ Cohort analysis, one of the important data visualization techniques, is used to predict the customer retention rate for a particular product.

2. Yes, data science is a continuously evolving field and every industry uses various techniques of data science to cope with the competition in the market because millions of data are created within a millisecond and data is the prime necessity of every industry.

3. a) **Text Analytics:** Text Analytics refers to a process of extracting meaningful or valuable information from unstructured data into meaningful data. Sometimes, text analysis is used to find out specific information from large amounts of data. This specific information may be keywords, names, topics, categories, semantics, tags, etc.

   b) **Image Analytics:** Image analytics is a technical term used for extracting meaningful information that exists in the images using digital processing

PMP

techniques. In recent times, we have seen various applications of image analytics in our day-to-day life which are as follows:

- ✓ Tagging a friend on a social networking site.
- ✓ Medical Imaging
- ✓ Real-time Merchandising Recommendation.

4. Deep learning, machine learning and artificial intelligence are interrelated with each other as deep learning is a subset of machine learning, and machine learning is a subset of AI.

5. The main goals of Artificial Intelligence are:

   a) **Knowledge Representation:** Knowledge representation is a term used to describe information that a machine or computer can understand. With the help of AI, the process of Knowledge representation became easy as it makes computers capable of describing objects easily.

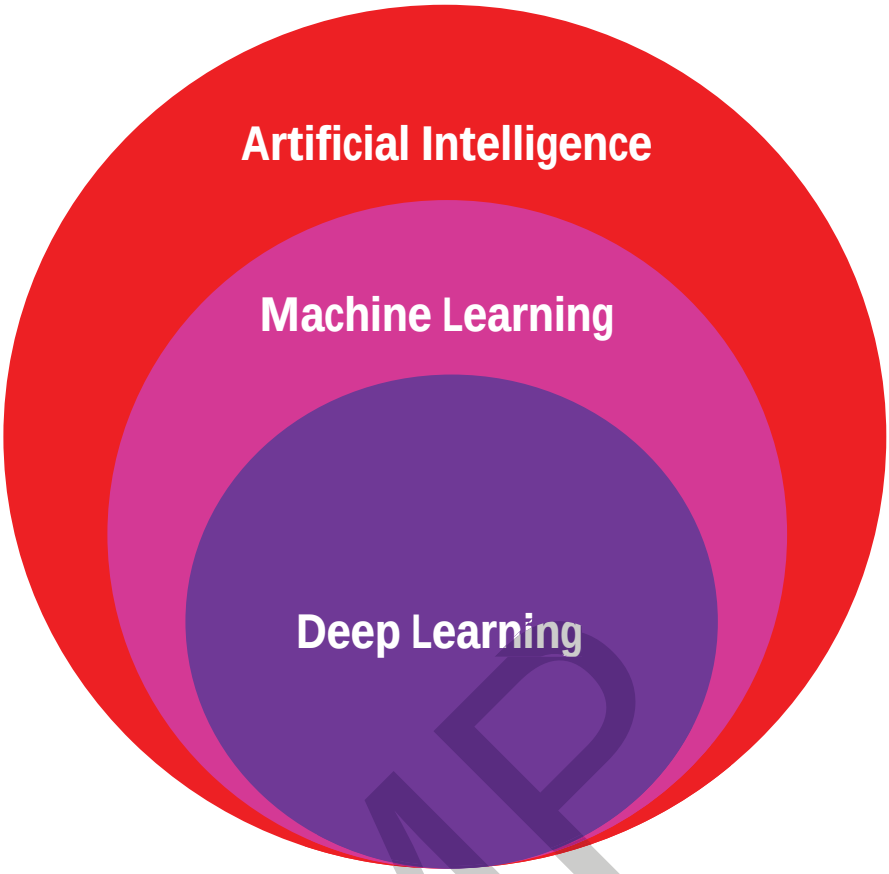
   b) **Development of intelligent Machine:** The development of intelligent machines that could learn on their own is the main goal of Artificial Intelligence. With AI, machines are independent of human intervention for feeding data into machines.

F. Application Based Questions.

1. The various applications of image analytics in our day-to-day life are:
   a) Tagging a friend on a social networking site
   b) Medical Imaging
   c) Real-time Merchandising Recommendation.
   d) ATM safety and security
   e) Google photos where photos can be categorized automatically.

PMP

2.



G. High Order Thinking Skills

- Do it Yourself

PMP