



DATA SCIENCE

Grade VIII

Version 1.0



DATA SCIENCE

GRADE VIII

Teacher Handbook



ACKNOWLEDGMENT

Patrons

- Sh. Ramesh Pokhriyal 'Nishank', Minister of Human Resource Development, Government of India
- Sh. Dhotre Sanjay Shamrao, Minister of State for Human Resource Development, Government of India
- Ms. Anita Karwal, IAS, Secretary, Department of School Education and Literacy, Ministry Human Resource Development, Government of India Advisory

Editorial and Creative Inputs

- Mr. Manuj Ahuja, IAS, Chairperson, Central Board of Secondary Education

Guidance and Support

- Dr. Biswajit Saha, Director (Skill Education & Training), Central Board of Secondary Education
- Dr. Joseph Emmanuel, Director (Academics), Central Board of Secondary Education
- Sh. Navtez Bal, Executive Director, Public Sector, Microsoft Corporation India Pvt. Ltd.
- Sh. Omjiwan Gupta, Director Education, Microsoft Corporation India Pvt. Ltd
- Dr. Vinnie Jauhari, Director Education Advocacy, Microsoft Corporation India Pvt. Ltd.
- Ms. Navdeep Kaur Kular, Education Program Manager, Allegis Services India

Value adder, Curator and Co-Ordinator

- Sh. Ravinder Pal Singh, Joint Secretary, Department of Skill Education, Central Board of Secondary Education



ABOUT THE HANDBOOK

In today's world, we have a surplus of data, and the demand for learning data science has never been greater. The students need to be provided a solid foundation on data science and technology for them to be industry ready.

The objective of this curriculum is to lay the foundation for Data Science, understanding how data is collected, analyzed and, how it can be used in solving problems and making decisions. It will also cover ethical issues with data including data governance and builds foundation for AI based applications of data science.

Therefore, CBSE is introducing 'Data Science' as a skill module of 12 hours duration in class VIII and as a skill subject in classes IX-XII.

CBSE acknowledges the initiative by Microsoft India in developing this data science handbook for class VIII teachers. This handbook introduces the concepts of data science, data visualizations and applications of data science in AI. The course covers the theoretical concepts of data science followed by practical examples to develop critical thinking capabilities among students.

The purpose of the book is to enable the future workforce to acquire data science skills early in their educational phase and build a solid foundation to be industry ready.



Contents

Introduction to Data

1. What is Data? 1
2. Real-World Examples of Data..... 2

Introduction to Data Science..... 7

1. A brief introduction to Data Science 7
2. Careers in Data science..... 8
3. What does Data Science help us achieve? 9

Data Visualization 14

1. Introduction..... 14
2. What is data visualization?..... 15
3. Examples of data visualization 15
4. Importance of data visualization 16
5. Asking the right question 17

Data Science and AI..... 22

1. Introduction..... 22
2. Applications of data science 23
3. Analytics on text data 23
4. Analytics on image data 24
5. Overview of AI 24

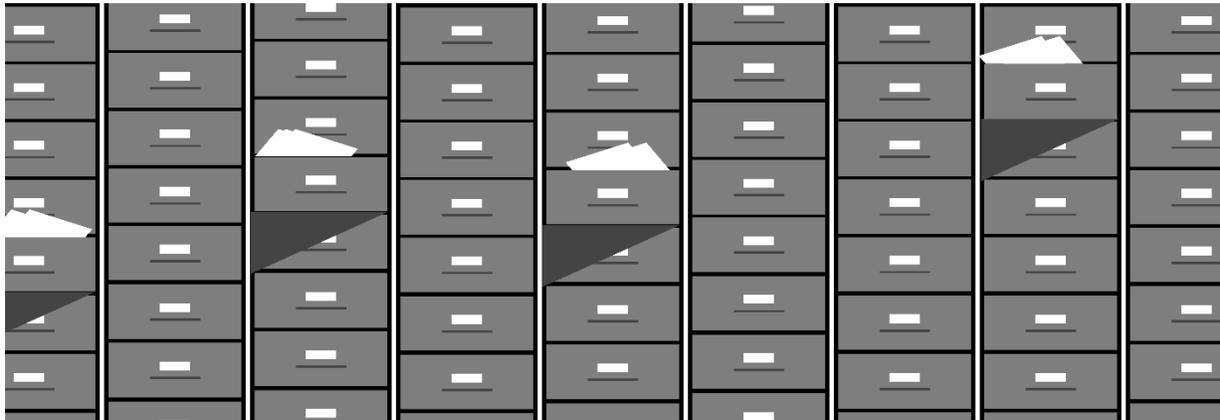
References..... 28



CHAPTER

1

Introduction to Data



Studying this chapter should enable you to understand:

- What is data?
- Real world examples of data

They can also ask the students to come up with such examples to encourage critical thinking.

The teacher should then ask the students to take part in the activities to test their understanding.

Lastly, teachers should elaborate on how data is used in the entertainment industry and ask students to discuss the airline industry.

Teacher's Notes

This chapter is essential for introducing a student to concepts of data and its types. The teacher should gently introduce the concept of data to the students.

Teachers can give real-world examples to the along with the examples given in the textbook.

1. What is Data?

We often use the term data to refer to computer information. This information is either transmitted or stored. Data comes in numerous forms. Any kind of information may it be in numbers or text or pictures is termed as data.

Data is gathered and translated for some purpose, usually analysis. However, if data is not put into context, it doesn't help in any way to humans or computers.



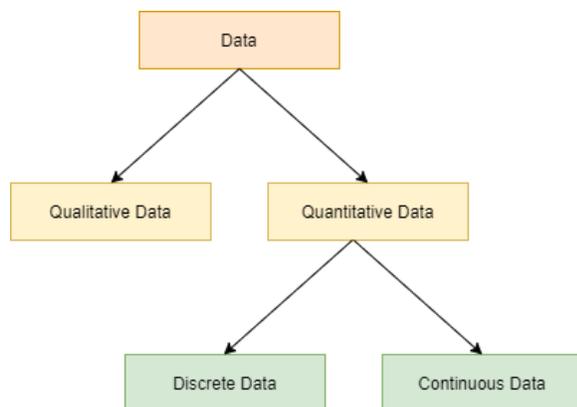
Data comes in different types. Some of the common types of data include:

- Text
- Image
- Video
- Numbers
- Spreadsheets
- Sound

Internally in computers, data is stored as a series of bits that have a value of either one or zero.

Data can be of two types:

1. **Qualitative** - Qualitative data is the data that is a descriptive piece of information. For example, "What a nice day it is"
2. **Quantitative** - Quantitative Data is the data that is numerical information—for example, "1", "3.65," etc.



Quantitative data can be further divided into subtypes "Discrete" and "Continuous."

| Discrete data | Continuous data |
|---|--|
| Discrete data is the data which is expressed as a specific value. This type of data can be counted. | Continuous data is the data that can take on any value in an interval. This type of data can be measured. |
| Examples of discrete data are: <ul style="list-style-type: none">•Number of months in a year.•Number of members in a family. | Examples of continuous data are: <ul style="list-style-type: none">•The amount of oxygen found in the atmosphere.•Age of the members in a family. |

Activity 1.1

Think of real world scenarios of discrete and continuous data.

2. Real-World Examples of Data

Now that we have understood what is data and what are types in which data is



categorized, an obvious question that strikes our mind is that, what is the application of this data in the real world?

Let us have a look at the below scenario from the entertainment industry to



understand the real-world application of data.

Many of you must be used to watching videos on the web. While you start watching a video, many a time you get the list of suggested videos to watch post the existing clip is completed. Have you ever noticed how relevant these videos are to the content that you are watching or the content that you like?

Yes, most of the time, the suggested clips are the ones that, if given a situation where you were not suggested any new videos, you could have searched and played the same clip by your choice.

Isn't it strange how the video platform knows your choices so well? Well, it is all because of data and data analysis. These platforms have many videos in their content. Combined with that, it analyses the videos that people usually play post-watching a video.

These people's preferences are stored and studied. Later an algorithm in the background creates a pattern of people's preferences and shows you the same content in suggested videos, which the majority of people watched post the existing clip.

This is how data analysis is applied in the entertainment industry in real life.

Some of the benefits of data in the entertainment industry are:

- Predicting interests of the audience
- Optimized or on-demand scheduling of media streams in digital media distribution platforms
- Getting insights from customer reviews
- Effective targeting of the advertisements

Recap

- We are surrounded by data. Every computer, every mobile device generates immense amount of data.
- Data comes in different types such as audio, video, text etc.
- Data can be qualitative or quantitative, continuous or discrete.
- Discrete data can take only a specific value.
- Continuous data can have a value within a specific range.



Exercises

Objective Type Questions

Please choose the correct option in the questions below.

1. Discrete data can take any value in a range.
 - a. True
 - b. False

Answer – b.

2. Continuous data cannot take decimal values.
 - a. True
 - b. False

Answer – a.

3. Information stored in a PDF is not considered data.
 - a. True
 - b. False

Answer – b.

4. Quantitative data cannot take numerical values
 - a. True
 - b. False

Answer – b.

5. Qualitative data is descriptive in nature.
 - a. True
 - b. False

Answer – a.

6. “How is the weather like?” is what kind of data
 - a. Quantitative
 - b. Qualitative

Answer – b.

7. Which of the following is considered data?
 - a. Speech
 - b. Video
 - c. Messages
 - d. All of the above

Answer – d.



8. How is data used in the entertainment industry?
- Predicting interests
 - Targeting ads
 - Both of the above

Answer – c.

9. Number of days in a week is an example of?
- Discrete Data
 - Continuous Data

Answer – a.

10. What are the types of quantitative data?
- Discrete
 - Continuous
 - Both a and b

Answer – c.

Standard Questions

Please answer the questions below in no less than 100 words.

1. Explain what data is, with the help of two real-life examples.

Answer – We often use the term data to refer to computer information. Two examples of data can be text data and audio data. See section 1.

2. How is the data categorized?

Answer – Data can be categorized as qualitative or quantitative. See section 1.

3. What is Discrete Data?

Answer - Discrete data is data that can take only specific values. See section 1.

4. What is Continuous Data?

Answer - Continuous data can have a value within a specific range. See section 1.

5. Give two examples of real-life applications of data.

Answer – Recommendations on streaming platforms and online stores.



Higher Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words.

1. How is data used by online streaming platforms?

Answer – See section 2.

2. Give five examples of discrete data around you.

Answer – Teacher should encourage students to give examples of discrete data that they see around them. They can take the reference of the concept of “Discrete Data” explained in the chapter for better understanding.

Applied Project

Data analytics has many applications in our life. Discuss how data analytics is applied in the airline industry to predict flight delays. Few factors which influence flight delays:

- Weather condition (Extreme weather)
- Route restriction/Air traffic
- Mechanical delays
- Availability of runways

Using these factors discuss with your classmates how data analytics can help predict flight delays.

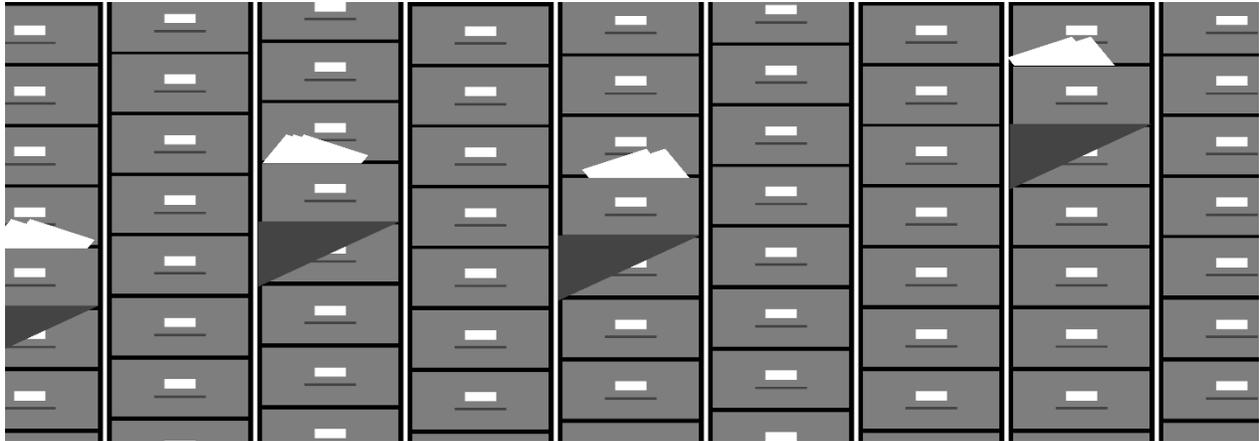
Refer to <https://gallery.azure.ai/Experiment/Binary-Classification-Flight-delay-prediction-3> for detailed implementation.



CHAPTER

2

Introduction to Data Science



Studying this chapter should enable you to understand:

- What is Data Science?
- Careers in Data Science.
- What questions does Data Science answer?

Teacher's Notes

This chapter introduces data science to the students. Teachers should give students a holistic idea of data science, the applications of the field today, and the upcoming new technologies as a result of research.

Teachers can give real-world examples along with the examples given in the textbook.

The teacher should then tell children about the careers in data science and help children decide if they want to pursue this field later.

The teacher should then explain the areas of data science to the students like classification, regression, etc. They can also ask the students to come up with their examples. The teacher should help students with the activities as that will help them understand the concepts better.

1. A brief introduction to Data Science

Every day, through various means of our lifestyle, a tremendous amount of data is generated. When you buy something from your local grocery store, somewhere in this world, someone keeps track of what you have purchased and in which quantity you have purchased.

Let's say you are withdrawing money from ATM for monthly expenditure. You might withdraw various amounts of cash through multiple ATMs at different



locations. ATMs of these Banks will generate data to manage your bank account correctly.

On your social media page, you are clicking the 'Like' button for the singer you like; on your browser, you are surfing through various tutorial videos in a video uploading site – all these activities are creating data.

This data can be investigated and can be organized, and through careful analysis, it will give a clearer picture of what you do and what might be offered to you to enrich your daily life.

This is what data science is about, to extract meaningful interpretations from the data.

The insight gained through the processing of the data using data science is meant for helping our

Activity 2.1

Try to find everyday used applications that depend on data science.

decision-making capabilities. It has various applications; for example, helps various industries to cater to us better and helping authorities to nab criminals, and learn cricket in a better way.

2. Careers in Data science

As we understand about Data, Data Analysis, and Data Science, one of the important questions that coins up is, what are the career options that we can take up in Data Science?

We have learned about the real-life applications of data and data science. Many of us may have found it interesting and may want to pursue this in our career to explore it further.

To help you nail through the right choice, let us understand which different careers we can take up in Data Science. Some common job titles for data scientists include:

1. Data Scientist
2. Business Intelligence Analyst
3. Data Mining Engineer
4. Data Architect
5. Senior Data Scientist

Let us now briefly go through these job titles to get a better understanding:

1. **Data Scientist** - Data Scientists are data enthusiasts who gather and analyze large sets of structured and unstructured data. A data scientist's role combines computer science, statistics, and mathematics. They analyze, process, and model data and later interpret the results to create actionable plans for companies and organizations.

Data Scientists are analytical experts who utilize their skills both in technology and social science to find trends and manage data. They use their industry knowledge and context specific understanding to find solutions to business challenges.

2. **Business Intelligence Analyst** - Business Intelligence Analysts use data to assess the market and find the latest business trends in the industry. This helps to develop a



clearer picture of how a company should shape its strategy.

3. **Data Engineer** - Data Engineer examines not only the Data for their own business but also that of third parties. In addition to mining data, a data engineer creates robust algorithms to help to analyze the data further.
4. **Data Architect** - Data Architects work closely with users, system designers, and developers to create a blueprint that data management systems use to centralize, integrate and maintain the data sources.
5. **Senior Data Scientist** - Senior Data Scientists anticipate the business's needs in the future. Although they might not be involved in gathering data, they play a high-level role in analyzing it. Using their vast experience, they can design and create new standards for analyzing

Activity 2.2

Which career path would be good for you? Discuss

data. They can also create ways to use statistical data and develop tools to further analyze the data.

3. What does Data Science help us achieve?

In simple words, data science helps us answer different types of questions that help us achieve various objectives.

Broadly, these questions can be divided into five types.

Which class does this belong to - A or B?

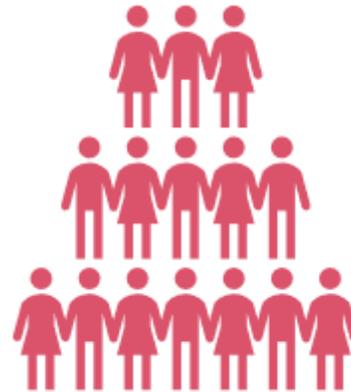
The answers to some questions can only be from a definite number of options.

For example, will it rain today?

A: Yes/No

Q: Will the weather be hot or cold?

A: Hot/Cold



To make such predictions, we use a family of algorithms called classification algorithms. In case we have only two choices, the mechanism is called binary classification. If we try to predict between more than two choices, we use a multiclass classification algorithm.

Is this an outlier?



In some cases, the objective is to find outliers or anomalies in data that is otherwise mostly consistent.

These so-called anomalies could be the cause of concern especially in cases where we need the data to be within a specific range all the time.

An unexpected change in data patterns can often be a sign of something going wrong or possible fraud.

For example, if an unexpected transaction is done from your debit card which does not match your regular transactions, there could be a case of fraud. Banking institutions track these records and alert the customer that an unexpected transaction has happened, and this helps in protecting the customer's money.

Some other examples of anomaly detections are:

Q: Is this email normal or spam?

Q: You are checking your car tyre pressure. Is the reading normal?

The algorithms that are used for these types of questions are called anomaly detection algorithms.

What will probably be the value of this variable?

Machine learning can also help us predict numerical values of continuous variables. There are scenarios in which we must predict numerical values of a variable based on historic data.

Some examples are:

Q: How much rainfall will we receive this year?

A: 100 mm

Q: How many runs will the winning team score?

A: 320

The kind of algorithms that can predict these values are called regression algorithms.

How is the data grouped?

Sometimes data may be separated into distinct groups based on some parameters. This approach is called clustering and is a type of unsupervised machine learning.

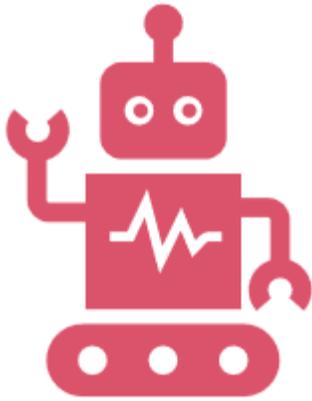
For example, consider the data of the heights and weights of three species of cats.

When we perform clustering, we will get an output that shows us the three different species of cats in three groups.

What should be done now?

This question usually solves the problems of autonomous robots or self-driving cars that need to make decisions based on changes in external factors. Machine learning helps to solve such problems with the help of reinforcement learning.

These models are trained by a process of reward every time a correct action is taken and punishment every time a wrong action is taken.



These systems are automated and can take decisions without human intervention.

Q: I am a robot vacuum. Should I continue cleaning or go to the charging station?

A: Continue

Recap

- Data science is about how to extract meaningful interpretation from the data.
- There are many careers in Data Science like Data Scientist, Data Engineer and Data analyst.
- Data Architect and Senior Data Scientist are two roles for experienced professionals.
- Classification helps us to predict if a new item belongs to class A or class B.
- Regression helps us to predict the value of a continuous variable.
- Clustering helps us to find patterns in the data.
- Reinforcement learning helps models to take decisions based on external factors.

Exercises

Objective Type Questions

Please choose the correct option in the questions below.

1. A school named ABC has recorded the total marks of every student in the class. This an example of:
 - a. Qualitative data
 - b. Quantitative data
 - c. Both qualitative and quantitative data
 - d. None of the above



Answer – b.

2. A food delivery app has asked for your feedback on the quality of the food. You have written two paragraphs to describe the food. This is an example of:
- Qualitative Data
 - Quantitative Data
 - Both qualitative and quantitative data
 - None of the above

Answer – a.

3. It would help if you predicted what the temperature would be for next Friday. Which algorithm will you use?
- Clustering
 - Regression
 - Anomaly detection
 - Binary classification

Answer – b.

4. You need to predict if your car tire will last for the next 1000 km. Which algorithm will you use?
- Clustering
 - Regression
 - Anomaly detection
 - Binary classification

Answer – b.

5. You want to build a way to segregate spam emails from good emails. Which algorithm will you use?
- Clustering
 - Regression
 - Anomaly detection
 - Binary classification

Answer – c.

Standard Questions

Please answer the questions below in no less than 100 words.

1. What are the common career paths for data science?

Answer – See section 2.

2. What does a Data Architect do?



Answer – See section 2.

3. What are the differences between classification and regression?

Answer – Classification is for categorical data; regression is for continuous data.

Higher Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words.

1. Discuss a recent innovation that makes use of reinforcement learning.

Answer – Many real-world innovations make use of reinforcement learning. The most important one is self-driving cars.

2. Write a short note on how data science is helping sports teams.

Answer – Player statistics and team statistics can be utilized to understand and improve the performance of sports teams. You can read more here - <https://analyticsindiamag.com/a-primer-to-using-data-science-in-sports/>

Applied Project

Emails are a part of daily communication. Sometimes we receive unwanted emails called spam. There are few techniques that email providers use to identify spam mails :

- Content-based filtering (Analyzing the words, occurrence, distribution of words to identify spam mail)
- Header filters (Reviewing the email header) Example: Promo!, Offer!

Provide 2 examples of each of the words/phrases in email content & header which marks an email as spam. Explain in detail, how email providers make use of clustering to mark an email as spam. Also, elaborate on how email providers create and update the words/phrases to mark an email as spam.

Answer – Email companies make use of clustering techniques, wherein they look at different sections of the email (header, sender, and content). The data is then grouped. These groups can then be classified to identify which are spam.



CHAPTER

3

Data Visualization



Studying this chapter should enable you to understand:

- What is data visualization?
- The importance of visualization
- Collecting relevant data
- Asking the right question
- Predict an answer
- Examples of data visualization

Teacher's Notes

This chapter introduces the important concept of data visualization. Teachers first explain the importance of data visualization and how getting good data ultimately helps to get better insights.

Teachers should then guide students on how they should go about conducting a

data analysis – what questions should they ask, and which factors they should consider before making a decision.

Teachers should stress on which visualization to choose for which kind of data. There are cases where a pie chart works well but cases where it would not make much sense. Teachers should focus on the point that it is their visualizations that ultimately tell their story and insights.

1. Introduction

In the previous chapters, we learned about how data is collected and how we can interpret the data by asking several types of questions on the data. In this chapter, we will learn to visualize data and make predictions.



2. What is data visualization?

Data visualization is the representation of data or information in a graph, chart, or other visual formats.

Data visualization provides a way to see and understand trends, outliers, and patterns in data. Charts and graphs make communicating data findings easier even if you can identify the patterns without them.

The goal of data visualization is to communicate information clearly and efficiently to users.

Common types of data visualizations are:

- Charts
- Graphs
- Tables
- Maps
- Histograms

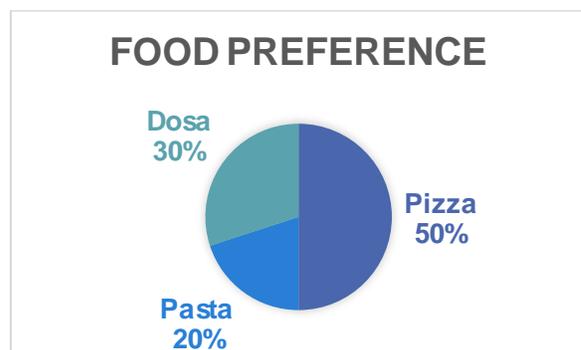
3. Examples of data visualization

Example 1: Using a pie chart displays the data of the food preferred by the students.

We have the food item preference of 50 students. Let us now visualize the data using a pie chart and find the most preferred and the least preferred food item.

| Food item | Number of students |
|-----------|--------------------|
| Pizza | 25 |
| Pasta | 10 |
| Dosa | 15 |

Let us now visualize the data using a pie chart:



The most preferred food item is pizza and the least preferred food item is pasta.

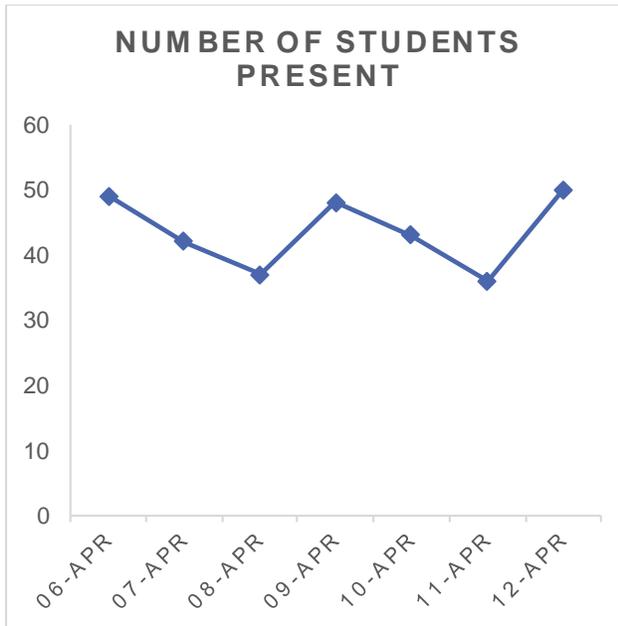
Example 2: Using a line chart displays the data of the number of students present in the class for one week.

Here is the data:

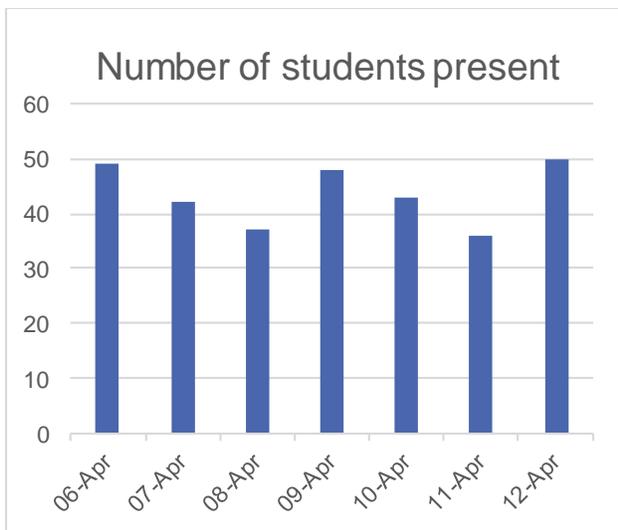
| Date | Number of students present |
|--------|----------------------------|
| 06-Apr | 49 |
| 07-Apr | 42 |
| 08-Apr | 37 |
| 09-Apr | 48 |
| 10-Apr | 43 |
| 11-Apr | 36 |
| 12-Apr | 50 |



Let's now visualize the data using a line chart



We can also visualize the same data using a bar graph:



4. Importance of data visualization

To make sure that we get the required outcome from the data, we must collect the right and relevant data.

It is essential to have correct and good quality data to make an analysis or to construct algorithms that can have an impact. Without relevant data, your analyses will not only be irrelevant, but they can also be misleading.

You cannot expect to find perfectly preprocessed raw data that be used directly for your needs. Hence, you need to understand how the data was gathered and what sources it was collected from.

Therefore, it is essential to understand how to collect relevant data for analysis.

Let us understand what steps we need to take to make sure that we collect the right set of data for analysis.

- Quality of the Data - Primary and most vital point to consider while collecting the data is the quality of data that is getting collected. If we collect incomplete data, build an unreliable database, and run analysis on skewed data sets, obviously we are not going to arrive at the required output. The quality of data that is collected should always be the top priority while assessing the data.
- Completeness of data - We need to make sure that the data that is getting collected is a complete set. Incomplete sets of data may cause



many discrepancies and wrong output on analysis.

- Format of data - The format of the Data that is collected for analysis should be right. Data should be accessible and readable for analysis. If the collected data is not in the right format, we should convert it to the required format for analysis.

5. Asking the right question

Once we have the required data ready with us, the next step is to ask the right question to the data. It is important to understand that if we don't ask the right questions, we will never get the right answers. To make sure we perform the proper analysis of data, we should ask the data the right set of questions.



Below are specific questions that you need to ask to your data set to get the right answer:

- What do you wish to find?

It is essential to consider what your goal is and what decision-making it will facilitate. What outcome from the analysis would you consider a success?

These initial analysis questions are important to guide you through the process and help focus on valuable insights. You can start by brainstorming and preparing a draft guideline for specific questions you want to find from the data. This will help you to dive deeper into the more specific insights you want to achieve.

- Which statistical techniques are applicable?

There are several statistical analysis techniques that you can use for analyzing data. However, in real-life scenarios, three statistical techniques are mostly used for analysis:

- a. Regression Analysis is a process for finding out the relationships and correlations among the different variables in the data.

Regression analysis helps us to figure out how the value of the dependent variable varies when one of the independent variables' changes. Thus, regression analysis helps to identify which independent variables affect the dependent variable.

- b. Cohort Analysis – it enables you to easily compare how different



groups, or cohorts, of customers, behave over time.

For example, you can create a cohort of customers based on the date when they made their first purchase. Subsequently, you can study the spending trends of cohorts from different periods in time to determine whether the quality of the average acquired customer is increasing or decreasing over time.

c. Predictive Analysis – Predictive analytics is the process of historical datasets to predict future possibilities. It can also be used for generating alternative scenarios and risk assessments.

- Who will be using the final results?

An important aspect of your data analytics refers to the end-users of our analysis. Who are they and how will they be using the reports you create? You must get to know your final users, including:

- a. What do they expect to learn from the data?
- b. What do they need?
- c. How advanced are their technical skills?
- d. How much time do they have?

If you know these answers, you can decide on how detailed your data visualizations should be and what areas of the data your report should be focused on.

You should keep in mind that technical and non-technical users have different needs. If the reports are designed for executives and non-technical staff, you know which insights will be useful for them and what level of data complexity they can handle.

If external parties will be using your reports, the reports you make should provide them with an easy-to-use and actionable interface. The final users should be able to read and interpret them independently, with no support needed.

- Which visualizations should you pick?

Once your data is clean and your statistical analysis is done, you need to pick your visualizations. You can have impactful and valuable insights from the data, but if they're presented badly, the end-users won't be able to understand the insights from them.

It is essential to convince executive and decision-makers that the data that you have gathered and analyzed are:

- a. Correct
- b. Important
- c. Urgent to act upon

Effective presentation aids in all these areas. There are several kinds of charts to pick from. You can improve your chances of getting good feedback by choosing the right data visualization type.



There are several data visualization software like Power BI that can perform most of the tedious aspects of data cleaning as well. They can not only help to prepare the data but also interpret the insights. Because they are so easy to use and test data hypotheses without the need for intensive training, these tools have become an invaluable resource in today's data management practice.

analyzing data, the tools can help to perform a voluminous analysis.

These tools are very flexible for the end-user and can easily adjust to your prepared questions for

Recap

- It is important to check the quality of data, completeness of data and format of data.
- Once we have the required data ready with us, next step is to ask right question to the data.
- There are a number of statistical analysis techniques that you can use for analyzing data – Regression, Predictive analytics and cohort analysis.
- Once your Data is clean and your statistical analysis is done, you need to pick your visualizations.
- Picking the right visualisation is important otherwise the results will not be interpreted well.
- You must analyse the end users of the visualisations in order to decide how to present your insights.



Exercises

Objective Type Questions

Please choose the correct option in the questions below.

1. Data can be visualized using:
 - a. Graphs
 - b. Maps
 - c. Charts
 - d. All of the above

Answer – d.

2. Which of the following statement is false?
 - a. Data visualization can absorb information quickly.
 - b. Data visualization decreases the insights and takes slower decisions.
 - c. Data visualization is a type of visual art.
 - d. None of the above

Answer -b.

3. Which of the following is a use case of data visualization?
 - a. Healthcare
 - b. Sales and Marketing
 - c. Politics/Campaigning
 - d. All of the above

Answer – d.

4. Which format of data is easiest for analysis?
 - a. Tabular data
 - b. Text data in a PDF
 - c. Data in an image
 - d. Speech data

Answer - a

5. Which visualization is best for representing a relation between two variables?
 - a. Scatter plot
 - b. Histogram
 - c. Pie chart
 - d. Gantt Chart

Answer – a



Standard Questions

Please answer the questions below in no less than 100 words.

1. What are the steps to make sure that the correct data is collected for analysis?

Answer – See section 1.

2. Write a short note on the statistical techniques which can be used for data analysis.

Answer – Regression Analysis, Cohort Analysis, etc. See section 3.

3. Is it important to assess the end-users for a visualization? Explain in your own words.

Answer - An important aspect of your data analytics refers to the end-users of our analysis. Who are they and how will they be using the reports you create? See section 3.

Higher Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words.

1. What are the things to consider before deciding on an appropriate visualization for your data?

Answer - See section 3.

2. If you find that the data collected has outliers, what steps can you take to ensure that your analysis is still accurate?

Answer – First you should determine whether the outliers are correct or incorrectly recorded data. We can then follow these methods to reduce the impact of outliers - <https://www.kdnuggets.com/2017/01/3-methods-deal-outliers.html>

Applied Project

Each student should write down the marks he/she had received in the examination for the subjects studied in the previous grade. Use these marks to plot on paper

- a. A bar graph to display marks of each subject.
- b. A line graph to display marks of each subject.
- c. A pie chart to show the percentage contribution of marks of each subject to the total marks obtained.



CHAPTER

4

Data Science and AI



Studying this chapter should enable you to understand:

- Applications of data science
- Analytics on text data
- Analytics on image data
- Overview of AI

6. Teachers can also show examples of some technologies as a demo – for example, teachers can show a mobile phone which has face recognition built into it. This will help arouse the children’s curiosity and help them understand better.
7. Teachers should give students a holistic idea of artificial intelligence and explain the various goals of AI.

Teacher’s Notes

This chapter introduces groundbreaking applications of data science in various fields. Some students might not have encountered technologies like speech recognition or facial recognition and so the teachers should take time to explain these terms.

1. Introduction

In the last chapter, we saw how we can visualize and make predictions from the data. In this chapter, we will learn about the applications of data science and the basics of artificial intelligence.



2. Applications of data science

Looking at the advantages of data science, it is obvious that many companies have applied data science to help their business grow. If we look around us, data science is everywhere. In numerous ways, it is affecting our day-to-day life.

Digital Advertisements - You must have noticed many times that if you do an internet search for a thing, say, you searched for a handbag, and close the browser once your surf is complete. Later, when you open any other applications or websites, you see advertisements for handbags from various brands in your window. Ever wondered, how this new application or website knows that you are looking to buy a handbag? Well, the answer to this is data science. Algorithms in data science help in tracking your searches and learn your preferences from them.

Speech Recognition - Speech recognition is now part of our everyday lives. Speech recognition has now become a part of phones, game consoles, and even smartwatches. Have you heard of Microsoft's Cortana? It uses speech recognition behind the scenes to take inputs from the user.

Speech recognition can also be found on many devices that can be used to automate our homes.

Speech recognition has been around for more than a decade. However, it is

gaining popularity now as machine learning is helping organizations make speech recognition much more accurate.

3. Analytics on text data

Text analytics can be defined as the process of collecting unstructured text from various sources and analyzing and extracting relevant information from it. It can also be used for transforming it into structured information that can then be used in various other ways.

There are several ways to analyze unstructured text. Most of these techniques can be divided under these technical areas - Natural Language Processing (NLP), data mining, and information retrieval.

Typically, we used text analytics technologies for four basic tasks - querying data, mining data, search data, and analyzing data to get insights.

For example, if we have a database with customer data, an end-user could query the database to find out how many customers have started using the company's services in the last quarter and how many have stopped using the service. They can do so by just entering a query in plain English instead of a query language like SQL.

Chatbots are also an important area that uses text analytics for both querying and searching data. Chatbots can use to query a database and give a reply based on the question. They can also use search based on text analytics to help in



retrieving a document based on what end users are looking for.

4. Analytics on image data

Image recognition can be described as a process by which we can process images for identifying people, patterns, logos, objects, or places.

Many machine learning tools can assist users with facial recognition of objects in a picture. These tools can also scan the objects in the picture and attempt to identify and name them based on a large database of images.

Mobile phones, for example, make use of computer vision technologies in combination with a camera to achieve image recognition. This advanced technology has a variety of applications like accessibility for the visually impaired and interactive advertising.

Facial recognition is also used by many organizations to check the attendance of workers and by government organizations for identification purposes.

Besides identifying faces and detecting objects in images, AI is also capable of recognizing special patterns, in the images and matched them with its database.

Another growing application of image analytics is in searching content based on images. Some search engines now allow users to upload images and search based on that.

5. Overview of AI

Artificial Intelligence is defined as the science and engineering of making intelligent machines. AI is a branch of Computer Science which deals with the research and design of intelligent systems that can take inputs from their environment and takes actions based on it as a human being would.

In technology, Artificial Intelligence, Machine Learning, and Deep Learning are widely used. While you may have seen these terms get used interchangeably, each carries its significance and application.

Artificial Intelligence, Machine Learning are subsets of each other, while Machine Learning is the superset of Deep Learning.



Artificial Intelligence aims at making machines as a smart human. This main goal of Artificial Intelligence can be explained using the below sub-goals:

- a. **Logical Reasoning:** AI aims at making computers capable of doing



all the intelligent and sophisticated tasks that we humans can do. For example, solving problems that require logical reasoning like switching on the fan because it is hot or solving complex mathematical problems.

basic vision for this goal is to enable machines to exhibit emotional intelligence, moral reasoning, and more.

- b. **Knowledge Representation:** Make computers capable of describing objects. For example, describing a car that just violated the traffic norms.
- c. **Planning and Navigation:** Making computers capable of traveling from Point X to Point Y. For example, a self-driving robot.
- d. **Natural Language Processing:** Make computers capable of understanding and processing a language. For example, a web translator that translates one language to another.
- e. **Perception:** Make computers capable of interacting with real-world objects by the sense of touch, sound, smell, and eyesight.
- f. **Emergent Intelligence:** Make computers capable of Intelligence that is not explicitly programmed but is derived from AI capabilities. The

Recap

- There are two important applications of data science – digital ads and speech recognition.
- Text analytics can be defined as the process of collecting unstructured text from various sources and analyzing and extracting relevant information from it.
- Chatbots are also an important area that uses text analytics for both querying and searching data.
- Image recognition can be said to be a process by which we can process images for identifying people, patterns, logos, objects or places.
- Artificial Intelligence is defined as the science and engineering of making intelligent machines.
- AI has many sub goals like – natural language processing, perception etc.

Exercises



Objective Type Questions

Please choose the correct option in the questions below.

1. Data Science can help with:
 - a. Speech Recognition
 - b. Digital Advertising
 - c. All of the above

Answer – c.

2. Which of the following is a goal of Artificial Intelligence?
 - a. Logical Reasoning
 - b. Knowledge Representation
 - c. Planning and Navigation
 - d. All of the above

Answer – d.

3. Which of the following is a use case of data science?
 - a. Facial recognition
 - b. Text analytics
 - c. Sentiment analysis
 - d. All of the above

Answer – d.

4. What does natural language processing help us with?
 - a. Text analytics
 - b. Video analytics
 - c. Image analytics

Answer – a.

5. What technologies are used by chatbots?
 - a. Text analytics
 - b. Speech recognition
 - c. Both of the above

Answer – c.

Standard Questions



Please answer the questions below in no less than 100 words.

1. How is data science used for speech recognition?

Answer - See section 1.

2. Write a use case for analyzing images.

Answer - Cameras used in mobile phones. See section 4.

3. What are some of the goals of AI?

Answer – Perception, logical reasoning, etc. See section 5.

Higher Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words.

1. How is text data analyzed?

Answer - See section 3.

2. What are some of the applications of image recognition?

Answer – Image recognition has applications in several industries like healthcare, retail, and manufacturing. You can find more examples here - <https://analyticsindiamag.com/8-uses-cases-of-image-recognition-that-we-see-in-our-daily-lives/>

Applied Project

Understanding the mood of the speaker can be very useful. Certain keywords can be associated with different sentiments.

Example 1: “The news continues to be gloomy.” If you read this sentence you will understand that the sentiment of the speaker is sad.

Example 2: “I was infuriated by his arrogance.” This sentence tells you that the sentiment of the speaker is angry.

Discuss with your classmates how text analytics can help us identify the sentiment of the speaker i.e. if the speaker is happy, angry, or sad. A sentence may have more than one keywords that highlight the sentiment of the speaker. Provide 2 examples of such scenarios for each of the sentiments discussed above.



References

Bargagliotti, A., Franklin, C., Arnold, P. and Gould, R., 2020. Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II). American Statistical Association

Luis Bermudez. 2021. Overview of Artificial Intelligence Buzz. [Online]. [25 February 2021]. Available from: <https://medium.com/machinevision/overview-of-artificial-intelligence-buzz-adb7a5487ac8>

Vivek Kumar. 2020. WHY DOES DATA SCIENCE MATTER IN ADVANCED IMAGE RECOGNITION? [Online]. [4 March 2021]. Available from: <https://www.analyticsinsight.net/data-science-matter-advanced-image-recognition/>