Great Learning
POWER AHEAD

Welcome! We will begin shortly

Live Virtual Class

Intro to the World of Data Science

Learning Outcomes

- Understand the key terminologies in the World of Data Science
- Understand how these key terminologies are connected

Guidelines

Listen only mode

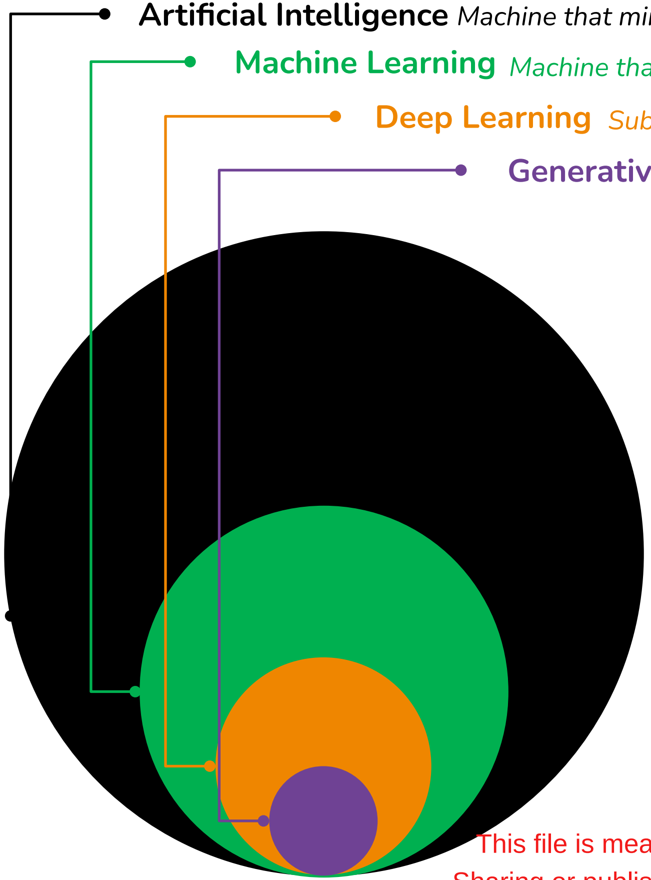
Ask questions at the interest of the larger audience

FAQ Questions in the Q&A Box

Thank you

Kindly utilize the chat box for **subject-relevant questions only** to maximize your learnings from the session.
Your questions are being managed by the academic team, and they will be answered

Key Terminologies in The World of Data



Artificial Intelligence *Machine that mimics human intelligence & behaviour*

Machine Learning *Machine that learns from the data automatically w.r.t a particular task and performance measure*

Deep Learning *Subset of ML, learning happens via algorithms inspired by neural networks in human brain*

Generative AI *Subset of DL, creating new content from existing ones using advanced algorithms*

Key Terminologies in The World of Data

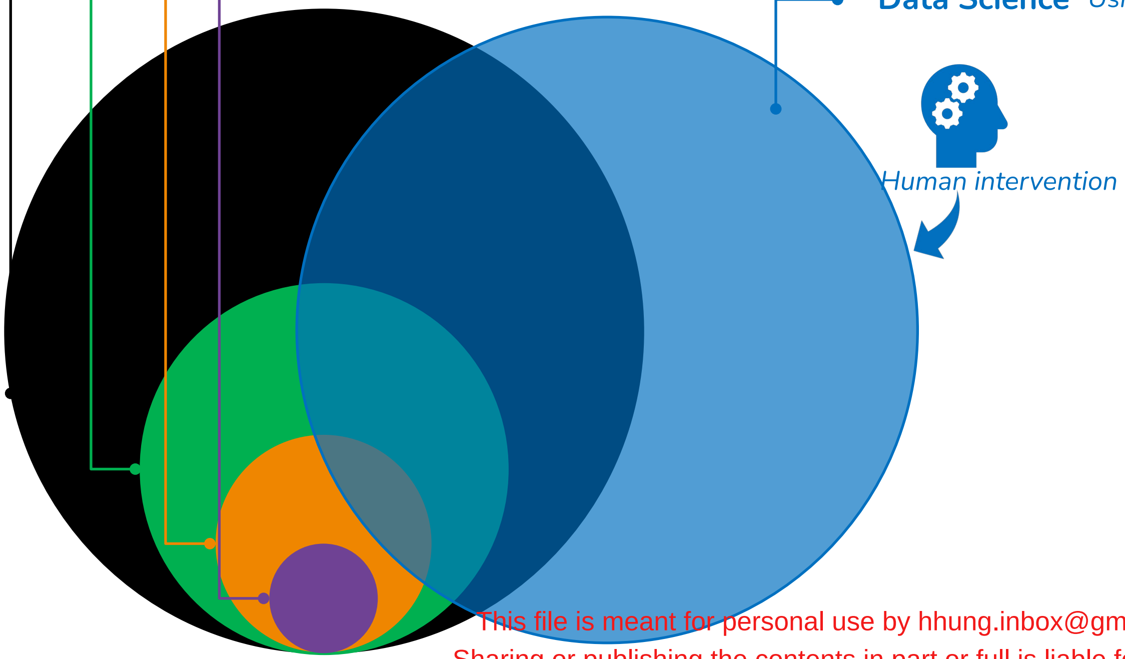
Artificial Intelligence *Machine that mimics human intelligence & behaviour*

Machine Learning *Machine that learns from the data automatically w.r.t a particular task and performance measure*

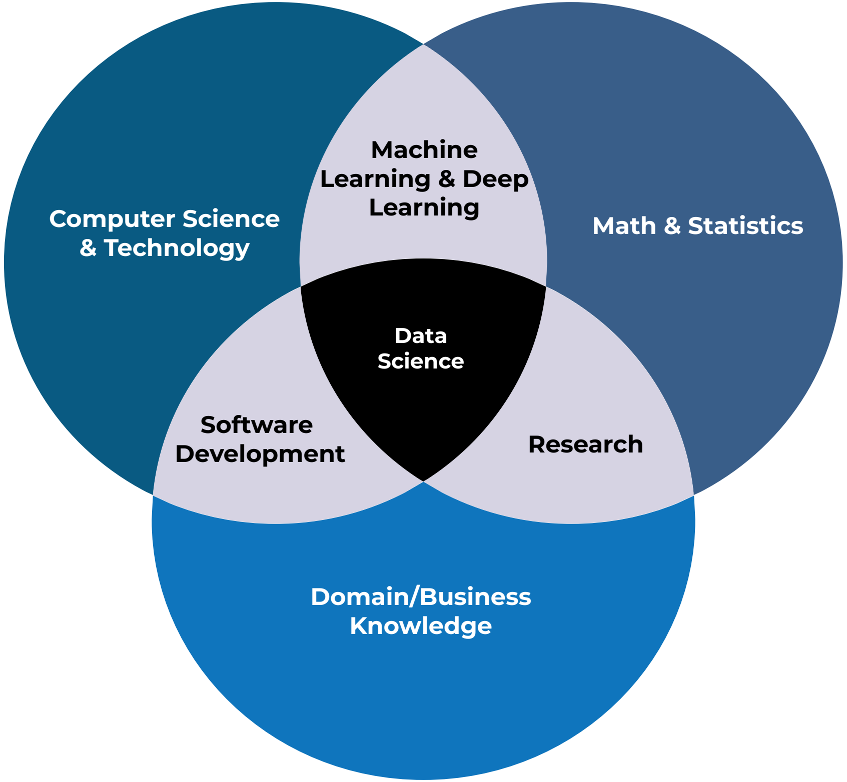
Deep Learning *Subset of ML, learning happens via algorithms inspired by neural networks in human brain*

Generative AI *Subset of DL, creating new content from existing ones using advanced algorithms*

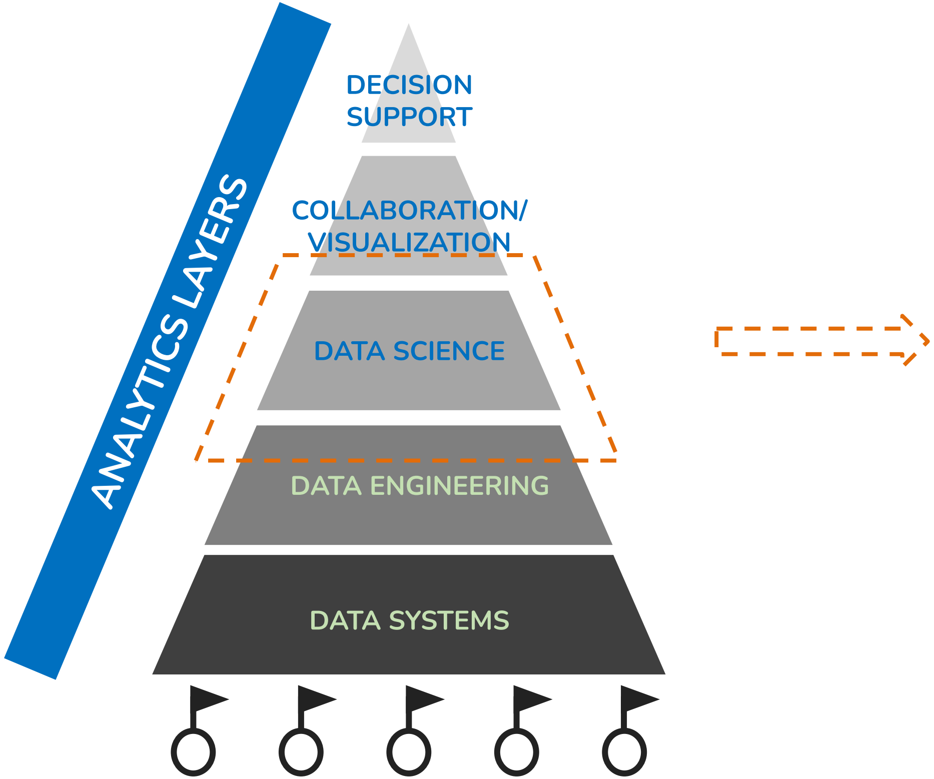
Data Science *Using AI to enable data driven decision making*

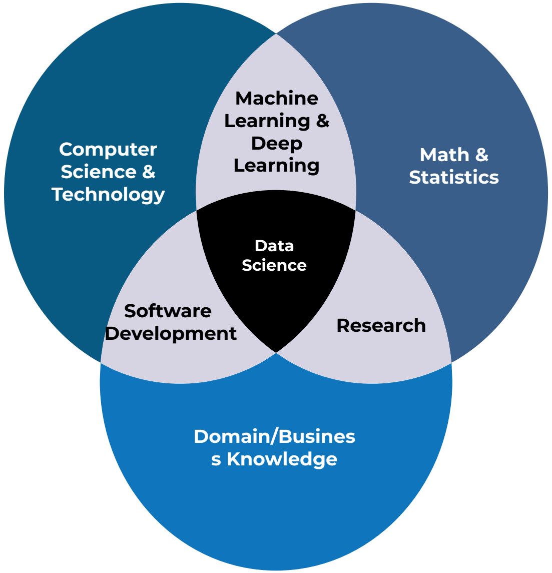


What is Data Science (DS)?

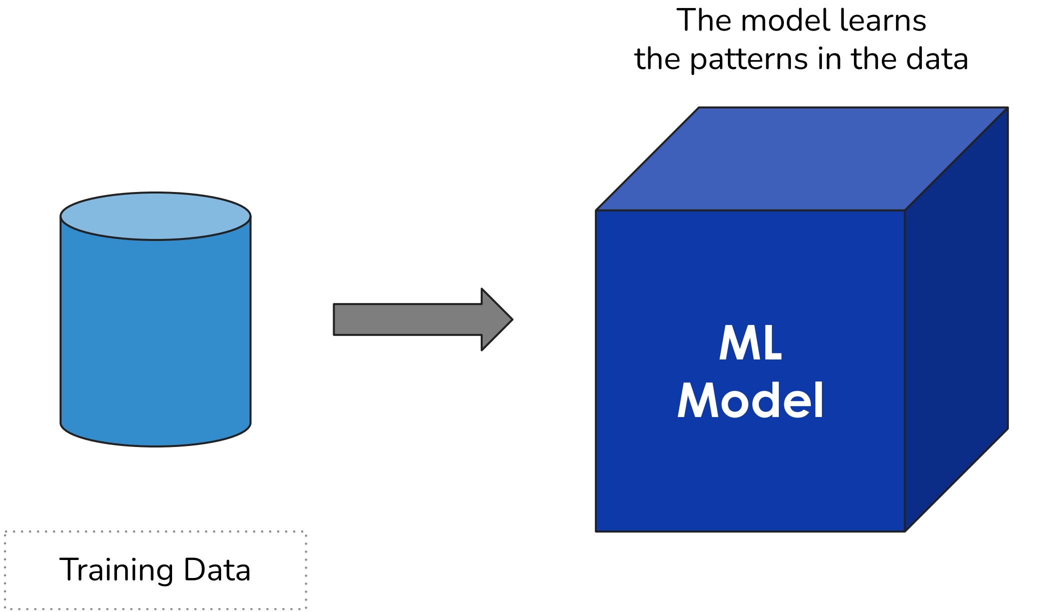

Computer Science & Technology
Machine Learning & Deep Learning
Math & Statistics
Data Science
Software Development
Research
Domain/Business Knowledge

Analytics vs Data Science

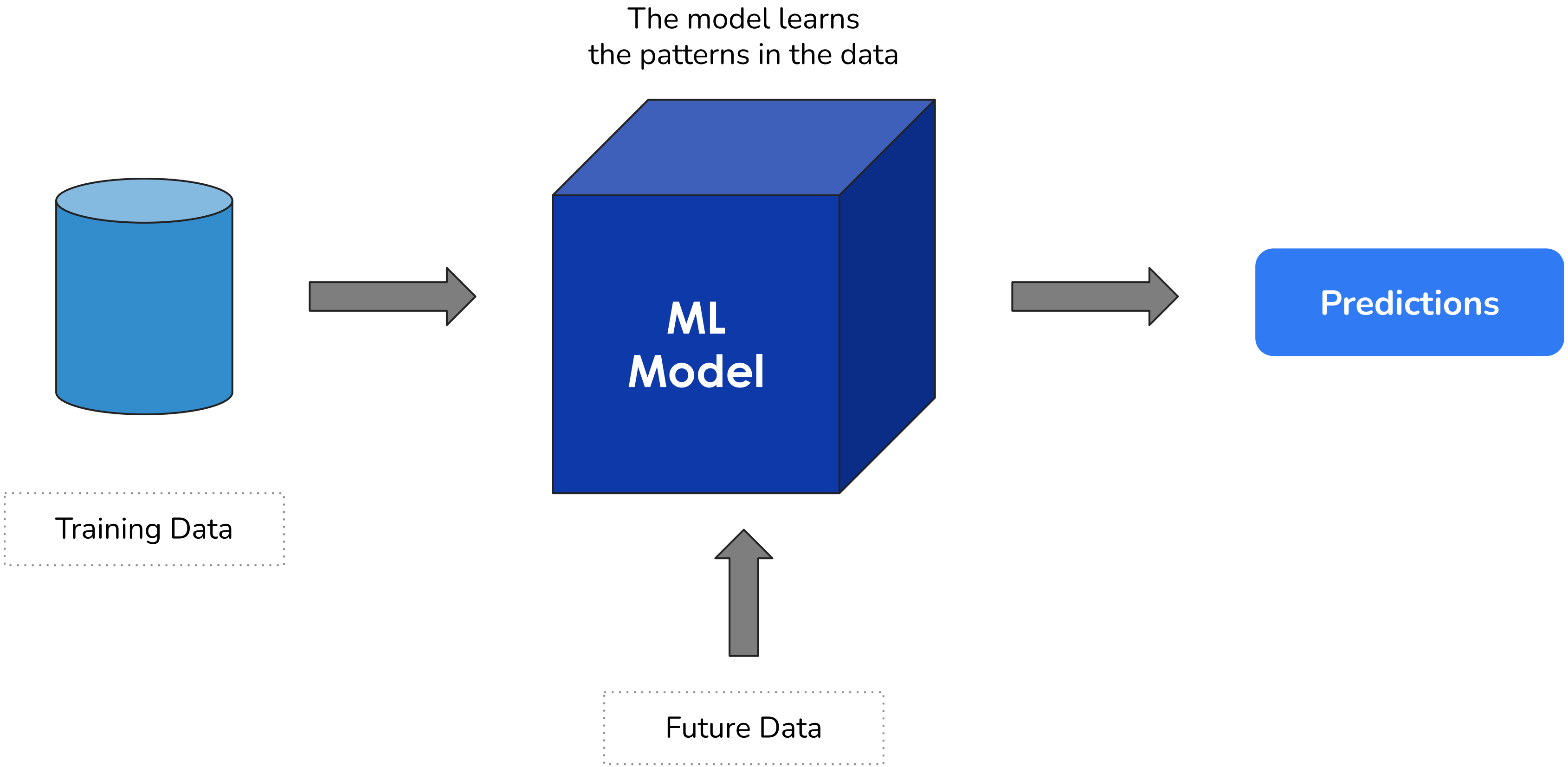



Computer Science & Technology
Machine Learning & Deep Learning
Math & Statistics
Data Science
Software Development
Research
Domain/Business Knowledge

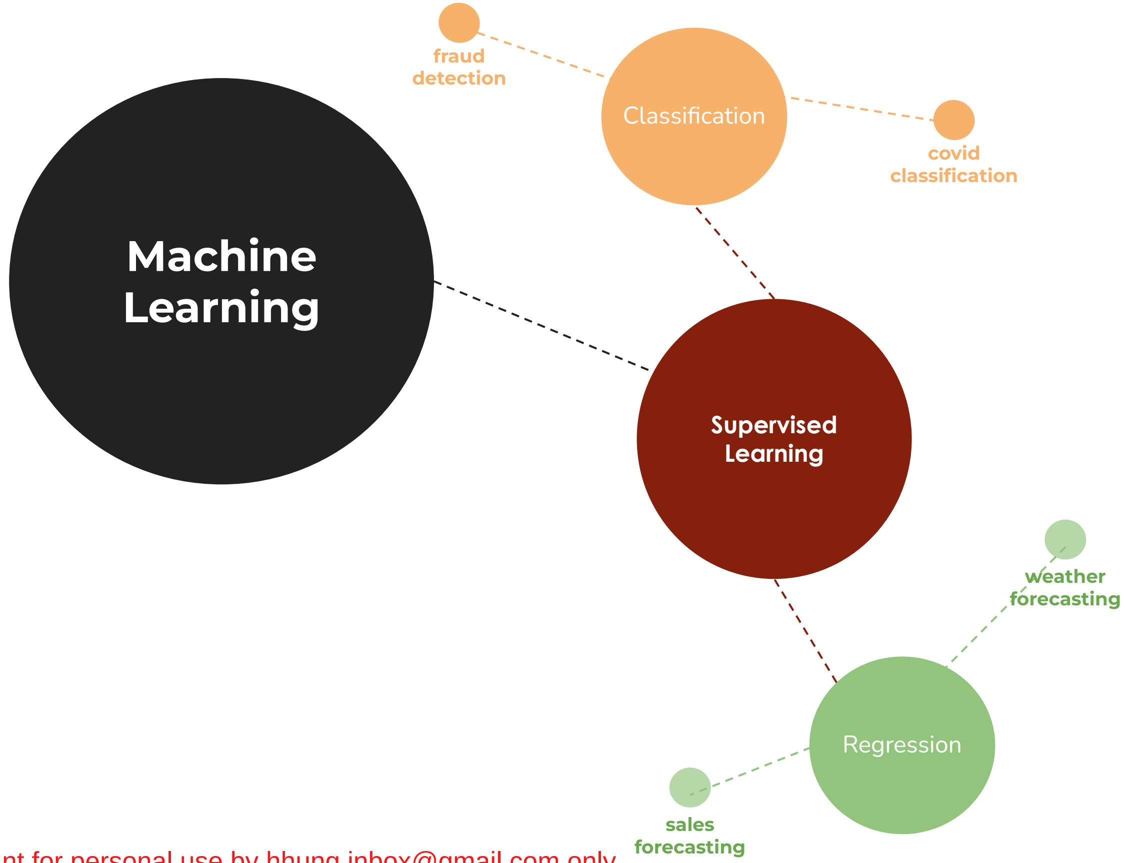
What is Machine Learning?



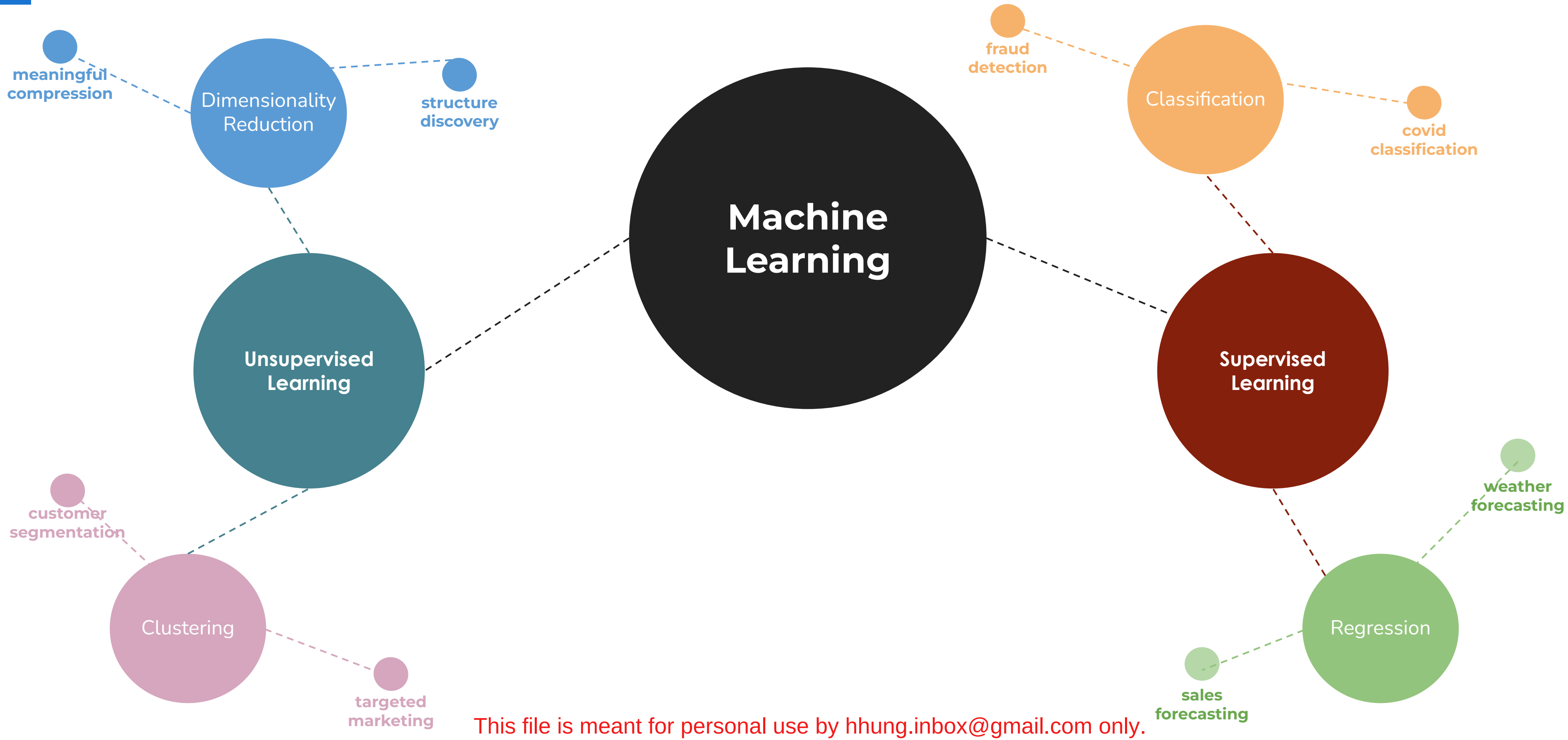
What is Machine Learning?



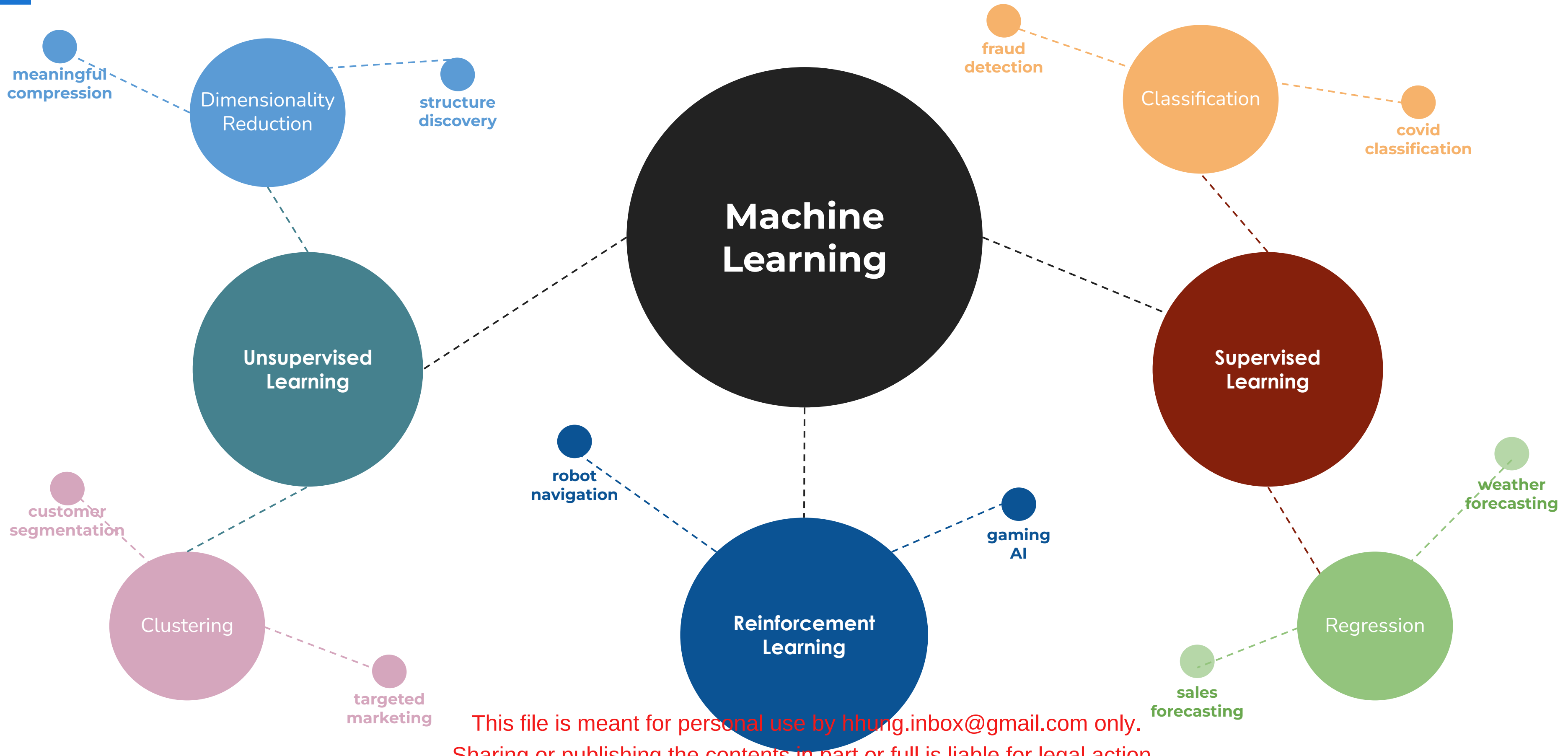
Types of Machine Learning



Types of Machine Learning

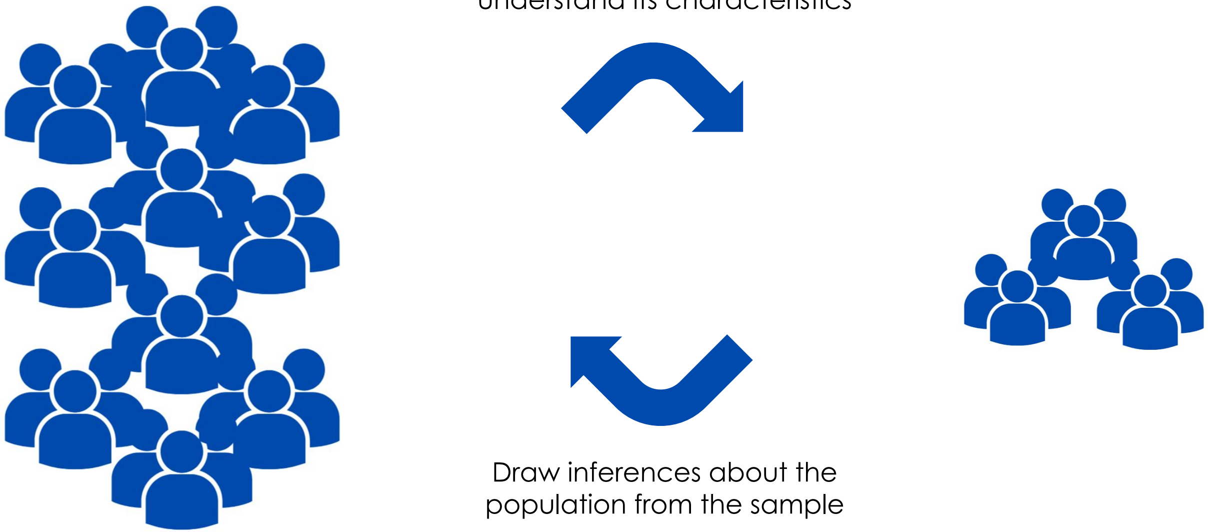


Types of Machine Learning



Types of Statistics

Draw samples from the population to understand its characteristics

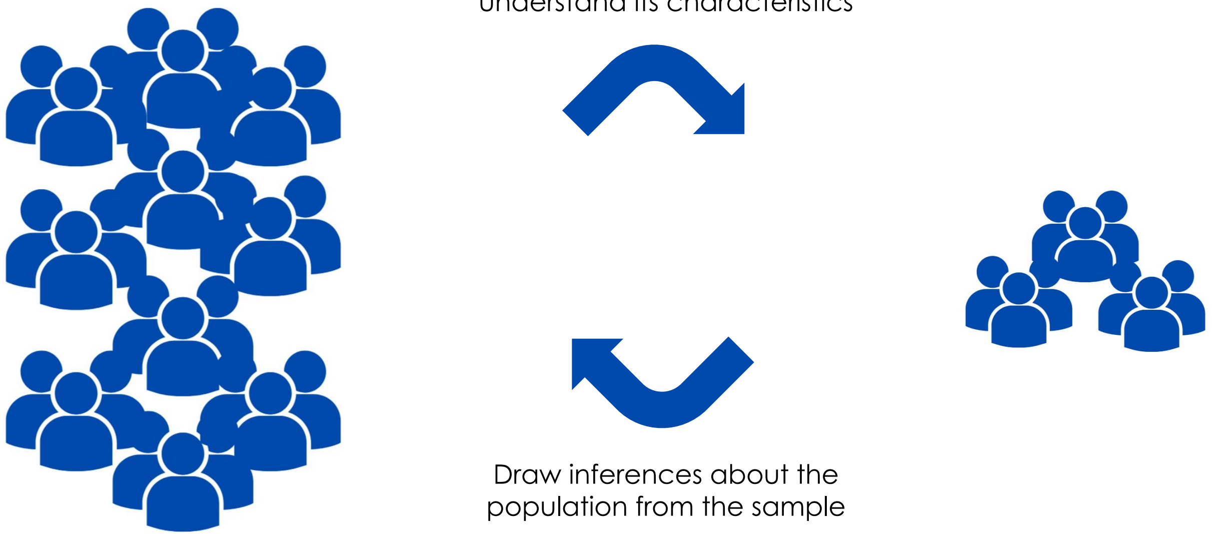


POPULATION

SAMPLE

Draw inferences about the population from the sample

Types of Statistics



POPULATION

Draw samples from the population to understand its characteristics

Draw inferences about the population from the sample

SAMPLE

Inferential Statistics

Confidence Intervals - The size of the transistor on a processor chip lies in the 95% confidence interval (4.95, 5.05) nm

Hypothesis Testing - Does the conversion rate of a marketing campaign vary with the font style of the infographic?

Descriptive Statistics

Central Tendency - Mean, Median, Mode

Dispersion - Variance, Range, Standard Deviation

Data Quiz

Which of the following are examples of the inferential paradigm of Data Science?

A Effectiveness of a new medication through randomized trial

B Weather forecasting based on historical and weather patterns

C Impact of a new policy on citizens

D Optimize routing of vehicles to minimize costs

Data Quiz

Which of the following are examples of the inferential paradigm of Data Science?

A Effectiveness of a new medication through randomized trial

B Weather forecasting based on historical and weather patterns

C Impact of a new policy on citizens

D Optimize routing of vehicles to minimize costs

Data Quiz

Effectiveness of a new medication through randomized trial

New medications are tested in a very controlled manner and amongst a specific, predetermined group to ensure that the we get a clear understanding of the effects and side-effects of the medications before rolling them out for large-scale production

Weather forecasting based on historical and weather patterns

Weather forecasting systems now use data from a variety of sources (like weather stations, satellites, etc), assimilate the data, and then use efficient computationally powerful mechanisms to provide accurate forecasts

Impact of a new policy on citizens

Impact of new policies on citizens is also conducted in an experimental format with careful considerations and comparative analysis to ensure that we arrive at the right decisions that would optimize the preset goals

Optimize routing of vehicles to minimize costs

Vehicle routes are optimized by taking into consideration a large number of factors (like geographical data, traffic data, etc.), identifying the constraints, and running computationally powerful algorithms to establish the most cost-effective route

Data Quiz

In World War II, which group of mathematicians played a crucial role in breaking German encryption codes?

A The Codebreakers

B The Enigma Team

C The Los Alamos Group

D The Navajo Code Talkers

Data Quiz

In World War II, which group of mathematicians played a crucial role in breaking German encryption codes?

A The Codebreakers

B [✓ The Enigma Team]

C The Los Alamos Group

D The Navajo Code Talkers

Data Quiz

The Codebreakers

A diverse group of mathematicians, cryptographers, and intelligence personnel who played a pivotal role in breaking complex codes and ciphers during World War II. They helped decipher encrypted messages and provided invaluable intelligence that contributed to Allied victories.

The Enigma Team

Enigma was a complex electro-mechanical device used by the German military during World War II to encrypt and decrypt secret messages. The machine's encryption was broken by Alan Turing and team, resulting in a significant intelligence advantage for the Allies and contributing to their eventual victory.

The Los Alamos Group

Stanislaw Ulam and his group of mathematicians at Los Alamos National Laboratory made substantial contributions to computational methods, including Monte Carlo simulations, which have become fundamental techniques in statistical analysis and machine learning.

The Navajo Code Talkers

A group of Native American soldiers who played the role of code talkers in World War II, using the Navajo language to create an unbreakable code. They were instrumental in securing military communications and contributing to Allied success

Data Quiz

Which of the following statements are NOT true regarding control charts?

A They are commonly used in Six Sigma projects to monitor and analyze process performance over time

B They consist of a centerline and upper and lower control limits

C They can indicate process variation and also provide direct insights into the root causes of the variation

D They rely on historical data collected from a process and are only as good as the data they rely on

Data Quiz

Which of the following statements are NOT true regarding control charts?

A They are commonly used in Six Sigma projects to monitor and analyze process performance over time

B They consist of a centerline and upper and lower control limits

C They can indicate process variation and also provide direct insights into the root causes of the variation

D They rely on historical data collected from a process and are only as good as the data they rely on

Data Quiz

They are commonly used in Six Sigma projects to monitor and analyze process performance over time

They are commonly used in Six Sigma projects, provide a visual representation of data collected from a process, and help determine whether the process is within statistical control or experiencing significant variation.

They consist of a centerline and upper and lower control limits

Control charts consist of a centerline and upper and lower control limits. The centerline usually represents the process average.

They can indicate process variation and also provide direct insights into the root causes of the variation

While control charts can indicate process variation, they do not provide direct insights into the root causes of the variation. Additional analysis and investigation are required to identify and address the underlying causes.

They rely on historical data collected from a process and are only as good as the data they rely on

Control charts are based on historical data collected from a process. If the historical data does not accurately represent the current process conditions or if the process has undergone significant changes, they may not be accurate.

Data Quiz

In 2011, IBM's AI system competed on the quiz show Jeopardy! and defeated human champions. What was the name of this AI system?

A Deep Blue

B Watson

C AlphaGo

D HAL 9000

Data Quiz

In 2011, IBM's AI system competed on the quiz show Jeopardy! and defeated human champions. What was the name of this AI system?

A Deep Blue

B Watson

C AlphaGo

D HAL 9000

Data Quiz

Deep Blue

A supercomputer developed by IBM that gained worldwide recognition for defeating reigning world chess champion, Garry Kasparov, in a six-game chess match in 1997. The victory marked a significant milestone in the field of AI, showcasing the potential of machine intelligence in complex strategic games.

Watson

An advanced AI system developed by IBM that showcased its ability to understand and process natural language and provide accurate answers to complex questions. It's capabilities have since been applied in various fields, including healthcare, finance, and customer service.

AlphaGo

An AI program developed by DeepMind, a subsidiary of Alphabet Inc. that made headlines in 2016 when it defeated the world champion Go player, Lee Sedol, in a five-game match. The success showcased the potential of AI in surpassing human expertise in strategic decision-making.

HAL 9000

A fictional sentient computer system designed to assist and manage operations on a spacecraft. It's malfunction and subsequent conflicts with the human crew highlight potential risks and ethical dilemmas associated with advanced AI

Data Quiz

Which of the following are examples of cloud computing services?

- A Amazon Web Services
- B Hadoop
- C Microsoft Azure
- D Apache Spark

Data Quiz

Which of the following are examples of cloud computing services?

- A 🟩 Amazon Web Services
- B Hadoop
- C 🟩 Microsoft Azure
- D Apache Spark

Data Quiz

Amazon Web Services

A cloud computing platform offered by Amazon. It provides a wide range of cloud-based services, including computing power, databases, analytics, machine learning, and more, enabling businesses and individuals to build, deploy, and manage applications on the cloud.

Hadoop

An open-source framework that enables distributed processing and storage of large datasets across clusters of computers. It provides a scalable and cost-effective solution for processing and analyzing big data.

Microsoft Azure

A cloud computing platform offered by Microsoft. It provides a wide range of cloud-based services, including virtual machines, storage, databases, AI, analytics, and more, enabling businesses to build, deploy, and manage applications and services with flexibility and scalability on the cloud.

Apache Spark

An open-source distributed computing system designed for processing and analyzing large-scale datasets. It provides a fast and flexible framework for in-memory data processing, supporting a wide range of applications on big data analytics.

Data Quiz

Which of the following statements is NOT a feature of Blockchain?

A Centralization

B Immutability

C Security

D Transparency

Data Quiz

Which of the following statements is NOT a feature of Blockchain?

[✓ A Centralization]

B Immutability

C Security

D Transparency

Data Quiz

Centralization

Immutability

Security

Transparency

Blockchain operates on a decentralized network, eliminating the need for a central authority or intermediary. This distributed nature enhances transparency, security, and resilience by allowing multiple participants to validate and maintain the integrity of the shared ledger.

Once data is recorded on the blockchain, it becomes nearly impossible to alter or tamper with. Each transaction or data entry is linked to previous ones through cryptographic hashes, creating an immutable chain of information, that enhances the trustworthiness and integrity of the data stored.

Blockchain employs advanced cryptographic techniques to ensure the security and integrity of data. Each transaction is digitally signed and encrypted, and the decentralized consensus mechanism prevents unauthorized modifications.

All participants in a blockchain network can view and access the entire transaction history stored on the blockchain. This transparency fosters trust among network participants, as they can independently verify and validate transactions.

Data Quiz

Inferential statistics can be for which of the following purposes?

A Computing the range of the price of a house in a locality with certain precision

B Determining whether a change in the website layout helps in increasing the number of subscribers

C Determining whether the quality of coffee beans varies with the type of roasting process

D Monitoring and controlling processes in an automobile part manufacturing unit

Data Quiz

Inferential statistics can be for which of the following purposes?

A Computing the range of the price of a house in a locality with certain precision

B Determining whether a change in the website layout helps in increasing the number of subscribers

C Determining whether the quality of coffee beans varies with the type of roasting process

D Monitoring and controlling processes in an automobile part manufacturing unit

Data Quiz

Computing the range of the price of a house in a locality with certain precision

Statistical concepts like confidence intervals can be used to estimate the prices of houses with certain levels of precision

Determining whether a change in the website layout helps in increasing the number of subscribers

A/B testing is commonly used to determine if a change a variable affects a business KPI. Inferential statistics can be used for effective measurements if one or more variables are changed simultaneously

Determining whether the quality of coffee beans varies with the type of roasting process

Hypothesis testing techniques like ANOVA (ANalysis Of VAriance) are commonly used to compare multiple groups to determine if there are significant differences amongst them

Monitoring and controlling processes in an automobile part manufacturing unit

Statistical Process Control (SPC) helps businesses maintain and improve the quality of their products or services. By identify potential issues in advance by detecting trends, it enables businesses to take a proactive approach to problem-solving

Data Quiz

Which of the following is a practical application of supervised learning?

A Dividing the customers of an e-commerce platform into different segments

B Visualizing high-dimensional equipment sensor data in lower dimensions

C Predicting the price of a used car based on the attributes of the car

D Predicting the likelihood of a hotel reservation getting cancelled

Data Quiz

Which of the following is a practical application of supervised learning?

A Dividing the customers of an e-commerce platform into different segments

B Visualizing high-dimensional equipment sensor data in lower dimensions

C Predicting the price of a used car based on the attributes of the car

D Predicting the likelihood of a hotel reservation getting cancelled

Data Quiz

Dividing the customers of an e-commerce platform into different segments

Customers can be segmented into different categories based on their purchase and demographic attributes using unsupervised learning techniques like clustering algorithms.

Visualizing high-dimensional equipment sensor data in lower dimensions

High-dimensional data can be efficiently brought down to lower dimensions (2 or 3) for visualization purposes using unsupervised learning techniques, like PCA and t-SNE, while retaining the most important information.

Predicting the price of a used car based on the attributes of the car

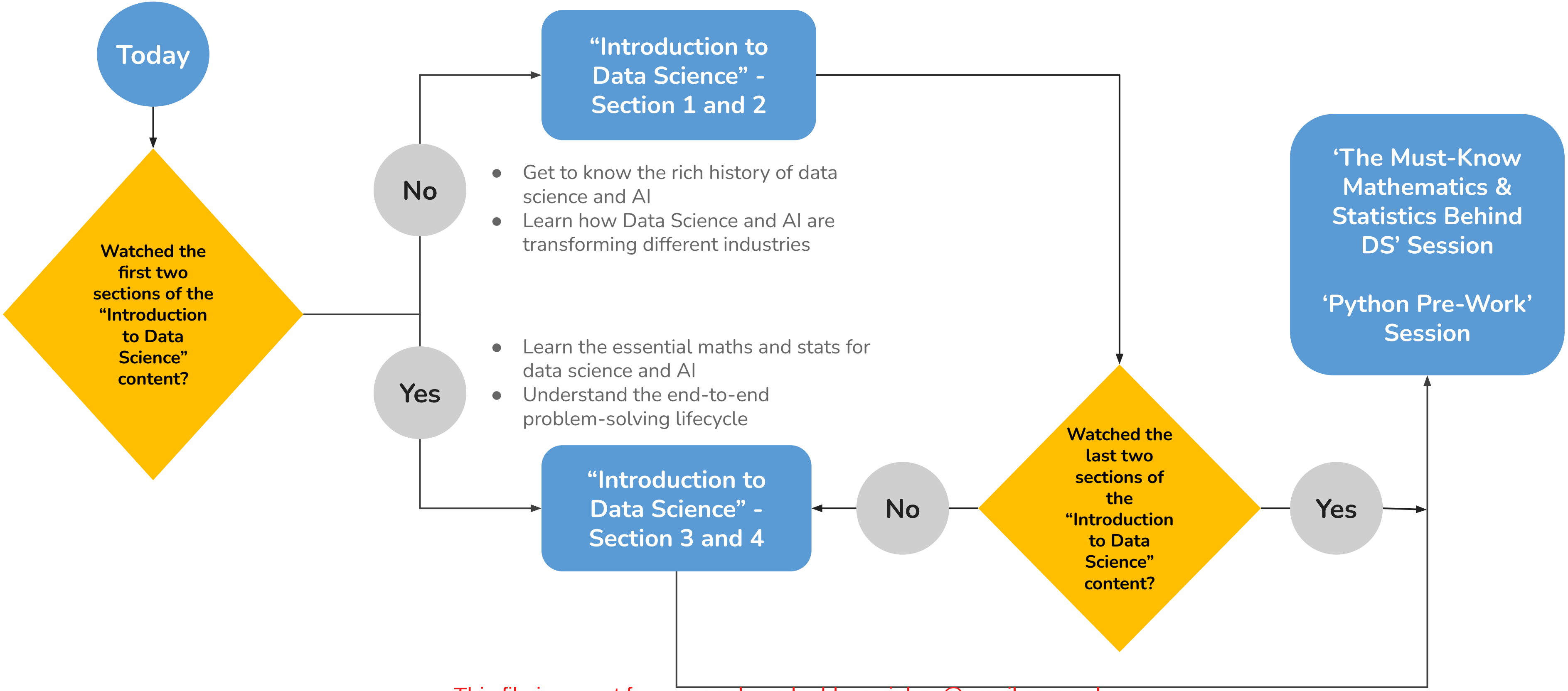
Algorithms from a subset of supervised learning, called regression, can be trained using historical data containing attributes like mileage, horsepower, manufacture year, distance driven, and more to determine the price of a used car

Predicting the likelihood of a hotel reservation getting cancelled

Algorithms from a subset of supervised learning, called classification, can be trained using historical data containing attributes like reservation lead time, room price, no. of guests, and more to determine the likelihood of cancellation

QnA

Next Steps


Today
Watched the first two sections of the "Introduction to Data Science" content?
No
Yes
"Introduction to Data Science" - Section 1 and 2
"Introduction to Data Science" - Section 3 and 4
Watched the last two sections of the "Introduction to Data Science" content?
No
Yes
'The Must-Know Mathematics & Statistics Behind DS' Session
'Python Pre-Work' Session

- Get to know the rich history of data science and AI
- Learn how Data Science and AI are transforming different industries

- Learn the essential maths and stats for data science and AI
- Understand the end-to-end problem-solving lifecycle

Thank you!

We'd love to hear your feedback!
Please share your feedback for the session

Wish you all the very best!

Please feel free to raise a Support Request through Olympus in case of any queries