

# Algorithms

Module 3

TECH TALENT  
ACADEMY |

WOMEN IN DATA  
ACADEMY |

# Session Content



WHAT IS AN  
ALGORITHM ?



ALGORITHMS FOR DATA  
SCIENCE



CASE STUDIES

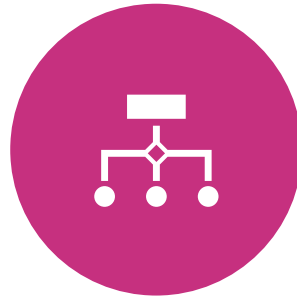
# What is an Algorithm?



ALGORITHMS ARE INSTRUCTIONS  
FOR PROBLEM SOLVING



IN COMPUTING, PROGRAMMERS  
WRITE ALGORITHMS THAT  
INSTRUCT THE COMPUTER HOW  
TO PERFORM A TASK.



**WHEN YOU WRITE AN  
ALGORITHM THE ORDER OF THE  
INSTRUCTIONS IS VERY  
IMPORTANT.**



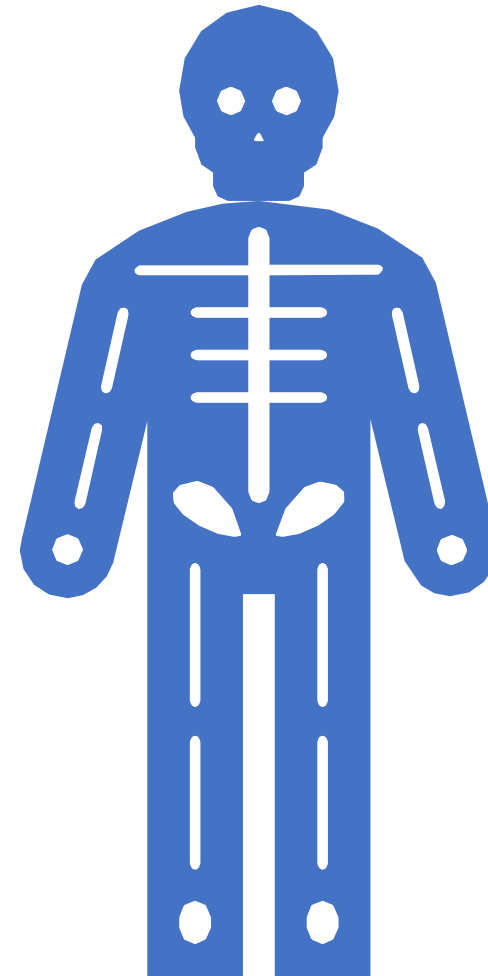
THINK ABOUT AN ALGORITHM  
FOR YOUR MORNING ROUTINE.  
YOU WOULDN'T GET DRESSED  
BEFORE TAKING A SHOWER!



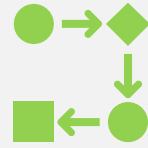
# The Brute Force Method

What approach would you use to find a name in a phonebook? The *brute-force algorithm*. On a good day, you go through the zillion names and find it quickly, that is, trying all possibilities till you get a solution that is good enough. A criminal can use the brute-force method, which is a trial-and-error approach, to crack data that is encrypted. This one doesn't work so well if the complexity (size) of the problem increases.

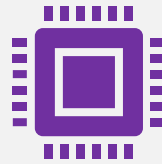
In the late 90s IBM's Deep Blue used brute force methods to defeat the then reigning chess champion, Garry Kasparov? We have moved on and computational abilities have progressed beyond belief.



# The Requirements of an Algorithm



Definiteness: Every step needs to be stated precisely



Effective Computability: Every step of the process can be carried out with a computer



Finiteness: In the end the program will successfully terminate

# Why do we need to use an Algorithm ?

An algorithm is a tool that provides a well-defined & efficient computational procedure to transform your data input into meaningful information that is output.

- Algorithm helps to **solve** computational problems.
- Algorithmic analysis will clear your vision about a computational problem & will let you go along way with it
- Algorithms are the core of most technologies that we see in modern electronic devices

(Quora)

# Types of Machine Learning

## Supervised Learning

### Classification

- Fraud Detection
- Image Classification
- Customer Retention
- Diagnostics

### Regression

- Forecasting
- Prediction
- Process Optimisation
- New Insights

## Unsupervised Learning

### Dimensionality Reduction

- Big Data Visualisation
- Meaningful Compression
- Structure Discovery
- Feature Elicitation

### Clustering

- Recommended Systems
- Targeted Marketing
- Customer Segmentation

## Reinforcement Learning

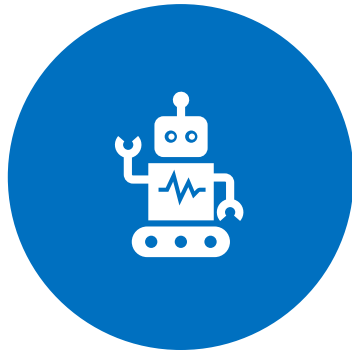
- Real-Time Decisions
- Gaming AI
- Learning Tasks
- Skill Acquisition
- Robot Navigation



# Supervised v Unsupervised Machine Learning

Machine Learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

## MAIN DIFFERENCES



IN SUPERVISED LEARNING, YOU TRAIN THE MACHINE USING DATA WHICH IS WELL LABELLED. IT MEANS SOME DATA IS ALREADY TAGGED WITH THE CORRECT ANSWER.



UNSUPERVISED MAINLY DEALS WITH THE UNLABELLED DATA. WHERE YOU DO NOT NEED TO SUPERVISE THE MODEL. INSTEAD, YOU NEED TO ALLOW THE MODEL TO WORK ON ITS OWN TO DISCOVER INFORMATION.



# Classification v Regression

Classification and Regression algorithms are both Supervised Learning algorithms. The algorithms are used for prediction in Machine learning and work with the labelled datasets.

## Key Differences



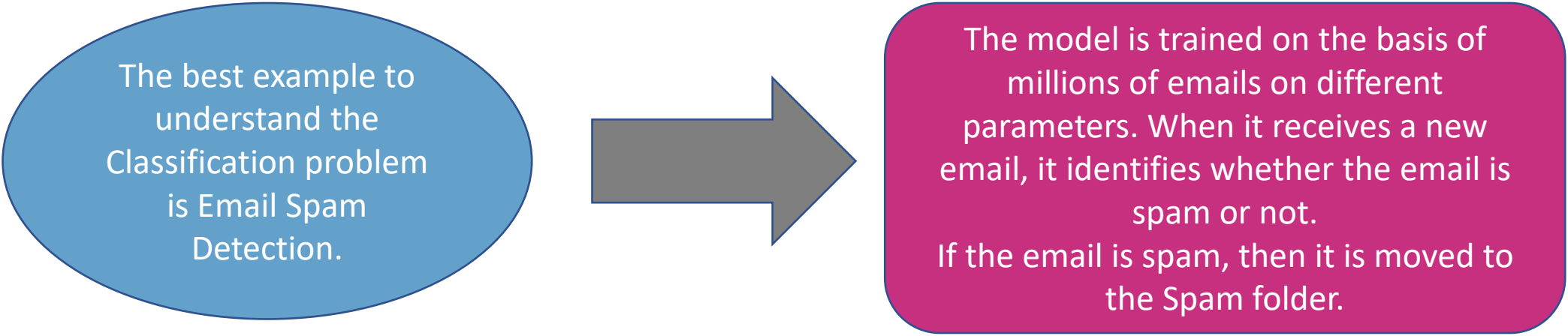
```
graph TD; A[Key Differences] --> B[Classification algorithms are used for predicting /classifying discrete values like Male/Female, True/False, Spam/Not Spam.]; A --> C[Regression algorithms are used for predicting continuous values such as price, salary, age.];
```

Classification algorithms are used for predicting /classifying discrete values like Male/Female, True/False, Spam/Not Spam.

Regression algorithms are used for predicting continuous values such as price, salary, age.

# Classification Algorithms

Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. The computer program is trained using a training dataset, based on that, it categorises the live data into different classes.



The best example to understand the Classification problem is Email Spam Detection.

A large grey arrow points from the blue oval on the left to the pink rounded rectangle on the right.

The model is trained on the basis of millions of emails on different parameters. When it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

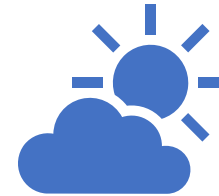
# Regression Algorithms



Regression is a process of finding the correlations between dependent and independent variables. It helps predict the continuous variables such as prediction of Market Trends and the prediction of House prices for example



Weather forecasting is a good example for the use of the Regression algorithm.



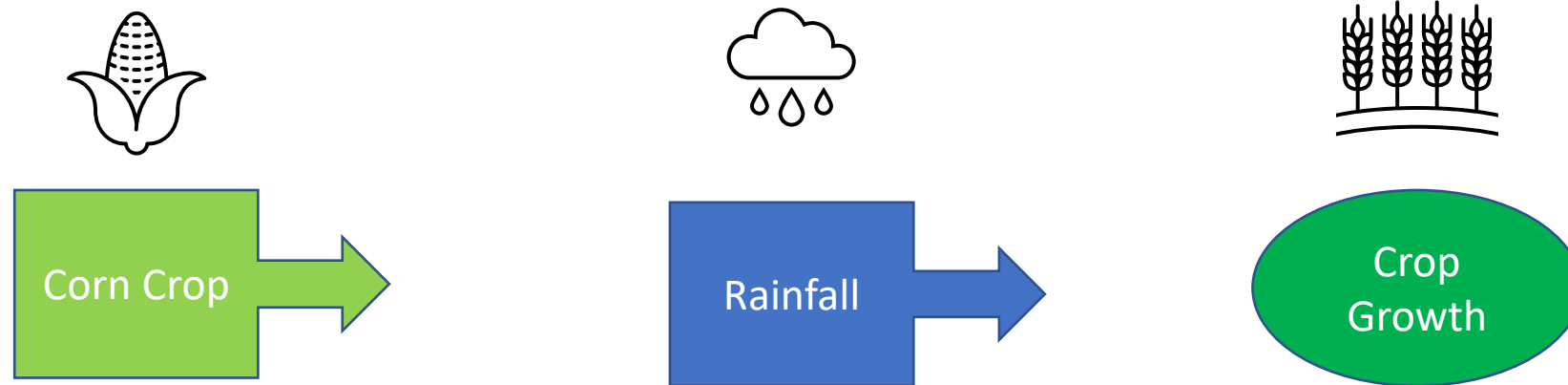
In weather prediction, the model is trained on the past data, and once the training is complete, it can easily predict the weather for future days.



# Linear Regression

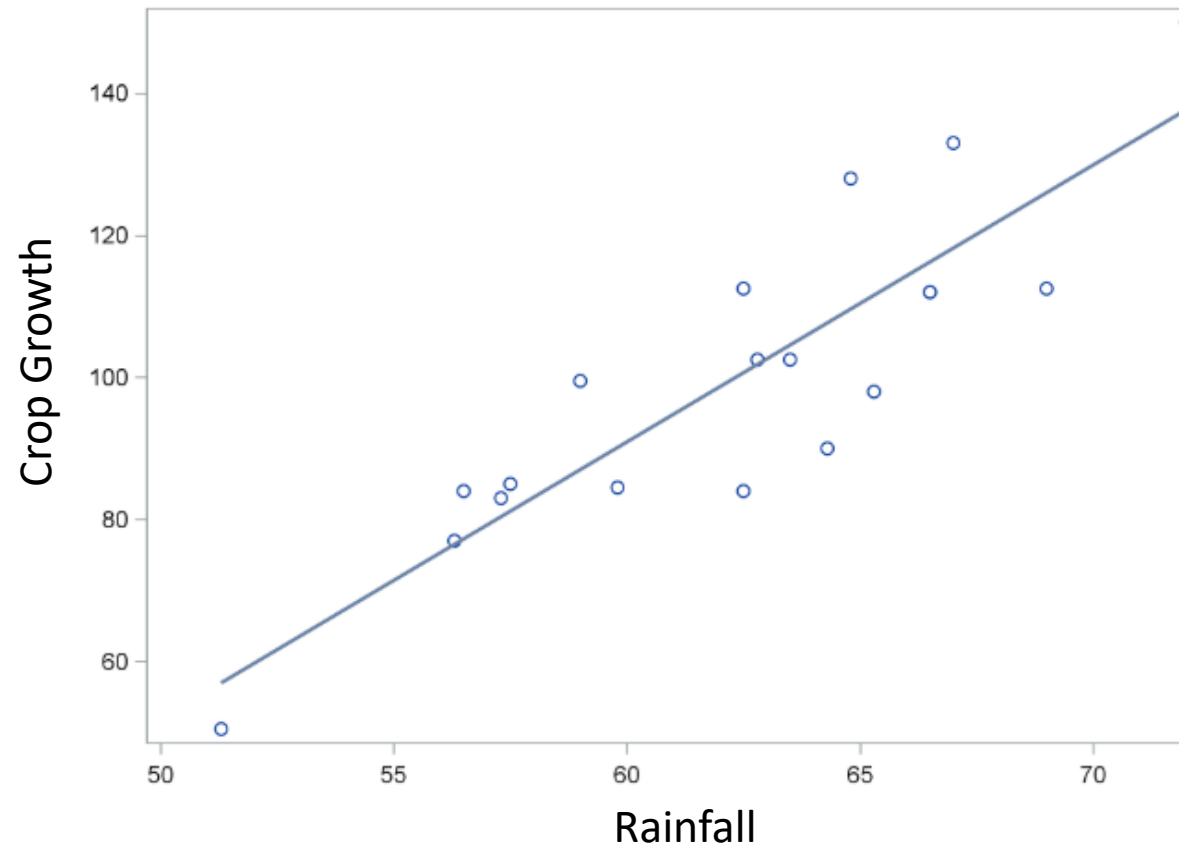
Linear regression is a basic and commonly used type of predictive analysis and modelling. It allows you to investigate if one variable is dependent on others.

## Example



# Linear Regression

Corn Crop Growth



# Dimensionality Reduction

Dimensionality Reduction is the process of making data digestible. It's the procedure of refining the relevant information from the disorder or getting rid of the unnecessary information. Before you can effectively analyse the data you need to clean it up and Dimensionality reduction helps you do that.

Some examples of techniques used for data-dimensionality reduction

- [Ratio of missing values](#)
- [Low variance in the column values](#)
- [High correlation between two columns](#)
- [Principal component analysis \(PCA\)](#)



# Clustering

Clustering is the term used to describe the exploration of data, where similar pieces of information are grouped.

These techniques are simple yet effective and although they can require some intense work, they often give us valuable insight into the data.

## Industry examples that use clustering techniques

**Biology**  
Genetic/Species  
Grouping

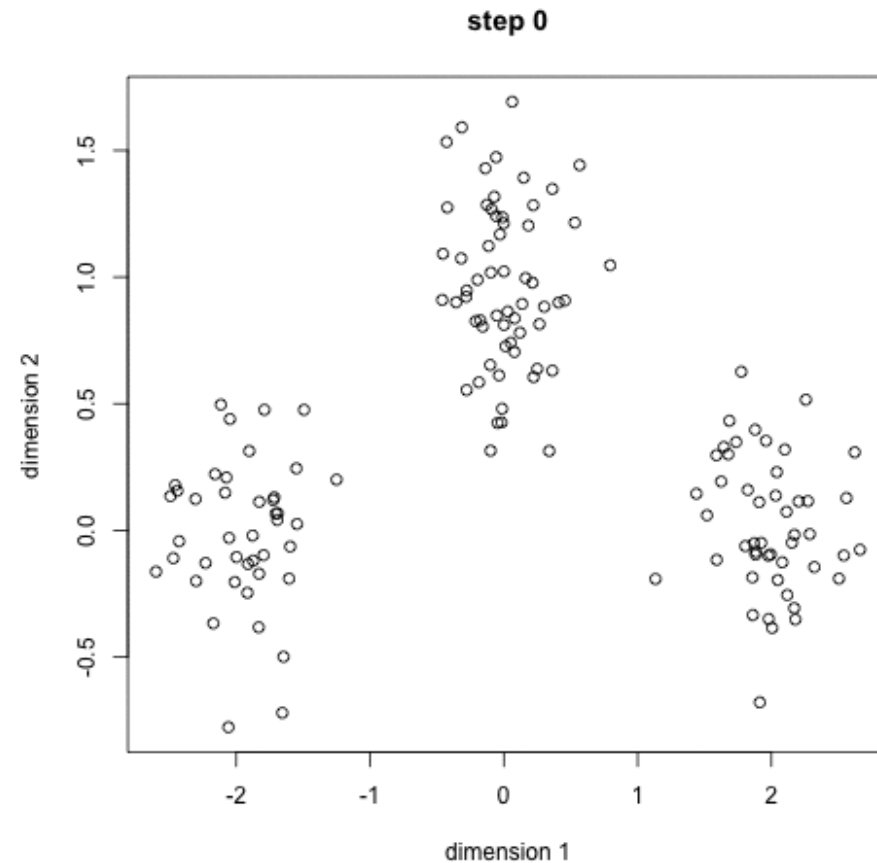
**Medical Imaging**  
Distinguishing tissue  
types

**Market Research**  
Differentiating groups of  
customers

**Recommended  
Systems**  
Amazon and Netflix  
suggestions

# K-Means Clustering

This is the most well-known **clustering algorithm**.



# Common Machine Learning Algorithms

- Linear Regression
- Logistic Regression
- Decision Tree
- SVM (Support Vector Machine)
- Naive Bayes
- KNN (K- Nearest Neighbours)
- K-Means
- Random Forest



# Lesson Tasks

Select one or more choices from the list of common Machine Learning Algorithms, do some investigations and write me a short summary. I am looking for the following:

- Is it Supervised/Unsupervised/Reinforcement learning?
- What does the algorithm do?
- In which situations will it be most useful?
- (Optional) Can you find any examples of where this algorithm has been used?





# WOMEN IN DATA ACADEMY

TECH TALENT  
ACADEMY