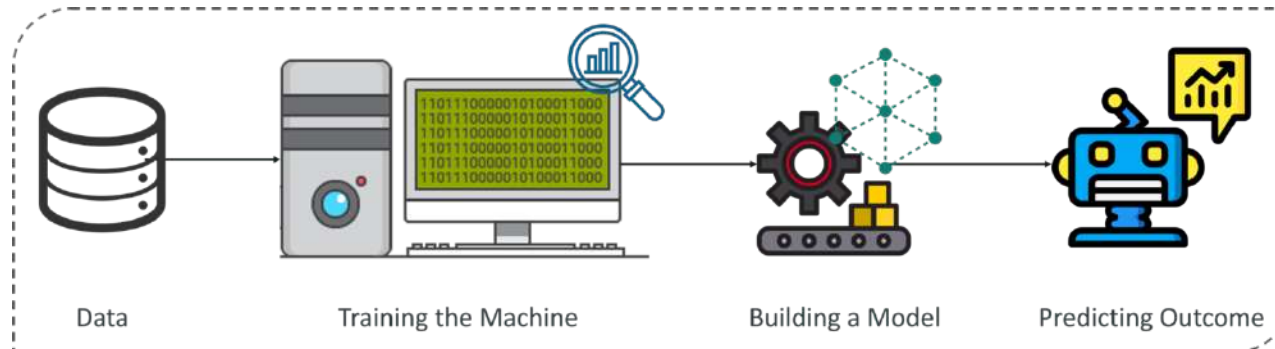


Artificial Intelligence and Machine Learning in Clinical Practice: Innovations, Challenges, and Future Directions



Dr. Upendra Kumar

Assistant Professor

Department of Computer Science & Engineering
&

Associate Dean of Student Welfare

Institute of Engineering & Technology, Lucknow

Agenda for today's session

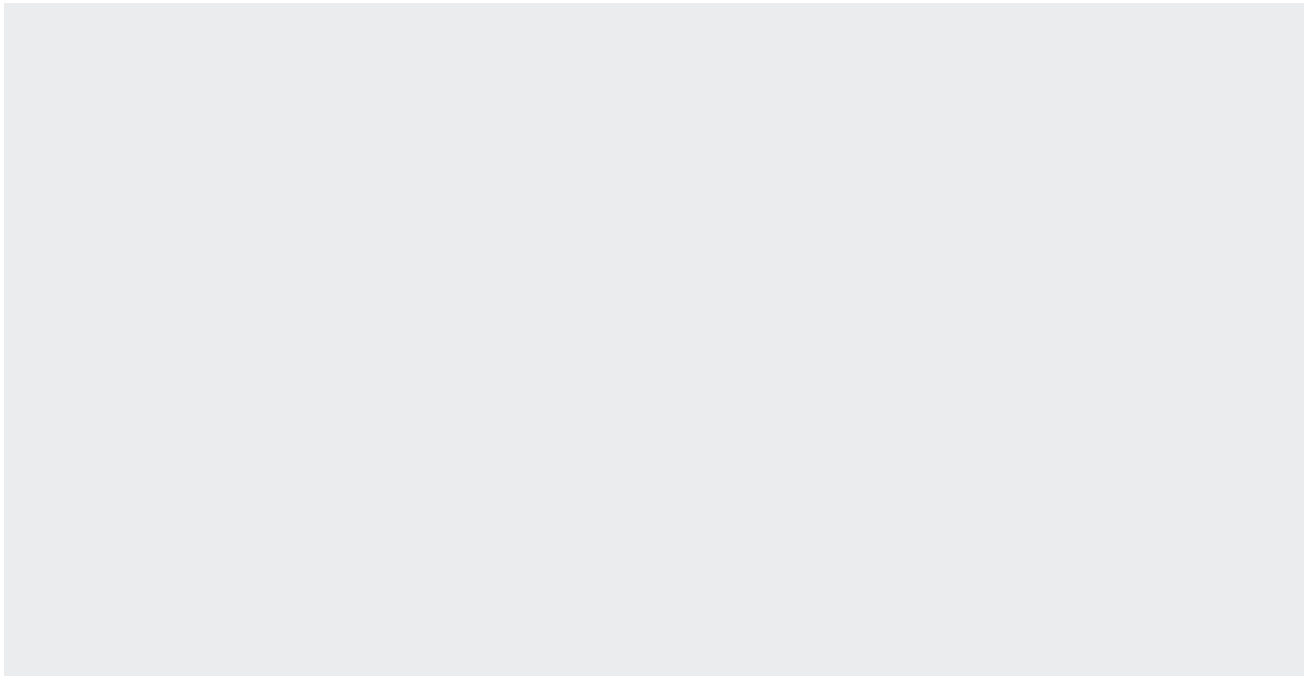
- Biomedical Science as healthcare science
- What is Machine Learning (ML)?
- What is a Machine Learning Algorithm?
- Why is Machine Learning Important for Healthcare Organizations?
- What are the Benefits of ML for Healthcare Providers and Patient Data?
- What are the types of Machine Learning Algorithms used in Healthcare?
- What are Supervised Learning, Unsupervised Learning and Reinforcement Learning Algorithm?
- List of Machine Learning Algorithms and their Applications
- Clustering Algorithms: K-Means and Hierarchical Clustering
- Ethics of Employing ML in Healthcare
- Challenges of ML in the Healthcare Industry
- Top companies using ML in healthcare
- ML based Biomedical Research Publications and Patent done by our Group
- Conclusion

What is Biomedical Science?

- Biomedical science is where healthcare meets science.
- It involves the study of the human body, examining the underlying causes of diseases to develop effective treatments and preventive measures.
- The field of biomedical science includes a wide range of disciplines including
 - anatomy (the study of the body's structure),
 - Physiology (scientific study of how living things function)
 - pharmacology (which looks at how drugs work in the body),
 - genetics (i.e., the role of genes in health and disease),
 - immunology (the study of the immune system and its role in defending the body against infection),
 - microbiology and pathology.
- By studying all these different aspects of the body and human health, biomedical scientists can better understand how to keep people healthy and treat illnesses.

Introduction to Artificial Intelligence

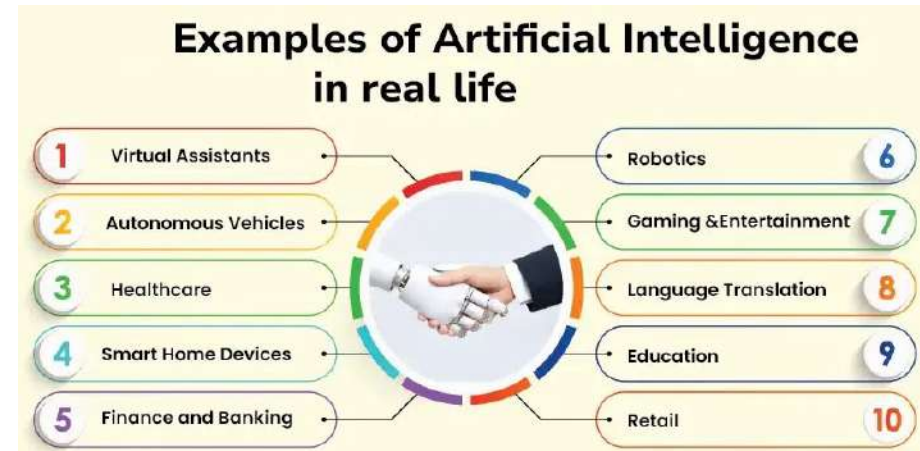
- AI term was coined by John McCarthy in 1955.
- Artificial intelligence (AI) is a field of study that uses machines and computers to perform tasks that usually require human intelligence.
- AI systems use algorithms, data, and computational power to learn, reason, and solve problems.
- AI can be used in many areas, including manufacturing, transportation, customer service, security, marketing, education, and healthcare, etc.



*Images and contents have been taken from either google server or AI tools

Examples of AI

- **Voice assistants:** Google Assistant, Siri, and Alexa
- **Recommendation systems:** Amazon, Netflix, and YouTube
- **Autonomous vehicles:** Waymo, Tesla
- **Chatbots:** Used in customer service
- **Smartphones:** Voice-assisted smartphones
- **Handwriting recognition:** Used for mail delivery
- **Facial recognition:** Used for payments (facial recognition payment (FRP)), ex. **SmilePay**
- AI has many applications, including healthcare, environmental sustainability, education, and public welfare, ex. **DeepMind AI, FireAid, GreenFi**

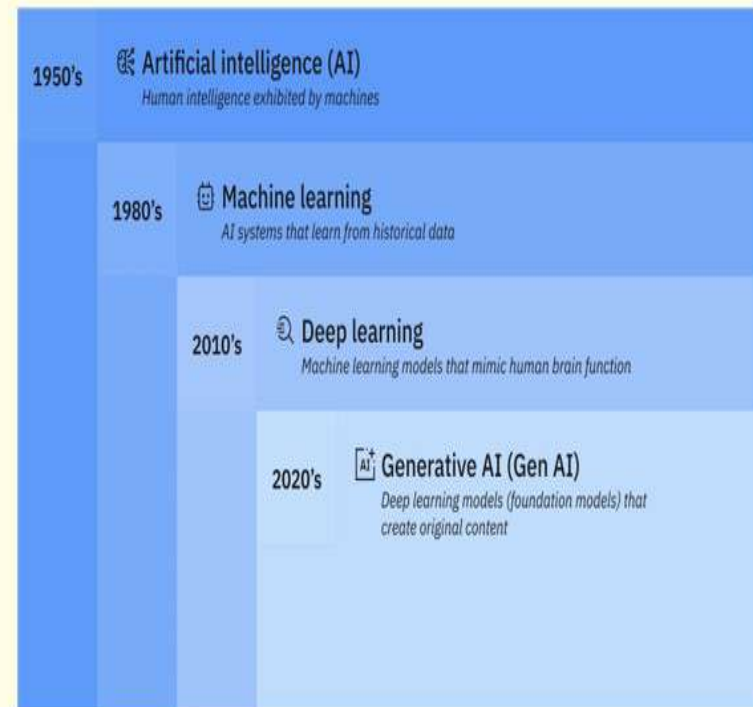
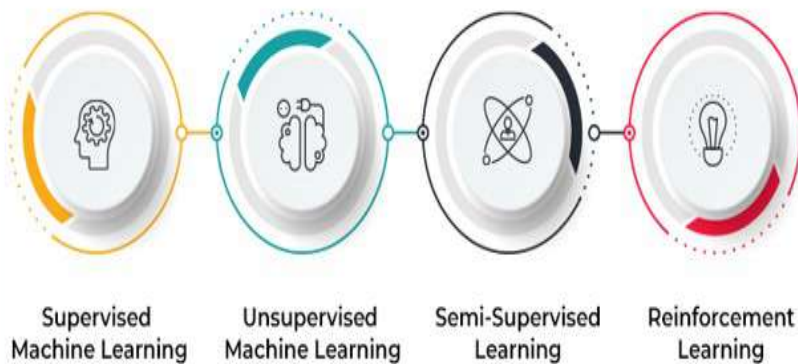


- Smart home devices that utilize AI include: smart thermostats, robot vacuums, smart security cameras, smart speakers with voice assistants like Alexa and Google Assistant, smart locks, video doorbells,
- and smart plugs, all of which can learn user patterns and adjust settings accordingly using machine learning algorithms to optimize convenience and efficiency in the home.

What is Machine Learning (ML)?

- **Machine learning** is a branch of artificial intelligence that enables algorithms to uncover hidden patterns within datasets, allowing them to make predictions on new, similar data without explicit programming for each task.

TYPES OF MACHINE LEARNING



What is Machine Learning (ML)?

- First coined by Arthur Samuel in the year 1959
- **Machine Learning (ML)** is a concept which allows the machine to learn from examples and experience, and that too without being explicitly programmed.
- Examples: learning by kids, student-teacher learning, self learning
- **Task:** How will you choose the best apples?
Now, imagine you were asked to write a computer program to choose your apples. You might write the following rules/algorithm:
if (bright red) and if (size is big): Apple is sweet.
if (crispy): Apple is juicy
- But every time you make a new observation (what if you had to choose **oranges**, instead) from your experiments, you have to modify the list of rules manually.
- **Issues:**
 - i) to understand the details of all the factors affecting the quality of the fruit
 - ii) difficult to make accurate rules by hand that covers all possible types of fruits.

This will take a lot of research and effort.

What is a Machine Learning Algorithm?

Alternative: This is where Machine Learning Algorithms come into the picture.

- So instead of you writing the code, what you do is you feed data to machine learning algorithm, and the algorithm/machine builds the logic based on the given data.
- It is an evolution of the regular algorithm. It makes your programs “**smarter**”, by allowing them to automatically learn from the data we provide.
- Two phases:
 - Training phase
 - Testing phase
- We can make our algorithm improve over time (**reinforcement learning**) so that it will improve its accuracy as it gets trained on more and more training dataset.
- In case it makes a wrong prediction it will update its rule by itself.
- The best part of this is, we can use the **same algorithm** to train different models (mangoes, grapes, bananas, or whichever fruit).

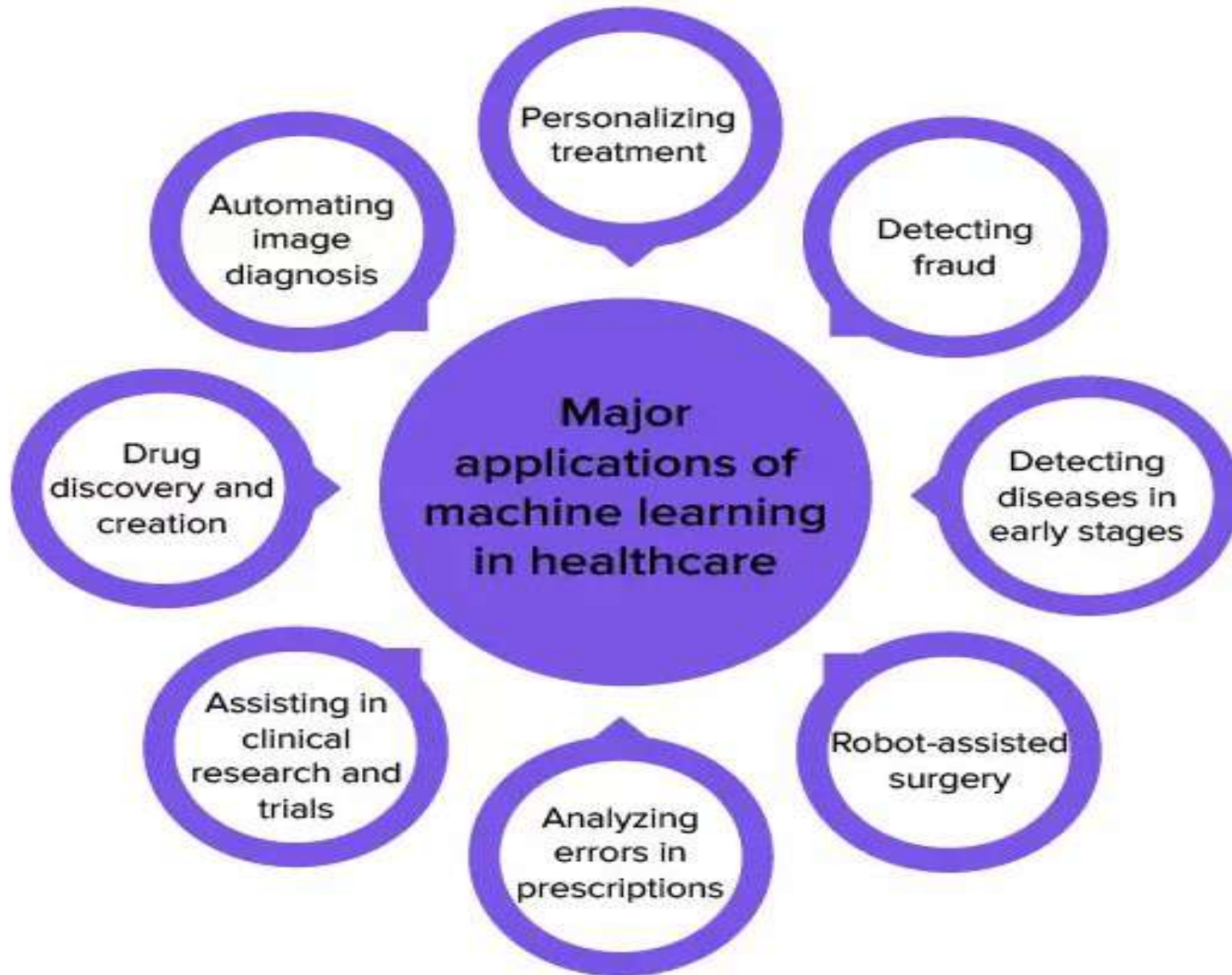
Why is Machine Learning (ML) Important for Healthcare Organizations?

- **ML** can help us make sense of the massive amounts of healthcare data that is generated every day within electronic health records (EHR- digital version of a patient's paper chart).
- **EHR-** Contain a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images (CT scan, MRI, X-Ray, Ultrasound etc.), and laboratory and test results.
- **ML** can help us find patterns and insights in medical data that would be impossible to find manually.
- **ML** allows healthcare professionals to focus on patient care rather than spend their time on information search or entry.
- **ML** helps in increasing diagnosing accuracy. For example, machine learning has proven to be 92% accurate in predicting the mortality of COVID-19 patients.
- **ML** in medicine can help to develop a more precise treatment plan. A lot of medical cases are unique and require a special approach for effective care and side effect reduction.
- **MD Anderson Cancer Center (Texas):** developed ML based software to predict acute toxicities in patients receiving radiation therapy for head and neck cancers (primary level screening)

Cont..

- Around 80% of medical data remains unstructured [1], without ML it is least useful despite of the great value of their application, such as
 - i) medical video/image data (endoscope, laparoscope, surgery robot, capsule endoscope, emergency video camera, thoracoscope, etc.)
 - ii) biosignal data (ECG, EEG, EMG etc.) that have been displayed on screen of patient monitor in operating rooms or intensive care units and wearable health monitoring devices.
 - iii) audio data that are verbally or nonverbally created from patients pathophysiologically and medical staffs for efficient communication in clinical procedures. Natural, unscripted speech can serve as a **biomarker for mental illnesses, neurological disorders** (e.g., Alzheimer's), and **respiratory conditions** (e.g., pulmonary hypertension). Analysis of voice, including **tone, pauses, and pitch**, can **indicate depression or Parkinson's**.
- Human language, or “natural language”, is very complex, lacking uniformity and incorporates an enormous amount of ambiguity, jargon, and vagueness.
- For better learning in ML, we need to convert these documents into more useful and analyzable data, ML relies on artificial intelligence like **natural language processing (NLP)** programs.

Benefits of Machine Learning In Biomedical and Healthcare



Cont..

1) Clinical Decision Support Systems (CDSS)

- It helps to analyze large volumes of data to **identify a disease, decide on the next treatment stage, determine any potential problems, and overall improve patient care efficiency.**
- CDSS is a powerful tool that helps the physician do their job efficiently and quickly, and it reduces the chances of getting the wrong diagnosis or prescribing ineffective treatment
- Examples of companies involved to make these tools: Epic, Change Healthcare, and Oracle Cerner.
- CDSS tools:

Tool	Functions
RAMPmedical	Alerts & Reminders
Medical Algorithms Company	Documentation Forms & Templates
HERA-MI	Diagnostic Decision Support

Cont..

2) Smart Recordkeeping:

- Making sure that all the patient records are updated regularly is challenging, as data entry is a monotonous task.
- However, it is also crucial for effective decision-making and better patient care.
- One of the uses of machine learning in healthcare is using **optical character recognition** (OCR) technology on physicians' handwriting, making the data entry fast and seamless.
- ML can significantly streamline some of the most time-consuming tasks, such as identifying and classifying documents. By looking at word patterns in a document, algorithms can interpret the **subject matter, entities involved, dates, record numbers, status, relationship to similar documents** and a variety of other characteristics.

3) Machine Learning in Medical Imaging:

- Medical images such as MRI, CT or X-ray images for automated pathology detection, a task that requires a trained specialist's eye for detection e.g. **labelling, localization, association with treatment or prognostics and personalized medicine**.
- According to a recent meta-analysis (Liu et al., 2019), ML algorithms do the job as well as (and, in some cases, even better) human specialists, with 87.0% sensitivity and 92.5% specificity for the **ML** algorithms and 86.4% sensitivity and 90.5% specificity for human physicians.
- Tumor detection, cancer detection, Medical imaging in pregnancy, identifying cardiovascular abnormalities, detecting fractures and other musculoskeletal injuries¹³

Cont..

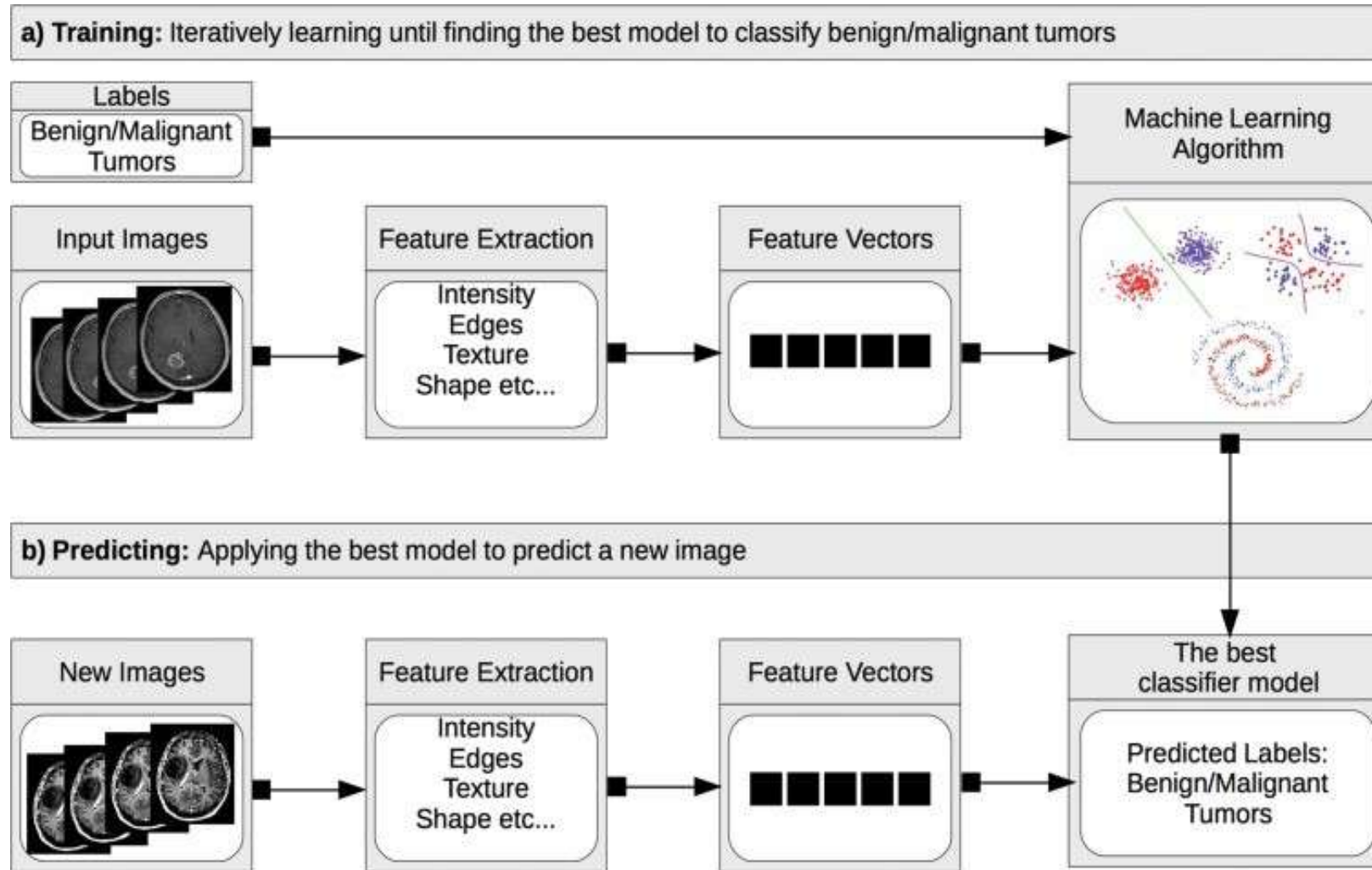


Fig. 1 Machine Learning in Medical Imaging to detect benign/malignant tumors [11]

Cont..

4) Personalized Treatment:

- ML allows healthcare organizations to deliver personalized patient care by analyzing patients' medical history, symptoms, and tests.
- Using ML, doctors can develop customized treatments and prescribe medicines that target specific diseases in individual patients.
- With ML, healthcare organizations can also have access to the analysis based on electronic health records for the patient. This helps doctors make faster decisions on what kind of treatment best suits the patient.
- ML can assist doctors in finding out if the patient is ready for necessary changes in medication. This helps induce the right treatment from the beginning.

3) Behavior adjustments:

- Prevention is as important in healthcare as disease treatment. One of the most important parts of preventive medicine is modifying one's behavior to get rid of unhealthy habits and establish a healthy lifestyle.
- ML can be used to point out something we don't notice.
- **Somatix-** ML based tool follows the patient's daily activity and points out their unconscious habits and routine so that they can focus on getting rid of them

Cont..

4) Detecting diseases in early stages

- A lot of diseases that you need to detect in the early stages to identify the treatment plan and help patients secure a good way of living.
- ML can be used to successfully predict some of the most dangerous diseases in at-risk patients.
- This includes the identification of signs of diabetes (using a Naïve Bayes algorithm), liver and kidney diseases, and oncology.

5) Elderly and low-mobility groups care

- ML and medicine can help low-mobility groups (including the elderly and people using wheelchairs) improve their day-to-day lives.
- ML helps them with smart reminders and scheduling help, predict and avoid potential injuries by identifying common obstacles and determining the optimal paths, and acquire help as soon as needed.

Cont..

6) Robot-assisted Surgery

- **ML-powered** surgical robots have revolutionized surgeries in terms of **accuracy** and **speed**.
- These systems can perform complicated surgical procedures with reduced blood loss, side effects, or pain risks. Additionally, **post-surgery** recovery is much faster and easier.
- **AIIMS New Delhi, Fortis Hospital Gurgaon, Manipal Hospital, Bangalore, Artemis Hospital, Gurgaon, Kokilaben Hospital, Mumbai, BLK Max Hospital New Delhi.**

7) Drug Discovery and Production

- **ML** has the ability to discover new drugs that offer great economic value for pharmaceuticals, hospitals, and new treatment opportunities for patients.
- **ML** also makes the process of drug creation faster and extremely cost-effective (**several years reduced to one day**). In the future, ML tool could be used in combination with **nanotechnology** for better drug delivery.
- In 2015, **Atomwise** used its **ML and super computer** to find out the current medicines in the market that could be redesigned for treating the **Ebola** virus. They successfully found two drugs that could help reduce the risks of the epidemic.

Cont..

8) Assistance in Clinical Research

- Clinical research and trials are costly and lengthy processes.
- There is a good reason behind this – new drugs and medical procedures should be proven to be safe before being used widely. However, there are cases when the solution needs to be released as soon as possible – like with the **vaccines for COVID-19**.
- Fortunately, there is a way to make the process shorter with the help of **ML** algorithms. It can be used to determine the best sample for the trial, gather more data points, analyze the ongoing data from the trial participants, and reduce the data-based errors.

9) Infectious Disease Outbreak Prediction

- **COVID-19** pandemic has shown us how unprepared we were to an infectious disease outbreak of this size.
- It is worth mentioning that experts in the area have warned the government about the possibility of such an event for years.
- Now, we have ML based tools that can help to detect the signs of an epidemic early on. The algorithms analyze the **satellite data, news, and social media reports, even video sources** to predict whether the disease has the potential to grow out of control.

*Some of the contents and images have been taken from google server or AI tools

Sensitivity vs. Specificity

- Sensitivity and specificity are two important metrics used in a statistical or ML model to assess if the model is performing successfully. As such, it is important to understand what each of these numbers tell us about what a trained model can do, and what the model cannot do.
- **Sensitivity** is the probability that a positive result occurs given that the sample is indeed positive. Mathematically,
- **Sensitivity**=(Number of True Positives)/Number of True Positives+Number of False Negatives

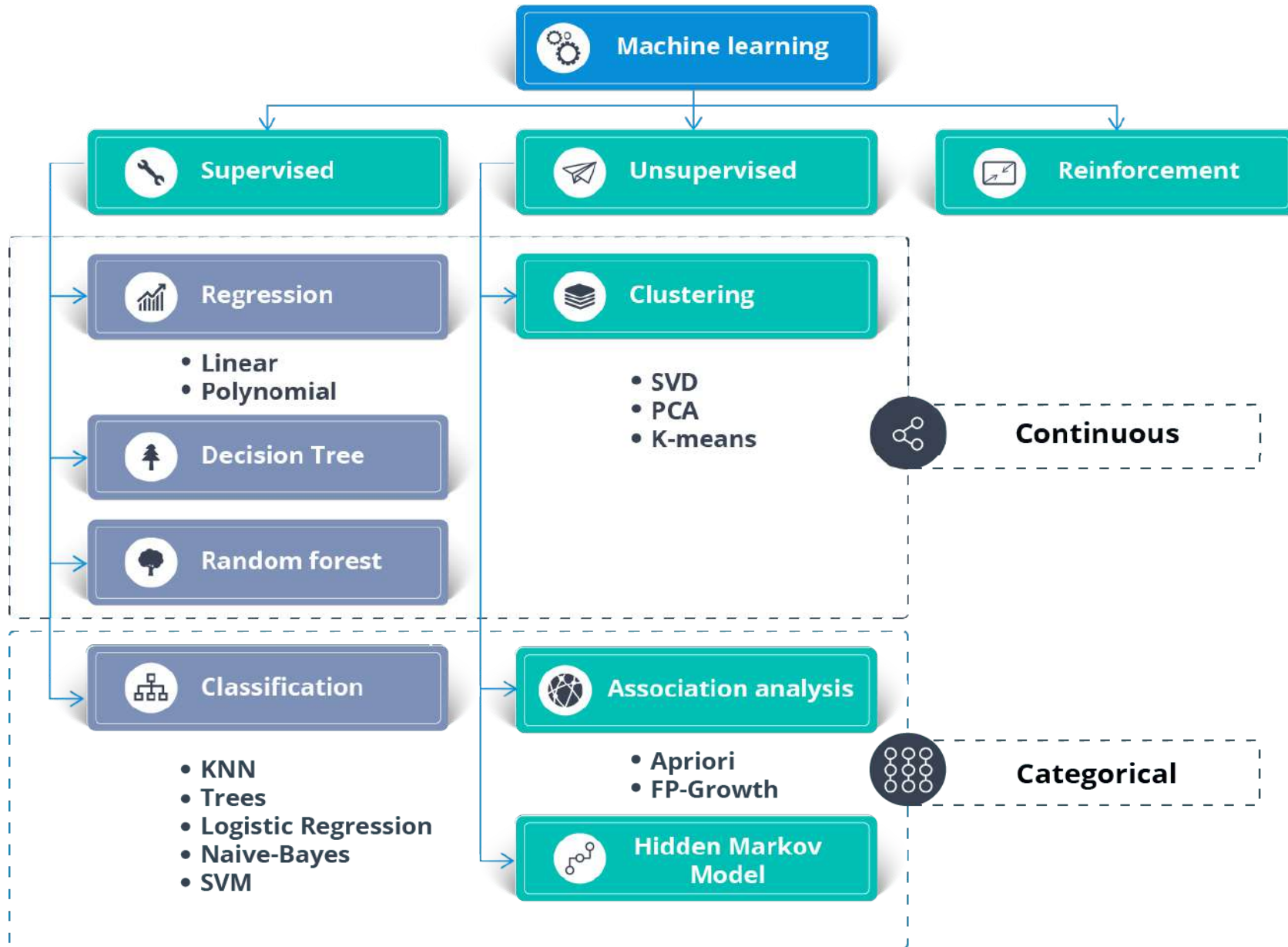
$$\text{Sensitivity} = \text{TP} / \text{TP} + \text{FN}$$

- This is also sometimes referred to as the **recall or hit rate**, or just simply the true positive rate, and the sensitivity is equivalent to 1–False Negative Rate
- **Specificity** is the probability of a negative result given that the sample is negative. Mathematically,
- **Specificity** = Number of True Negatives / Number of True Negatives+Number of False Positives

$$\text{Specificity} = \text{TN} / \text{TN} + \text{FP}$$

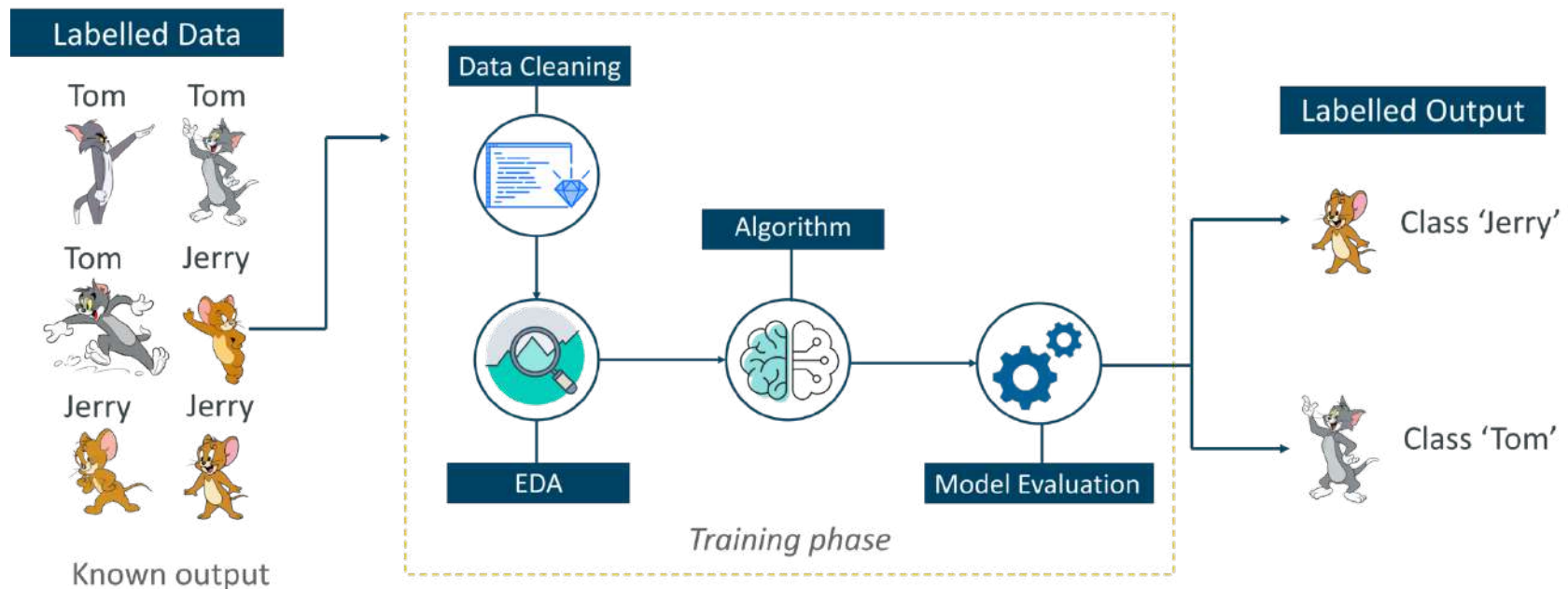
- This value is also referred to as the selectivity of the test. This is equivalent to 1–False Positive Rate

Types of Machine Learning Algorithms Used in Healthcare



What is Supervised Learning?

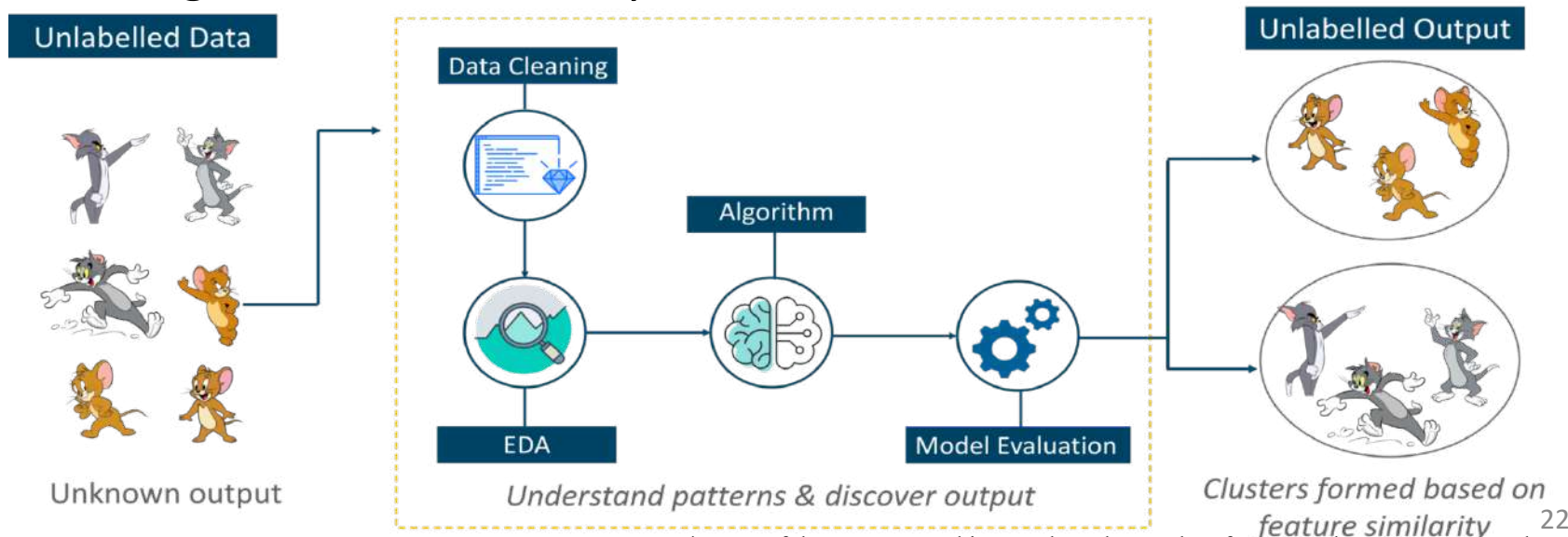
- Supervised learning is a technique in which we teach or train the machine using data which is well labelled.
- The process of an algorithm learning from the training dataset can be thought of as a teacher teaching his students.
- The algorithm continuously predicts the result on the basis of training data and is continuously corrected by the teacher.



*Some of the contents and images have been taken from google server or AI tools

What is Unsupervised Learning?

- Unsupervised learning involves training by using unlabelled data and allowing the model to act on that information without guidance.
- Algorithms are left on their own to discover and return the interesting structure in the data.
- Think of unsupervised learning as a smart kid that learns without any guidance.
- Example: it classifies the images into two different classes without knowing who Tom is or Jerry is.



*Some of the contents and images have been taken from google server or AI tools

What is Reinforcement Learning?

- It can be thought of like a hit and trial method of learning.
- The machine gets a Reward or Penalty point for each action it performs. If the option is correct, the machine gains the reward point or gets a penalty point in case of a wrong response.
- It is all about the interaction between the environment and the learning agent.
- The learning agent is based on exploration and exploitation.
- **Exploration** is when the learning agent acts on trial and error and **Exploitation** is when it performs an action based on the knowledge gained from the environment.

Cont..

- The environment rewards the agent for every correct action.
- With the aim of collecting more rewards obtained, the agent improves its environment knowledge to choose or perform the next action.
- Imagine that you were **dropped off at an isolated island**! What would you do?
- **Panic? Yes**, of course, initially we all would. But as time passes by, you will learn how to live on the island.
- You will explore the environment, understand the climate condition, the type of food that grows there, the dangers of the island, etc.
- This is exactly how Reinforcement Learning works, it involves an Agent (you, stuck on the island) that is put in an unknown environment (island), where he must learn by observing and performing actions that result in rewards.

List of Machine Learning Algorithms

- Linear Regression
- Logistic Regression
- Decision Tree
- Naive Bayes
- kNN (k-Nearest Neighbor)
- Artificial Neural Network (ANN)
- Support Vector Machines (SVM)
- Random Forest
- Self Organization Map (SOM)
- So many others (CNN, DBN, etc.)

Linear Regression

- It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on continuous variables.
- Here, we establish a relationship between the independent and dependent variables by fitting the best line.
- This best fit line is known as the *regression line* and represented by a linear equation

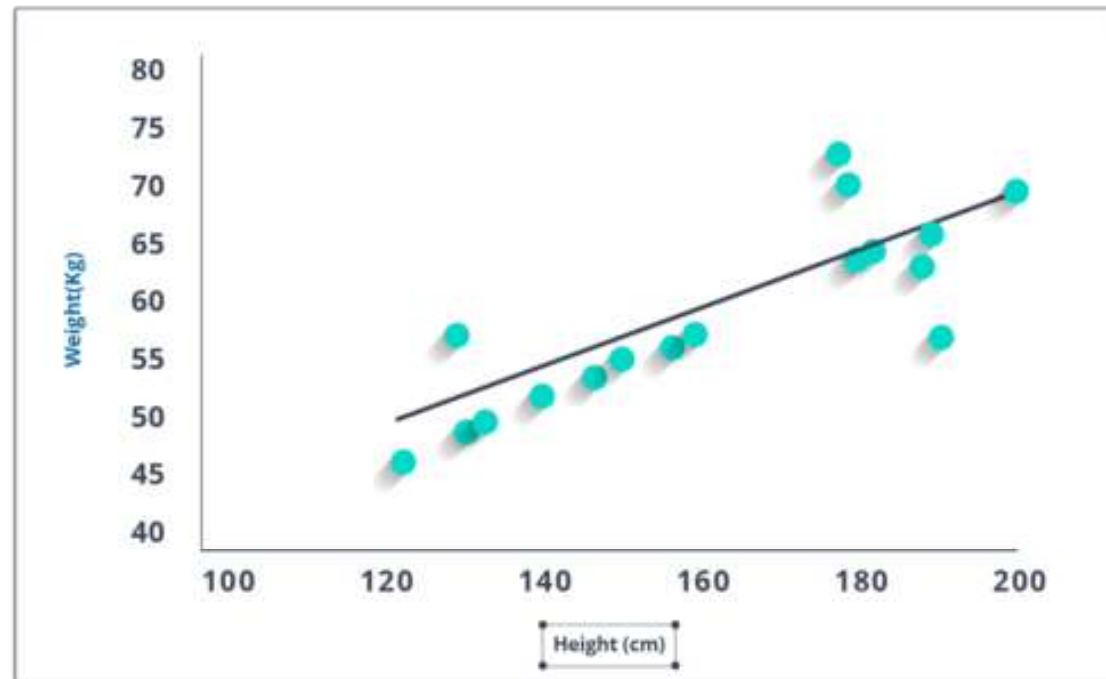
$$Y = aX + b.$$

Y – Dependent Variable, a – Slope, X – Independent variable

b – Intercept

- *Example:* Ask a child in fifth grade to arrange people in his class by increasing order of weight, without asking them their weights!
- This is a **linear regression** in real life! The child has actually figured out that height and build would be correlated to the weight by a relationship, which looks like the equation above.

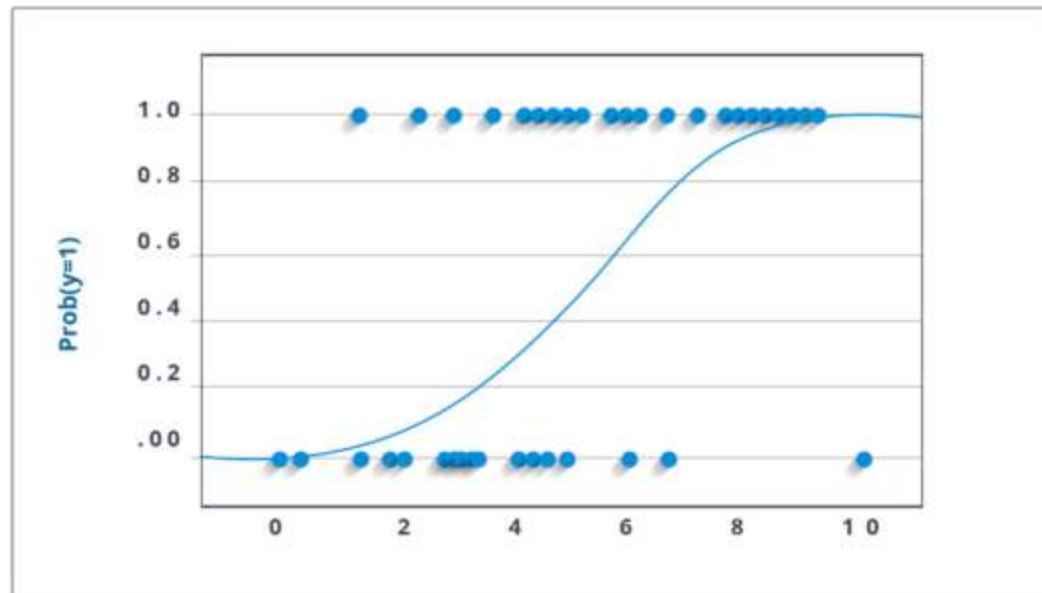
Cont..



- These coefficients **a** and **b** are derived based on minimizing the 'sum of squared differences' of distance between data points and regression line.
- Multiple Linear Regression: $y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon$

Logistic Regression

- Don't get confused by its name! It is a classification, and not a regression algorithm.
- It is used to estimate discrete values ([Binary values like 0/1, yes/no, true/false](#)) based on a given set of independent variable(s).
- In simple words, it predicts the probability of occurrence of an event by fitting data to a *logit function*. Hence, it is also known as [logit regression](#). (fraud detection, disease diagnosis, spam-non spam)
- Since it predicts the probability, its output values lie between 0 and 1.

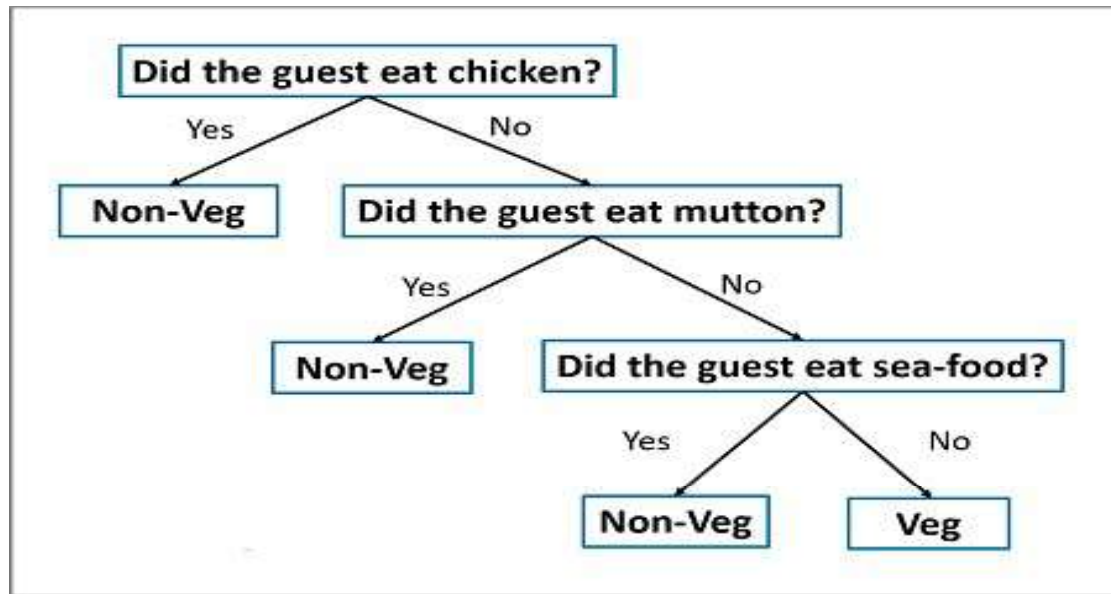


*Some of the contents and images have been taken from google server or AI tools

Decision Tree

- It is a type of supervised learning algorithm that is mostly used for classification problems.
- Surprisingly, it works for both categorical and continuous dependent variables
- In this algorithm, we split the population into two or more homogeneous sets.
- A Decision Tree looks like an **inverted tree**, wherein each node represents a predictor variable (feature), the link between the nodes represents a Decision and each leaf node represents an outcome (response variable).
- *Example*: you hosted a huge party and you want to know how many of your guests were non-vegetarians.

Cont..



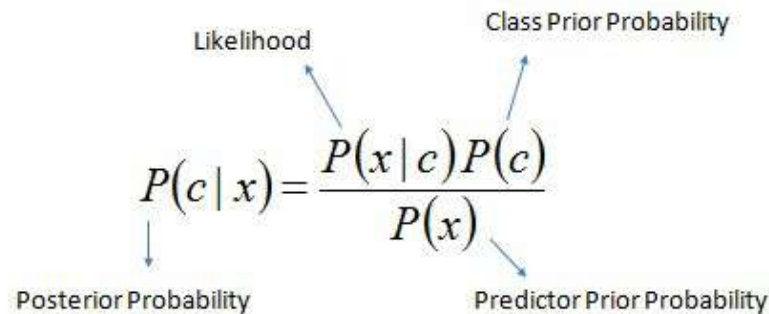
- Decision tree that classifies a guest as either vegetarian or non-vegetarian.
- Each node represents a **predictor variable** that will help to conclude whether or not a guest is a non-vegetarian.
- As you traverse down the tree, you must make decisions at each node, until you reach a dead end (**leaf node-gives output**).

Naive Bayes Classifier

- Based on *Bayes' theorem* with an assumption of independence between predictors.
- It assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature
- The assumption that all features are independent makes naive bayes algorithm very fast compared to complicated algorithms. In some cases, speed is preferred over higher accuracy.
- It works well with high-dimensional data such as text classification, email spam detection.
- Naive Bayesian model is easy to build and particularly useful for very large data sets.
- Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Cont..

- Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$.



The diagram shows the Bayes' Theorem formula: $P(c|x) = \frac{P(x|c)P(c)}{P(x)}$. Arrows point from labels to the corresponding terms in the formula: 'Likelihood' points to $P(x|c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c|x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Here,

$P(c/x)$ is the posterior probability of *class (target)* given *predictor (attribute)*.

$P(c)$ is the prior probability of *class*.

$P(x/c)$ is the likelihood which is the probability of *predictor* given *class*.

$P(x)$ is the prior probability of *predictor*.

Example: We have a training data set of weather and corresponding target variable 'Play'. Now, we need to classify whether players will play or not based on weather condition.

Cont..

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Problem: Players will play if the weather is sunny, is this statement is correct?

We can solve it using above discussed method,

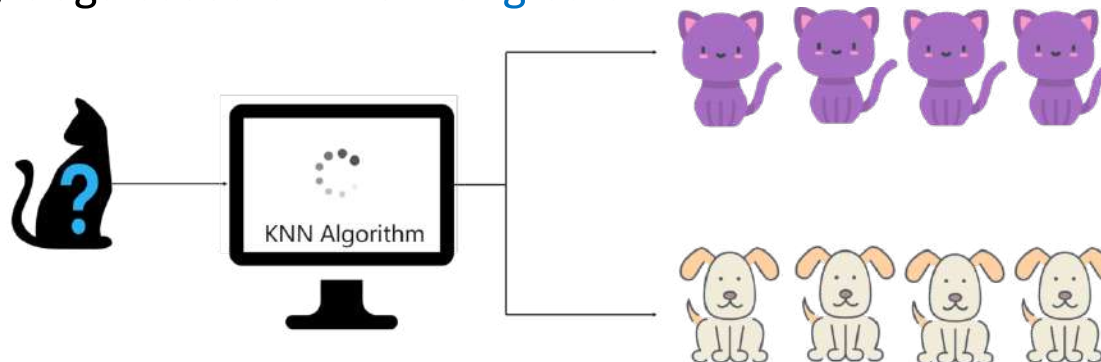
so $P(\text{Yes} \mid \text{Sunny}) = P(\text{Sunny} \mid \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$

Here we have $P(\text{Sunny} \mid \text{Yes}) = 3/9 = 0.33$, $P(\text{Sunny}) = 5/14 = 0.36$, $P(\text{Yes}) = 9/14 = 0.64$

Now, $P(\text{Yes} \mid \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$, which has higher probability.

kNN (k- Nearest Neighbours)

- It can be used for both classification and regression problems.
- It is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbours.
- The case being assigned to the class is most common amongst its K nearest neighbours measured by a distance function ([Euclidean](#), [Manhattan](#), [Minkowski](#) and [Hamming distance](#)).
- **Example:** If you want to learn about a person, of whom you have no information, you might like to find out about his close friends and the circles he moves in and gain access to his/her information!
- Let's say we want a machine to distinguish between images of cats & dogs.
- Features such as [pointy ears](#) can be used to identify cats and similarly we can identify dogs based on their [long ears](#).



*Some of the contents and images have been taken from google server or AI tools

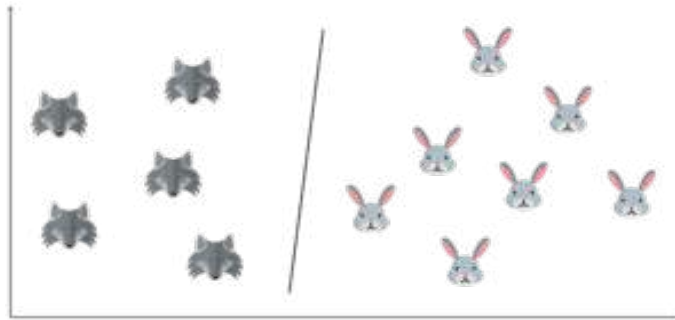
kNN (k- Nearest Neighbours)



As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

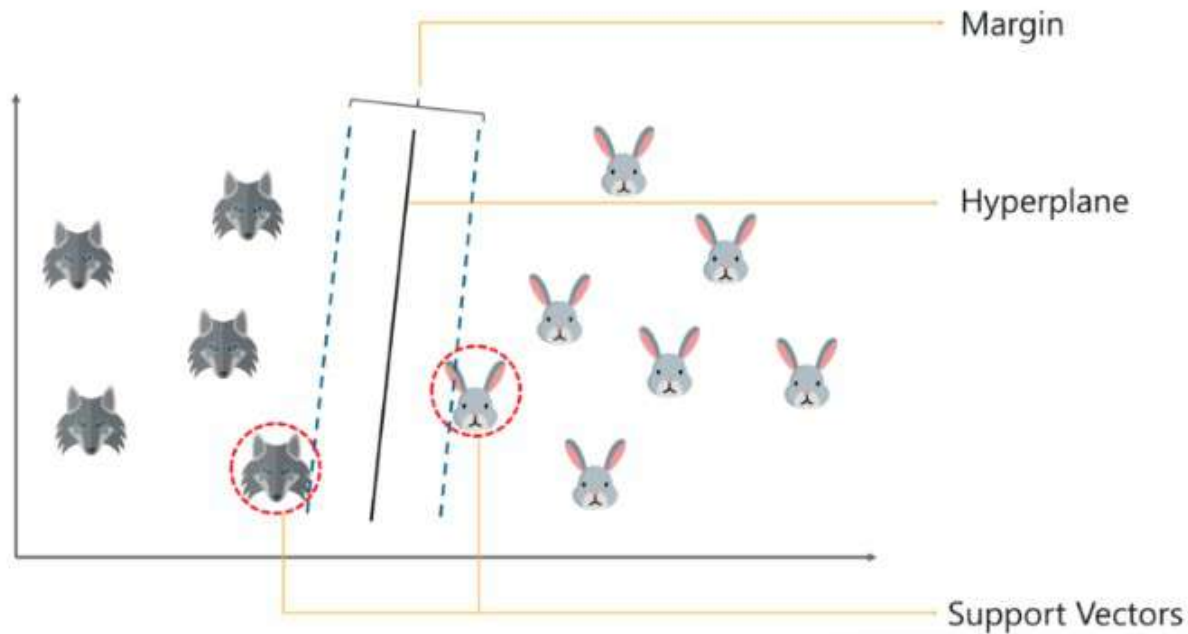
Support Vector Machine (SVM)

- It is a supervised machine learning algorithm which is mainly used to classify data into different classes.
- Unlike most algorithms, SVM makes use of a hyperplane which acts like a decision boundary between the various classes.
- A main advantage of SVM is that it can be used for both classification and regression problems.
- **Example:** suppose we own a farm and we have a problem—we need to set up a fence to protect your rabbits from a pack of wolves. But where do we build our fence?
- So if I do that, and try to draw a decision boundary between the rabbits and the wolves, Now we can clearly build a fence along this line. this is exactly how SVM works

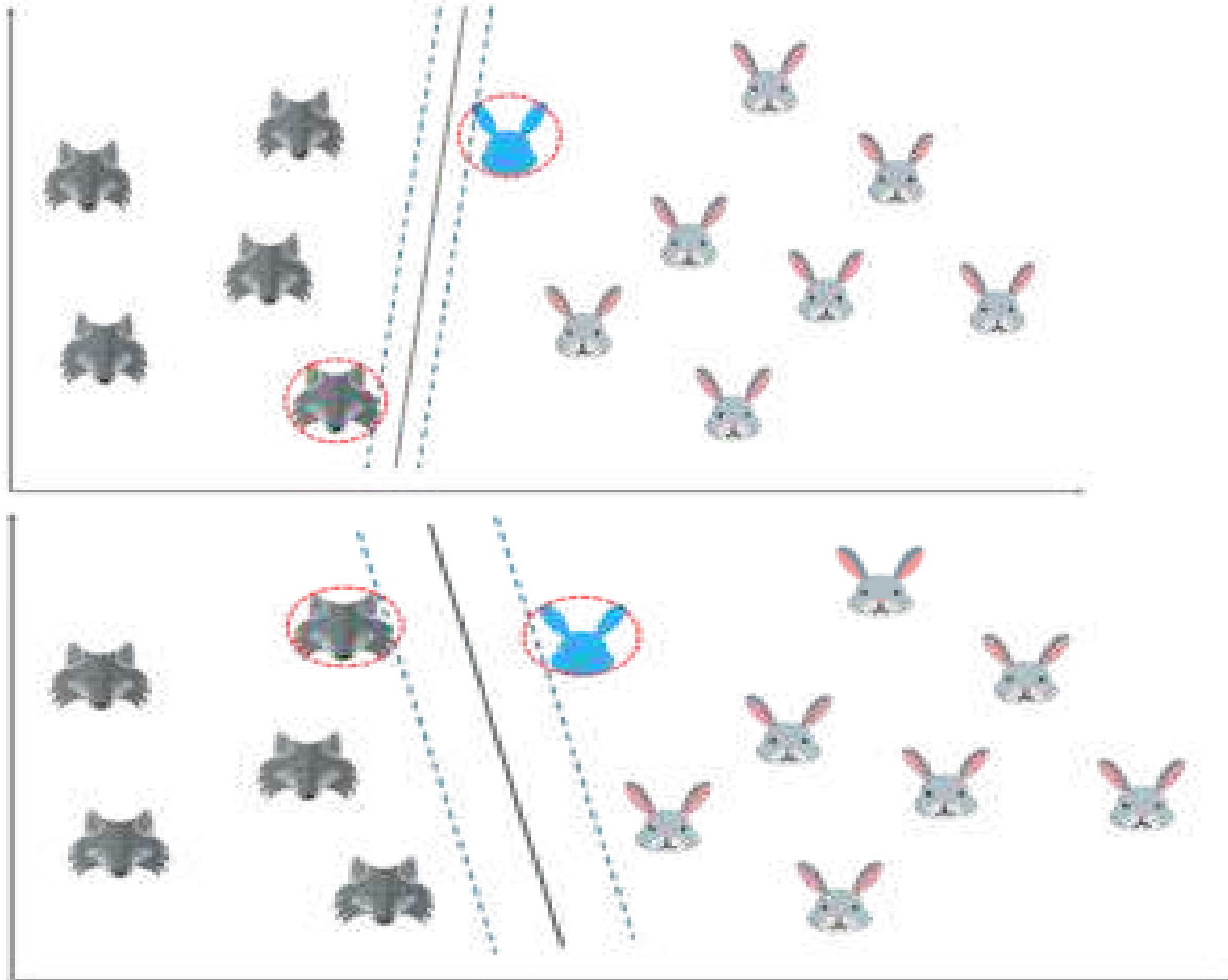


What is a Support Vector in SVM?

- Draw a random hyperplane and then check the distance between the hyperplane and the closest data points from each class.
- These closest data points to the hyperplane are known as support vectors.
- SVM is used to classify data by using a hyperplane, such that the distance between the hyperplane and the support vectors (**margin**) is maximum.



Cont..



- But what if I draw the hyperplane like this? The margin for this hyperplane is clearly more than the previous one. So, this is our optimal hyperplane.

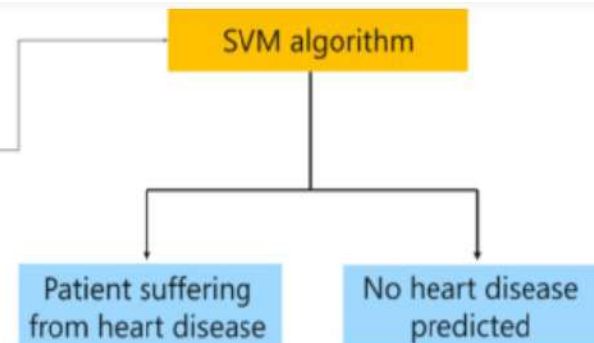
Cont..

- But what will you do if the data set is like this?
- This is where Non-linear SVM is implemented.



- Application of SVM

	A	B	C	D	E	F	G	H	I
1	63	1	1	145	233	1	2	150	0
2	67	1	4	160	286	0	2	108	1
3	67	1	4	120	229	0	2	129	1
4	37	1	3	130	250	0	0	187	0
5	41	0	2	130	204	0	2	172	0
6	56	1	2	120	236	0	0	178	0
7	62	0	4	140	268	0	2	160	0
8	57	0	4	120	354	0	0	163	1
9	63	1	4	130	254	0	2	147	0
10	53	1	4	140	203	1	2	155	1
11	57	1	4	140	192	0	0	148	0
12	56	0	2	140	294	0	2	153	0
13	56	1	3	130	256	1	2	142	1
14	44	1	2	120	263	0	0	173	0
15	52	1	3	172	199	1	0	162	0

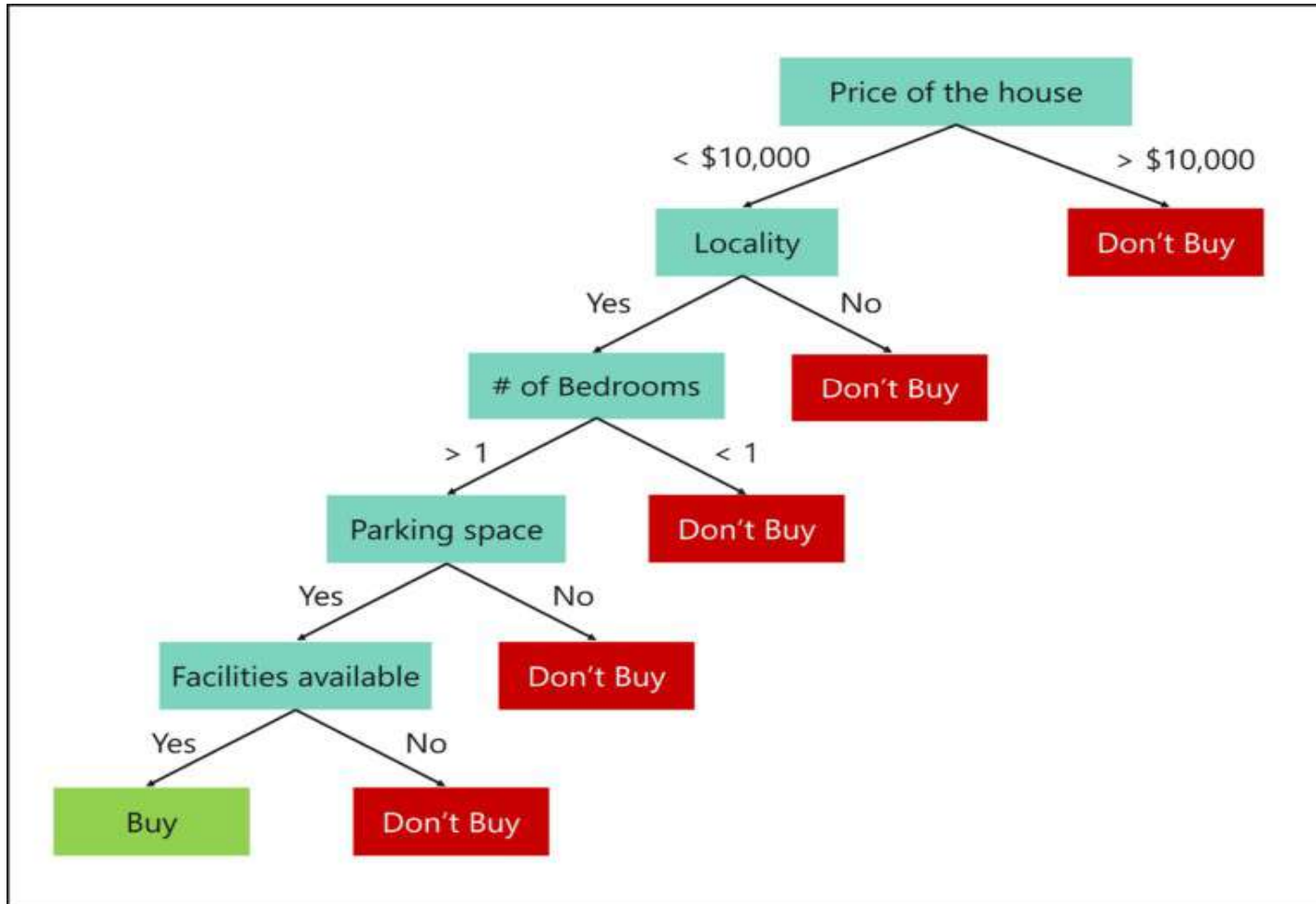


Random Forest

- Supervised classification and regression algorithm
- As the name suggests, this algorithm randomly creates a forest with several trees.
- Decision trees are built on the entire data set using all the predictor variables, whereas Random Forests are used to create multiple decision trees, such that each decision tree is built only on a part of the data set.
- **Example:** Looking to buy a house, consult a few agents and they give you a list of parameters:
 - Price of the house
 - Locality
 - Number of bedrooms
 - Parking space
 - Available facilities
- These parameters are known as *predictor variables*, which are used to find the response variable.

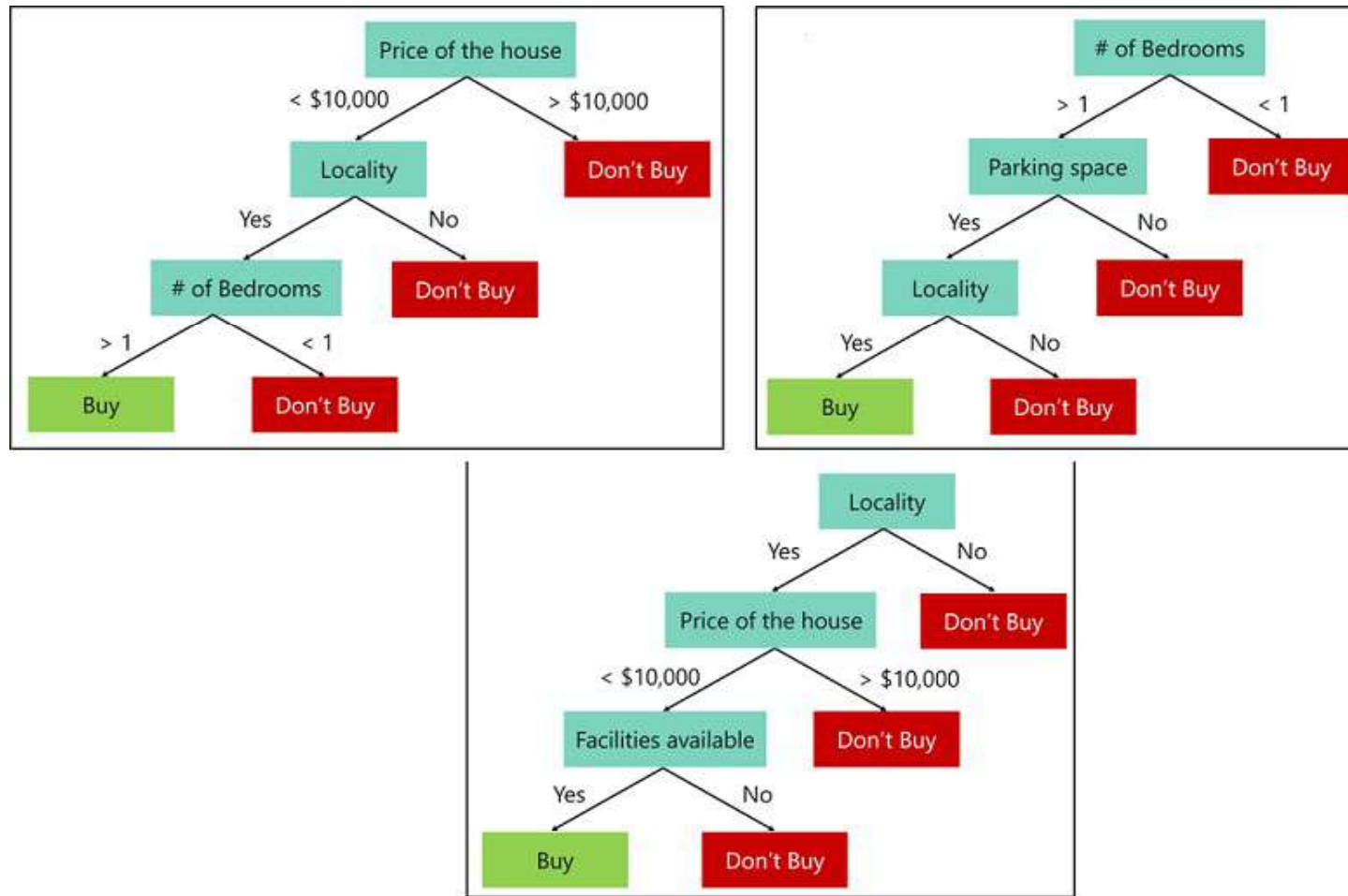
Cont..

- By using **Decision Tree**:
Decision trees are built on the entire data set, by making use of all the predictor variables.



Cont..

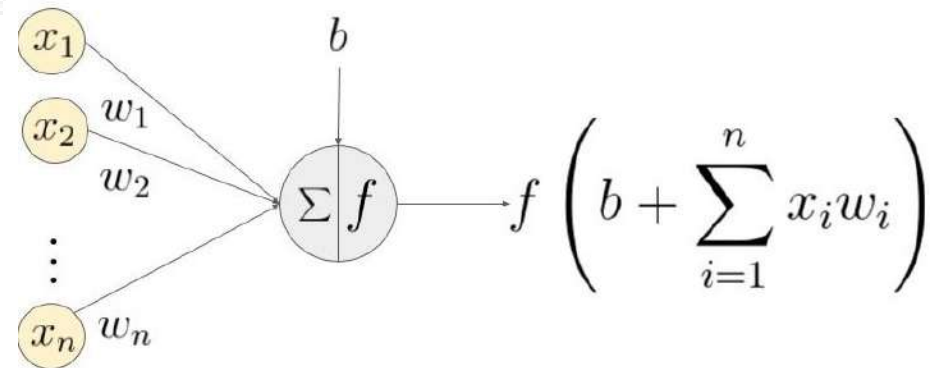
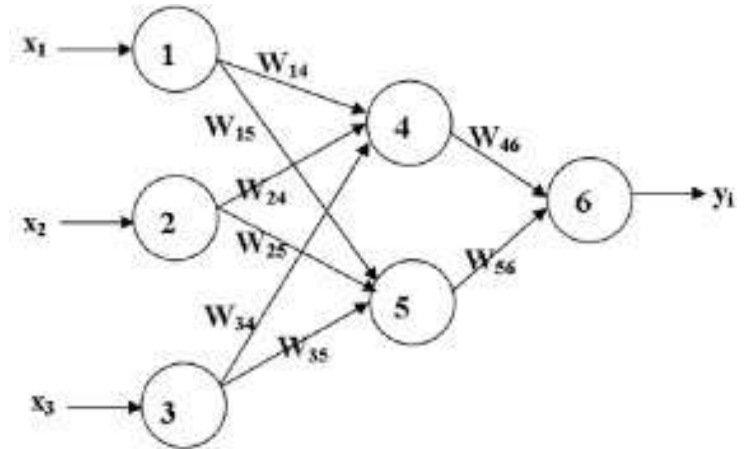
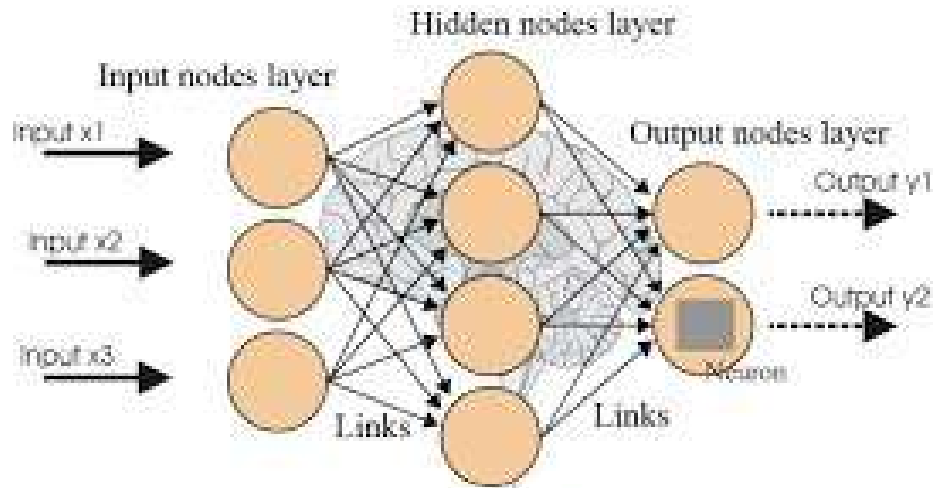
- By using **Random Forest**:
In simple words, after creating multiple Decision trees using this method, each tree selects or votes the class, and the class receiving the *most votes by a simple majority* is termed as the predicted class.



Why use Random Forest?

- Even though Decision trees are convenient and easily implemented, they lack accuracy.
- Decision trees work very effectively with the training data that was used to build them, but they're not flexible when it comes to classifying the new sample.
- Which means that the accuracy during testing phase is very low.
- This happens due to a process called Over-fitting.
- **Over-fitting** occurs when a model studies the training data to such an extent that it **negatively** influences the performance of the model on new data
- This means that the disturbance in the training data is recorded and learned as concepts by the model.
- But the problem here is that these concepts do not apply to the testing data and negatively impact the model's ability to classify the new data, hence **reducing the accuracy** on the testing data. ***This is where Random Forest comes in.***
- (Bagging)- Reduce the variation in the predictions by combining the result of multiple Decision trees on different samples of the data set.

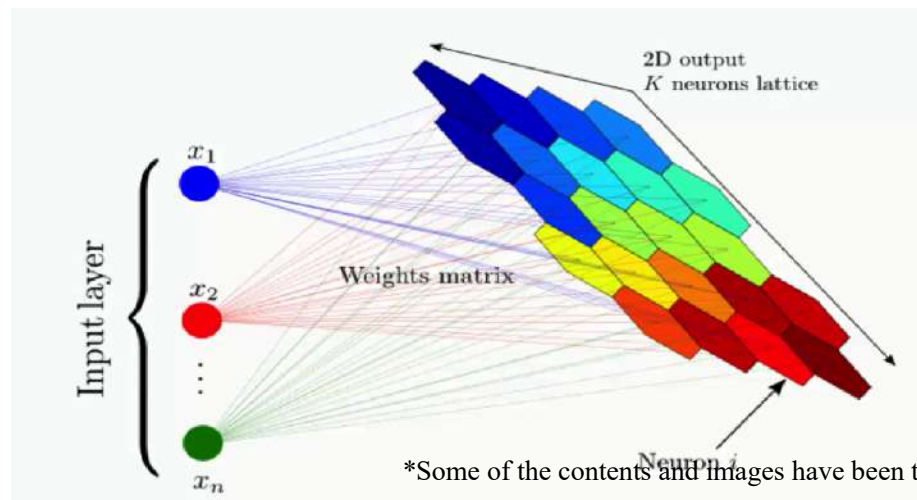
Artificial Neural Network (ANN)



A layer in a neural network without a bias is nothing more than the multiplication of an input vector with a matrix.

Self Organization Map (SOM)

- Self Organizing Map(SOM) by **Teuvo Kohonen** in 1980s provides a data visualization technique that is trained using [unsupervised learning](#) and helps to understand high dimensional data by [reducing the dimensions](#) of data to a map.
- SOM **reduces data dimensions** and **displays similarities among data**.
- Unlike other learning technique in neural networks, training a SOM requires no target vector.
- A SOM learns to classify the training data without any external supervision.
- Self-organizing maps differ from other artificial neural networks as they apply [competitive learning](#) as opposed to error-correction learning (such as back-propagation with gradient descent).



*Some of the contents and images have been taken from google server or AI tools

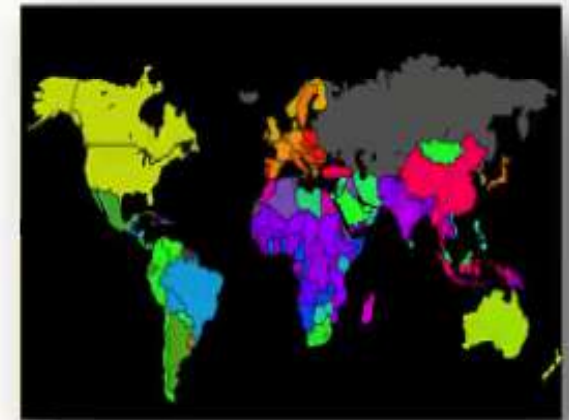
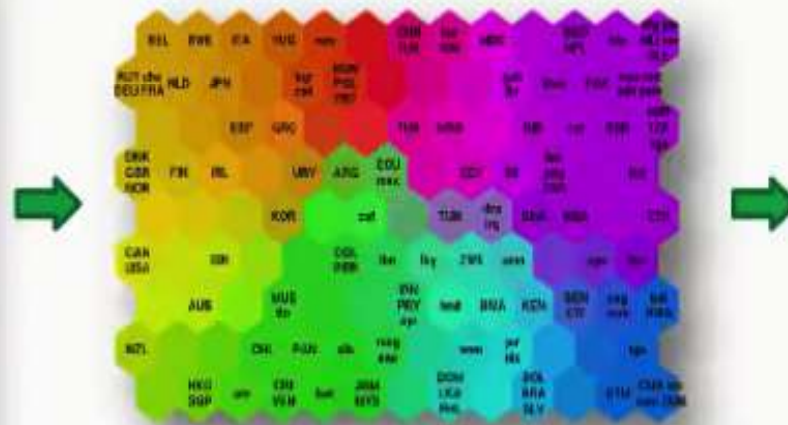
Cont..

- Complex data set consisting of a massive amount of columns and dimensions and demonstrates how that data set's dimensionality can be reduced.
- So, instead of having to deal with hundreds of rows and columns, the data is processed into a simplified map (**two-dimensional representation easier to read and analyse**).
- In following case, a data set from the World Bank containing all 39 indicators of human development, processed them into a SOM, and then applied that to a world map.
- You can imagine this data set of 200+ countries to have started off with 39 different columns. **It's tough to imagine such a bulk of data, isn't it?**
- **That's exactly why we use SOMs.**
- Instead of presenting you with each indicator separately, the SOM lines up the countries inside this cluster to make up a spectrum of wellbeing.

Cont..

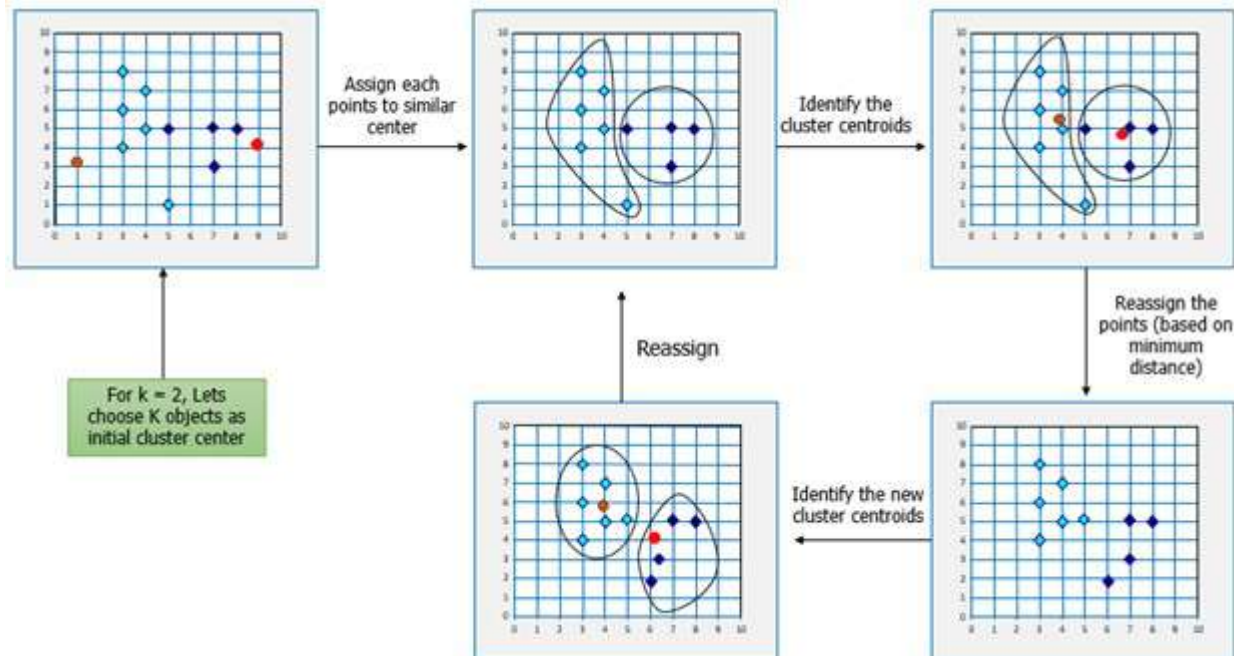
- As you can see, the countries are lined up in a cluster, and their order inside that cluster is based on various indicators (e.g. **health conditions, quality of education, average income per capita, etc.**). -spectrum of wellbeing

	A	B	C	D	E
1	Country	Country C	Health Ex	Education E	Inflation
2	Aruba	ABW	9.418971	5.92467022	-2.13637
3	Afghanistan	AFG	4.371774		-8.28308
4	Angola	AGO	5.791339		13.73145
5	Albania	ALB	6.75969		2.280502
6	Andorra	AND	4.57058	3.1638701	
7	Arab Wor	ARB	4.049924		3.524834
8	United Ar	ARE	7.634758		
9	Argentina	ARG	4.545323	4.88997984	6.282774
10	Armenia	ARM		3.84079003	3.406767
11	American	ASM	4.862062		
12	Antigua a	ATG	9.046056	2.55447006	-0.55016
13	Australia	AUS	11.19444	5.09262991	1.820112
14	Austria	AUT	5.85024	5.7674098	0.306313
15	Azerbaija	AZE	6.964187	3.22430992	1.401056
16	Burundi	BDI	10.39434	6.3197999	10.98147
17	Belgium	BEL	4.46431	6.41335997	-0.05315
18	Benin	BEN	7.405431	4.22204018	2.15683



K-Means Clustering

- K-means ([Macqueen, 1967](#)) is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem.
- A cluster refers to a small group of objects. Clustering is grouping those objects into clusters.
- Clustering is dividing data points into homogeneous classes or clusters:
 - Points in the same group are as similar as possible
 - Points in different group are as dissimilar as possible



*Some of the contents and images have been taken from google server or AI tools

Cont..

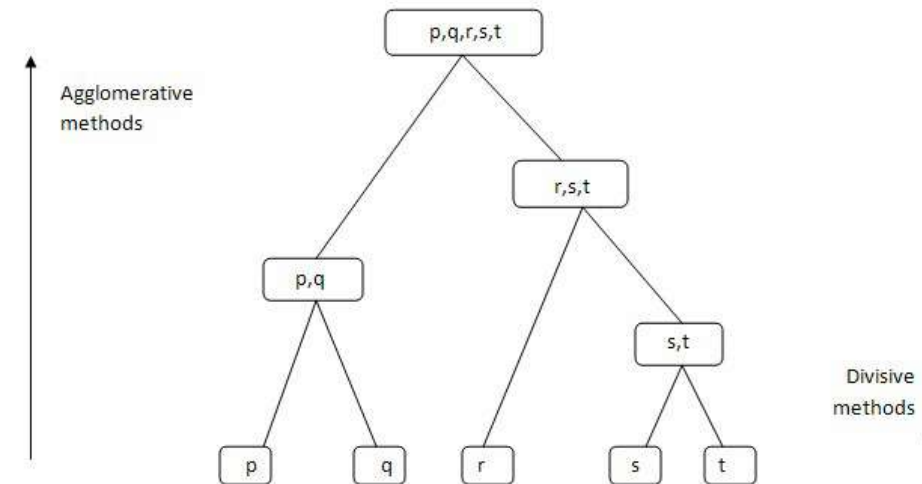
- Clustering helps marketers improve their customer base and work on the target areas.
- Clustering helps in identification of groups of houses on the basis of their value, type and geographical locations
- **Example:**

A pizza chain wants to open its delivery centres across a city. What do you think would be the possible challenges?

 - They need to analyse the areas from where the pizza is being ordered frequently.
 - They need to understand as to how many pizza stores has to be opened to cover delivery in the area.
 - They need to figure out the locations for the pizza stores within all these areas in order to keep the distance between the store and delivery points minimum.

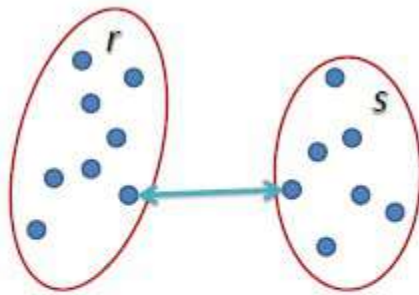
Hierarchical Clustering

- It involves creating clusters that have a predetermined ordering from top to bottom. For example, all files and folders on the hard disk are organized in a **hierarchy**.
- There are **two types** of hierarchical clustering, **Divisive** and **Agglomerative**.
- In **divisive or top-down clustering** method we assign all of the observations to a single **cluster** and then partition the **cluster** to **two** least similar clusters.



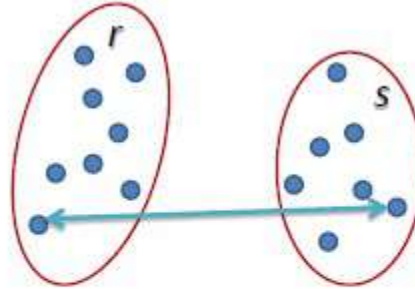
Cont..

- In **agglomerative** or *bottom-up clustering* method we assign each observation to its own cluster. Then, compute the similarity (e.g., distance) between each of the clusters and join the two most similar clusters (least distance). Finally, repeat this until there is only a single cluster left.
- Methods to measure distance: **single linkage, complete linkage, average linkage**



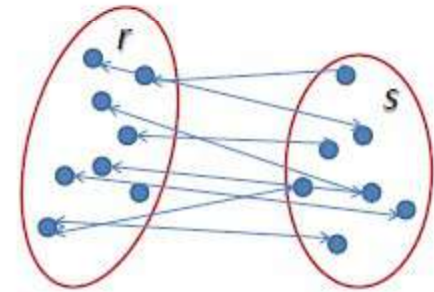
$$L(r, s) = \min(D(x_{ri}, x_{sj}))$$

single linkage



$$L(r, s) = \max(D(x_{ri}, x_{sj}))$$

complete linkage



$$L(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

average linkage

Ethics of Employing ML in Healthcare (morally correct or acceptable)

1) Privacy and Data Security

- HIPAA and other privacy regulations ensure the security of the patient's information. Everybody should have a right to keep information about their health private.
- Nevertheless, a lot of healthcare data leaks are happening every day that result in up to **\$16 million penalties** for healthcare providers.
- However, data is the **blood of the ML** organism. How can these points effectively coexist?
- ML doesn't require a full spectrum of information on the patient (like name, email, phone number, and insurance policy number); thus, it can be effectively anonymized so that the **person's identity cannot be revealed**, while the precision of the ML-algorithm won't be discounted.
- For others, special data security approaches have to be implemented to ensure patient anonymity.

2) Autonomy Issues:

- ML can be effectively used to help the elderly and people with psychological issues make decisions to improve their health.
- **This concerns taking the right medications, creating healthy habits, and referring to the specialist whenever needed.**

Cont..

- However, the ethical issue behind this is that people will potentially give up their **autonomy** and act as they are told.
- It limits their potential choice range to certain recommended options.
- So, a **clear balance** between the instructions from the algorithm and freedom of personal choice should be provided.

3) Patients' Safety:

- The decisions made by the machine learning algorithm completely rely on the data it has been learned on.
- If **the input is unreliable or wrong**, the result will be wrong as well. The flawed decision can harm the patient or even cause their death.
- The ethical dilemma here is who would be responsible for the death of a patient because of the decision made by an algorithm? Right now, this remains an open question.
- The final decision on the treatment method is behind the patient, who should be informed about all the benefits and risks of each treatment method.

4) Transparency and Informed Consent

- ML algorithms rely on data. The more relevant data is available, the better they work, and the more precise results and predictions could be achieved.

Cont..

- Many countries have legislation that restricts the use of patient data without their informed consent.
- So, the use of ML in healthcare should go hand-in-hand with informing the patient on it and on the data security efforts applied to keep their data safe.

5) Representation and Inclusivity

- When developing a comprehensive healthcare software solution you should assure its algorithms work effectively on a bright variety of patients.
- According to PMC, different ethnic and racial groups could have varying responses to pharmaceutical treatment and require special care.
- Thus, a ML solution should be “learned” on sufficiently wide and diverse patient cases and backgrounds. Moreover, it’s recommended to prevent and warn physicians about instances the ML algorithm might lack of research data and probably drive less accurate results, therefore.

Challenges of ML in the Healthcare Industry

1) Lack of Quality Data to Build Precise Algorithms

- The results you get from ML algorithms depend on the quality of data put into them.
- Unfortunately, medical data is not always as precise and standardized as it often needs to be. There are gaps in records, inaccuracies in profiles, and other difficulties.
- Overall, electronic health records (EHRs) were not built to be used as a data source for an algorithm.
- So, before apply for a ML tool, we need to **spend time gathering, cleaning, validating, and structuring** data for its purposes.

2) Building ML Tools Friendly to Medical Workflow

- There are multiple highly specific ML use cases that can help with patient diagnostics and treatment.
- Even if an ML tool works well on paper, it does not necessarily mean that it will be adopted by physicians. Thus, it's critical to develop and roll-out ML tools that would be intuitive and easy to use in the everyday medical workflow.
- Without the necessary feedback from people who will work with the tool, it will not be as effective, and the professionals will not trust it.

Cont..

3) Gather Big Teams with Broad Skill Sets in One Place

- Besides hands-on healthcare specialists, an effective ML development team should include such roles:
 - Business analyst
 - Data architect
 - Data engineer
 - Data scientist
 - Machine learning expert.
- Besides, it's important to facilitate effective cooperation proceedings in the team so that it's possible to deliver value and prove the product's viability at the earliest opportunity.

Top companies using ML in healthcare

1) Viz.ai

- The Viz.ai tool helps care teams to optimize care coordination and improve communication between healthcare professionals with machine learning algorithms.
- Viz.ai streamlines care by quickly connecting frontline healthcare providers to specialists, resulting in faster case resolution and improved patient outcomes.

2) Deep Genomics

- The Deep Genomics' AI-powered platform accelerates research by helping healthcare professionals quickly find candidates for the development of drugs for specific disorders.

3) Oncora Medical

- Oncora Medical is a Philadelphia-based startup that streamlines cancer research and treatment.
- By collecting large amounts of medical data, Oncora's platform can assess care quality and suggest better treatment methods.

4) Intuitive Surgical

- Intuitive Surgical are the developers of the most widely used ML-powered surgical system called Da Vinci.
- Da Vinci Surgical System allows surgeons to perform robotic-assisted, minimally invasive surgeries that significantly improve surgery outcome.

ML based Biomedical Research Publications and Patent done by our Group

- “Analysis of ECG signal by chaos principle to help Automatic Diagnosis of Myocardial Infarction”, Journal of Scientific & Industrial Research (JSIR), Vol. 68, pp. 866-870, 2009. ISSN: 2320 – 4818, (SCI).
- "Significant Enhancement of Segmentation Efficiency of Retinal Images Using Texture-Based Gabor Filter Approach Followed by Optimization Algorithm", International Journal of Computer Vision and Image Processing, Volume 7 (1), pp. 44-58, January-March 2017, ISSN: 2155-6997, DOI: 10.4018/IJCVIP.2017010103. (UGC CARE)
- “A Noval Hybrid Method for Segmentation and Analysis of Brain MRI for Tumor Diagnosis”, Advances in Science, Technology and Engineering Systems Journal, Volume 5, Issue 3, Page No 16-27, ISSN: 2415-6698, 2020. (Scopus)
- “A Novel Approach to Brain Tumor Detection using Texture Based Gabor Filter Followed by Genetic Algorithm”, International Journal on Communications Antenna and Propagation (IRECAP), Vol. 11 (4), 2021, pp. 233-241, ISSN 2039-5086. DOI: <https://doi.org/10.15866/irecap.v11i4.20766>. (Scopus).
- “Image splicing detection system using intensity-level multi-fractal dimension feature engineering and twin support vector machine based classifier” (10.1007/s11042-022-13519-2), Multimedia Tools and Applications, ISSN: 13807501, published on 13 July, 2022. IF-2.57 (SCI)

Cont..

- “A Paradigm Shift in Diagnosis of Mental Health Problems Towards Integrated Methods of Soft Computing”, Journal of Optoelectronics Laser, Vol. 41 (8), pp. 126-130, ISSN: 1005-0086, 2022. (Scopus)
- “Enhancing Feature Selection Accuracy Using Butterfly and Lion Optimization Algorithm With Specific Reference to Psychiatric Disorder Detection & Diagnosis”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 11, Nov. 2022, pp. 107-14, ISSN: 2321-8169, doi:10.17762/ijritcc.v10i11.5786. (Scopus)
- “An Efficient Approach for Automated System to Identify the Rice Crop Disease Using Intensity Level based Multi-fractal Dimension and Twin Support Vector Machine”, Archives of Phytopathology and Plant Protection, Vol. 56 (10), pp. 806-834, 2023. (Scopus and ESCI).
- Analysis of Meta-Heuristic Feature Selection Techniques on classifier performance with specific reference to psychiatric disorder. International Journal of Experimental Research and Review, Vol. 31 (Spl Volume), 51-60. ISSN: 2455-4855 (Scopus).
- “An Integrated Method using Neuro-fuzzy classifier and Intelligence Feature Selection Methods for Diagnosis for Psychiatric Disorder”, Soft Computing, Springer, IF- 3.732, Accepted on 04 August, 2023. (SCIE)

Books and Book Chapters

- 1)"Applications of Machine Learning in Diseases Pre-screening", Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention, ISSN: 2475-6628, IGI Global Publisher, Hershey PA, USA, Published on 07/03/2018, Advances in Medical Diagnosis, Treatment, and Care (AMDTC) (ISSN: 2475-6628; EISSN: 2475-6636) (Scopus)
- 2)"Deep learning for healthcare biometrics", book title: Design and Implementation of Healthcare Biometric Systems, Publisher: IGI Global Publisher, Hershey PA, USA, Approved on 30/08/2018, Advances in Medical Technologies and Clinical Practice (AMTCP), (ISSN: 2327-9354; EISSN: 2327-9370)(Scopus)
- 3)"Artificial Intelligence and Biotechnology: The golden age of Medical Research", book title: Biotechnology in the Modern Medicinal System: Advances in Gene Therapy, Immunotherapy, and Targeted Drug Delivery, Publisher: Apple Academic Press Inc. by CRC Press, a Taylor & Francis Group, Hard ISBN: 9781771889728, Oct 2021. (Scopus)
- 4)"Application of Machine Learning to Analyze Biomedical Signals for Medical Diagnosis", Handbook of Research on Disease Prediction through Data Analytics and Machine Learning, Publisher: IGI Global Publisher, Hershey PA, USA, ISBN13: 9781799827429 | ISBN10: 1799827429, Oct 2020. (Scopus)
- Book-1: "Advanced Research in Electronic Devices for Biomedical and mHealth", Hard ISBN: 9781774915165, Apple Academic Press, Inc.,CRC Press, Taylor & Francis Group, 2023.
- Book-2: "Smart Electronic Devices and Systems for Biomedical and Healthcare Applications", Apple Academic Press, Inc.,CRC Press, Taylor & Francis Group, 2023 (Under Process). (Scopus)

Conclusion

- The future of ML in the healthcare sector looks bright.
- Despite some challenges, ML is already enhancing the patient experience, medicine practice of clinicians, and the pharmaceutical industry operations.
- And the journey has just begun. According to **Grand View Research**, the global AI and ML in the healthcare market are expected to expand at a compound annual growth rate (CAGR) of 38.4% from 2022 to 2030.
- The **growing datasets of patient health-related** digital information, **increasing demand for personalized medicine**, and the **rising demand for reducing care expenses** are some of the major driving forces of the market growth.
- Additionally, in the coming years, there could be programmed robots that would assist doctors in the surgery room.
- ML-powered technologies in healthcare can enable doctors to minimize the risk during operations by going down to the minutest details of the treatment.

Cont..

- ML can also help healthcare industry by enabling “virtual biopsies” (non-invasive approach) and advancing the innovative field of radiomics.
- While machine learning has a lot of potential in the healthcare industry, it also comes with certain challenges, like healthcare data quality, building physician-friendly products, and gathering a huge team of data experts.
- There also are certain ethical concerns, including patient safety and accountability.

References

1. Kong HJ. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res. 2019 Jan;25(1):1-2.Epub 2019 Jan 31. PMID: 30788175; PMCID: PMC6372467.
2. Gose, E., Johnsonbaugh, R. & Jost, S. (1996). Pattern recognition and image analysis. Prentice Hall.
3. Tom Mitchell, Machine Learning, McGraw Hill, 1997.
4. Sachi Nandan Mohanty, G. Nalinipriya, Om Prakash Jena, Achyuth Sarkar, "Machine Learning for Healthcare Applications", 2021, Print ISBN:9781119791812, DOI:10.1002/9781119792611.
5. Bikesh Kumar Singh, G.R. Sinha, "Machine Learning in Healthcare Fundamentals and Recent Applications", 2022 by CRC Press, ISBN 9780367564421.
6. Vishal Jain and Jyotir Moy Chatterjee, "Machine Learning In Biomedical Science And Healthcare Informatics", Scrivener Publishing LLC.
7. Bharat Bhushan, Om Prakash Jena, Utku Kose, "Machine Learning and Deep Learning in Medical Data Analytics and Healthcare Applications (Biomedical and Robotics Healthcare)", 25 February 2022.
8. Habehh H, Gohel S., "Machine Learning in Healthcare". Curr Genomics. 2021 Dec 16;22(4):291-300. doi: 10.2174/1389202922666210705124359. PMID: 35273459; PMCID: PMC8822225.
9. Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, "Significance of machine learning in healthcare: Features, pillars and applications", International Journal of Intelligent Networks, Volume 3, 2022, Pages 58-73, ISSN 2666-6030.
10. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019 Oct;1(6):e271-e297. Epub 2019 Sep 25. Erratum in: Lancet Digit Health. 2019 Nov;1(7):e334. PMID: 33323251.
11. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. Radiographics. 2017 Mar-Apr;37(2):505-515. doi: 10.1148/rq.2017160130. Epub 2017 Feb 17. PMID: 28212054; PMCID: PMC5375621



Amazon.in · In stock
Deep Learning in M...



Amazon.in · In stock
Healthcare Applicat...



Amazon.in · In stock
Deep Learning in Medical ...



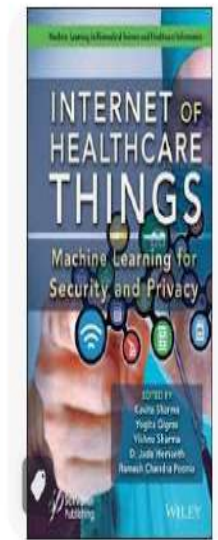
Amazon.in · In stock
Buy MACHINE LEA...



Amazon.in · In stock
Explainable AI in He...



Routledge · In stock
Machine Learning i...



Amazon.in · In stock
Buy Internet of Heal...

