AI Series for Teachers

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Dr. Amit Bhagat
Associate Professor
Deptt. of Mathematics, Bioinformatics & Computer Applications,
Maulana Azad National Institute of Technology, Bhopal
amitbhagat@manit.ac.in

AI Series for Teachers

Artificial Intelligence

“Intelligence" refers to the ability to perform tasks that typically require cognitive abilities, such as learning from experience, understanding language, recognizing patterns, making decisions and solving problems.

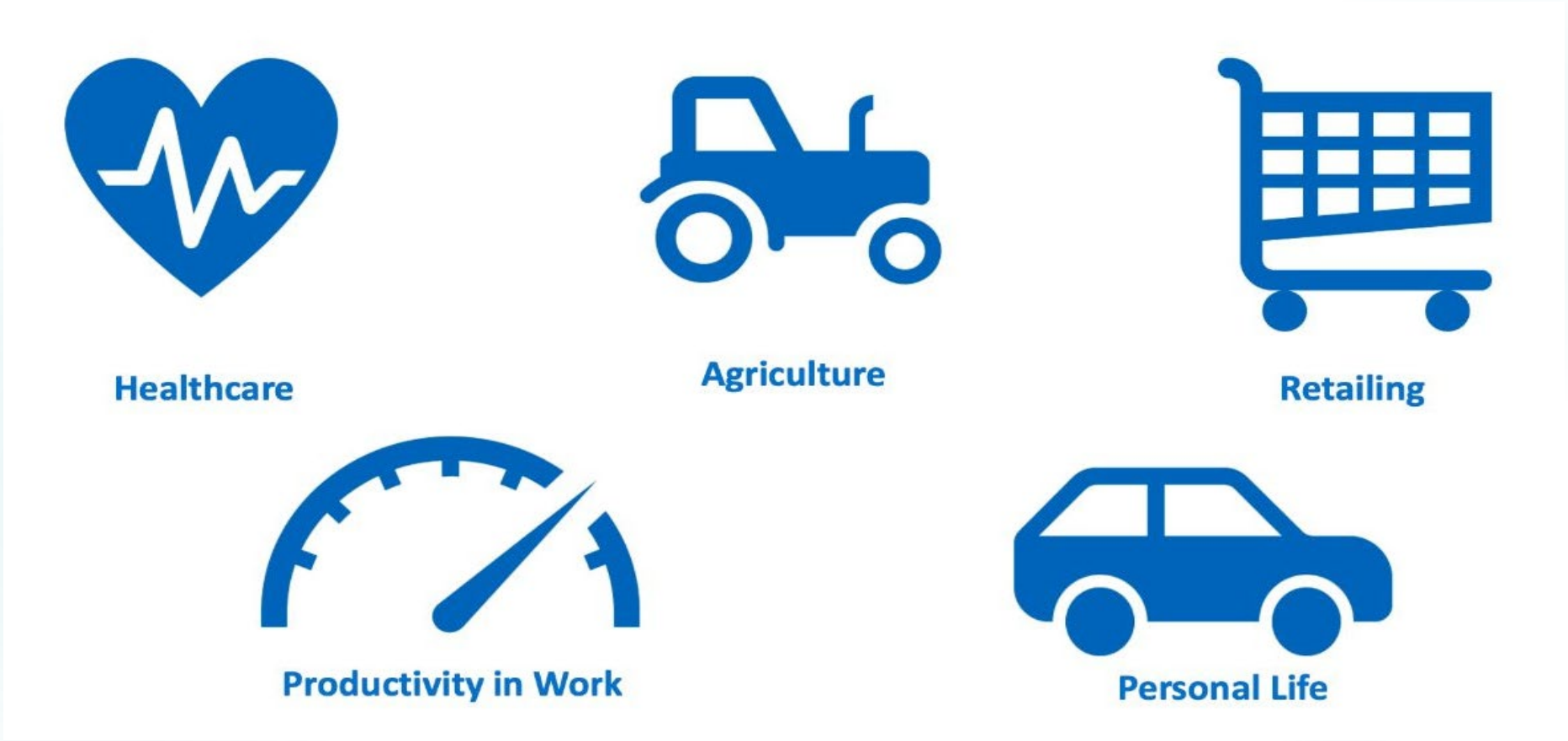
AI Series for Teachers

Artificial Intelligence

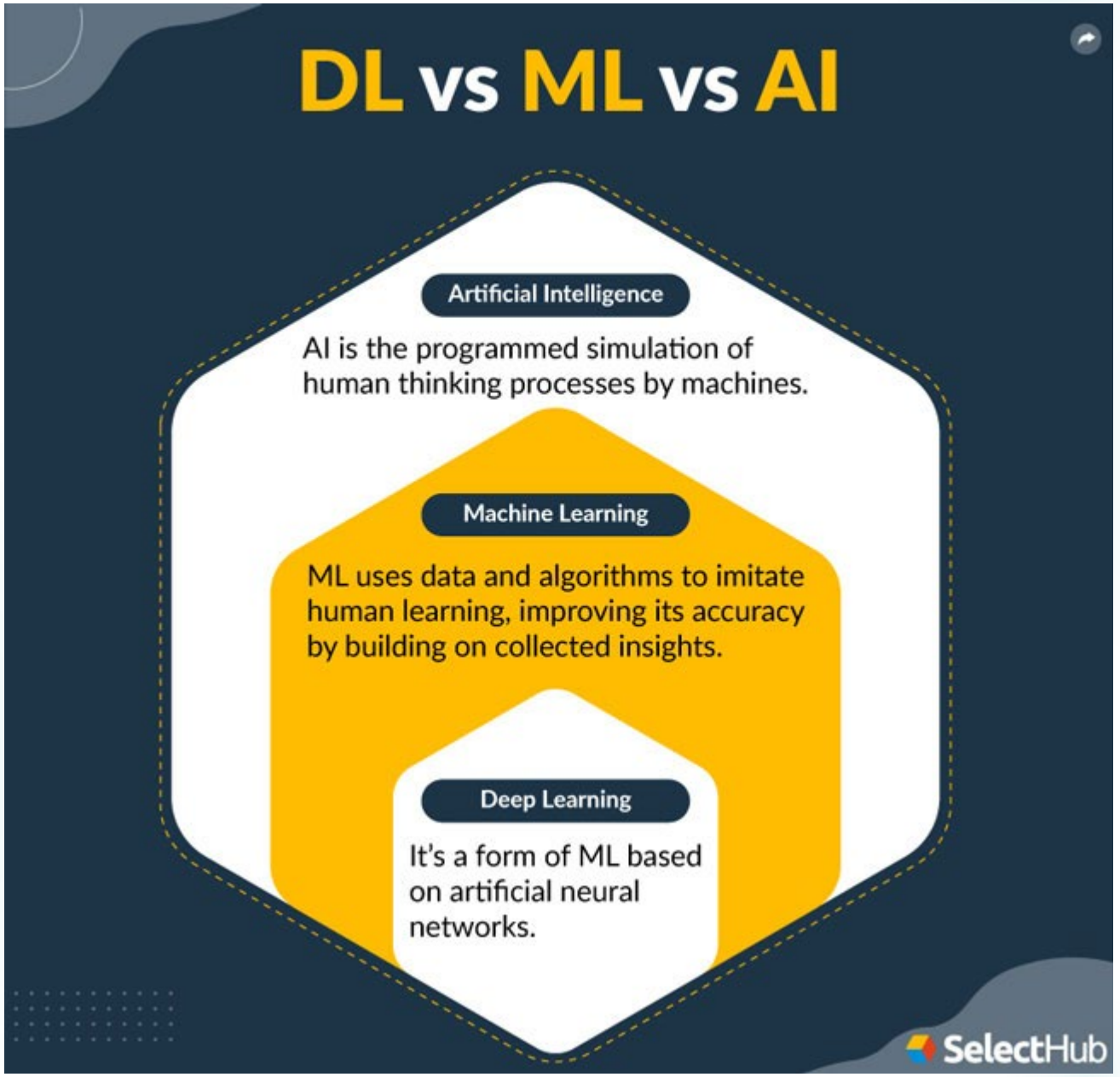
Artificial Intelligence (AI) refers to the development of computer systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, decision-making, and perception. AI encompasses a wide range of technologies, including machine learning, deep learning, and natural language processing (NLP).

AI Series for Teachers

Applications of AI in real life



AI Series for Teachers



https://www.selecthub.com/

Artificial Intelligence Vs Machine Learning

	Artificial Intelligence	Machine Learning
Definition	Artificial intelligence is human intelligence exhibited by machines	It is an approach to achieve AI
Subset	Not the subset of a machine or deep learning	Subset of Artificial intelligence
Programming	Requires full programming thing to make the system	Does not require any hard code algorithms
Complex	More complex as one has to know everything	Less complex than AI

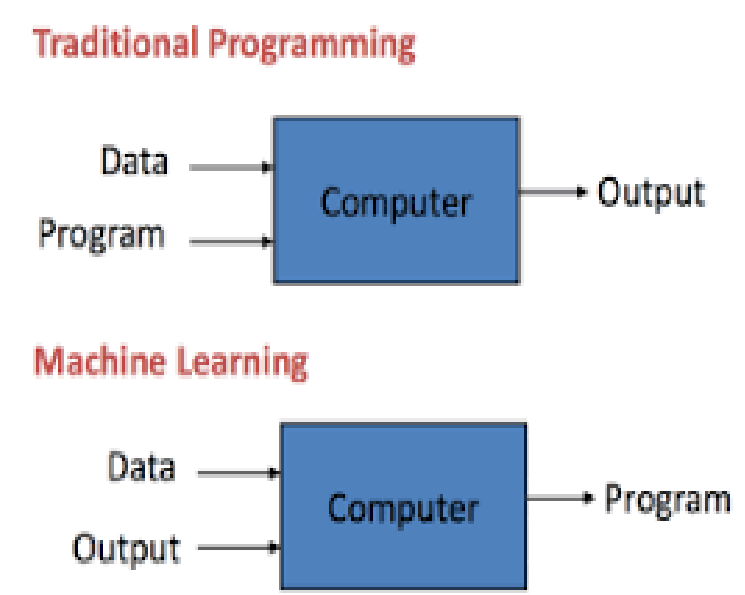
AI Series for Teachers

Machine Learning

Arthur Samuel, a pioneer in the field of artificial intelligence and computer gaming, coined the term **"Machine Learning"** as – **"Field of study that gives computers the capability to learn without being explicitly programmed"**.

How it is different from traditional Programming:

- In Traditional Programming, we feed the Input, Program logic and run the program to get output.
- In Machine Learning, we feed the input, output and run it on machine during training and the machine creates its own logic, which is being evaluated while testing.



AI Series for Teachers

ML Definition

Tom Mitchell's well-known definition of machine learning states:
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".

In simpler terms, Machine Learning is a technique that enables computers to **learn from data** and **improve their performance** on a **specific task**. It involves teaching machines to identify patterns and relationships in data that they can use to make decisions or predictions. The process of Machine Learning involves three key elements: **the model, the learning algorithm, and the data.**

AI Series for Teachers

Terminologies that one should know before starting Machine Learning:

- **Model:** A model is a **specific representation** learned from data by applying some machine learning algorithm. A model is also called **hypothesis**.
- **Feature:** A feature is an individual measurable property of our data. A set of numeric features can be conveniently described by a **feature vector**. Feature vectors are fed as input to the model. For example, in order to predict a fruit, there may be features like color, smell, taste, **etc.**
- **Target(Label):** A **target variable** or label is the **value** to be **predicted** by our model. For the fruit example discussed in the features section, the label with each set of input would be the name of the fruit like apple, orange, banana, etc.
- **Training:** The idea is to give a **set of inputs(features)** and it's **expected outputs(labels),** so after training, we will have a model (hypothesis) that will then map new data to one of the categories trained on.
- **Testing(Prediction):** Once our model is ready, it can be fed a **set of inputs** to which it will provide a **predicted output(label).**

AI Series for Teachers

Types of Learning

- **Supervised Learning**
- **Unsupervised Learning**
- **Semi-Supervised Learning**
- **Reinforcement Learning**

1. **Supervised Learning**: Supervised learning is when the model is getting trained on a labelled dataset. **Labelled** dataset is one which **have both input and output parameters**. In this type of learning both training and validation datasets are labelled as shown in the figures below.

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0

Classification

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867544	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680246	988.9386597	53.74138803	68.15406036	1.650191426

Regression

AI Series for Teachers

Types of Supervised Learning:

- **Classification**
- **Regression**

Classification : It is a **Supervised Learning** task where output is having defined labels(discrete value). For example in above Figure A, Output – Purchased has defined labels i.e. 0 or 1 ; 1 means the customer will purchase and 0 means that customer won't purchase. It can be either binary or multi class classification.
In **binary** classification, model predicts either 0 or 1 ; yes or no but in case of **multi class** classification, model predicts more than one class.

Example: Gmail classifies mails in more than one classes like social, promotions, updates, offers.

Regression : It is a **Supervised Learning** task where **output** is having **continuous value**. Example in before regression Figure, Output – Wind Speed is not having any discrete value but is continuous in the particular range. The goal here is to predict a value as much closer to actual output value as our model can and then evaluation is done by calculating error value. The smaller the error the greater the accuracy of our regression model.

AI Series for Teachers

Example of Supervised Learning Algorithms:

- ❑ Linear Regression
- ❑ K- Nearest Neighbor
- ❑ Gaussian Naive Bayes
- ❑ Decision Trees
- ❑ Support Vector Machine (SVM)
- ❑ Random Forest

AI Series for Teachers

Linear Regression

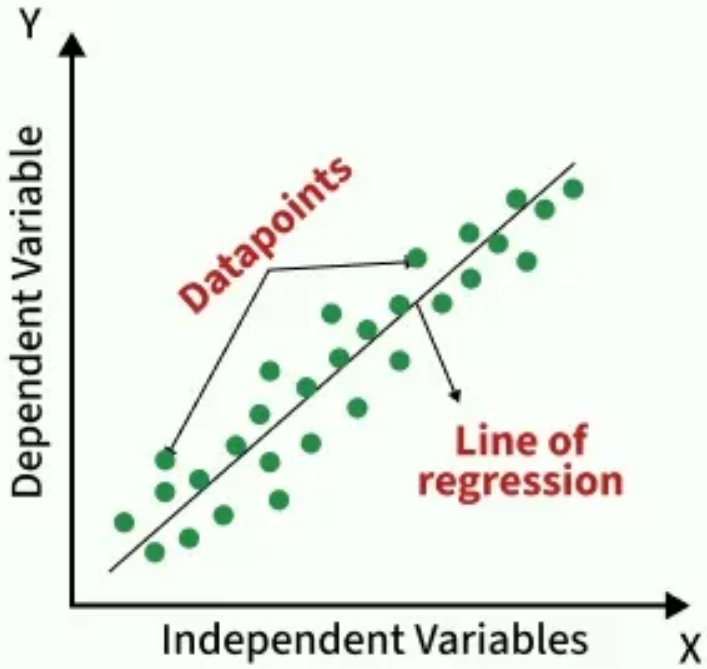
- It assumes that there is a linear relationship between the input and output, meaning the output changes at a constant rate as the input changes. This relationship is represented by a straight line.
- **For example** we want to predict a student's exam score based on how many hours they studied. We observe that as students study more hours, their scores go up. In the example of predicting exam scores based on hours studied. Here
- **Independent variable (input):** Hours studied because it's the factor we control or observe.
- **Dependent variable (output):** Exam score because it depends on how many hours were studied.

AI Series for Teachers

How Does Linear Regression Work?

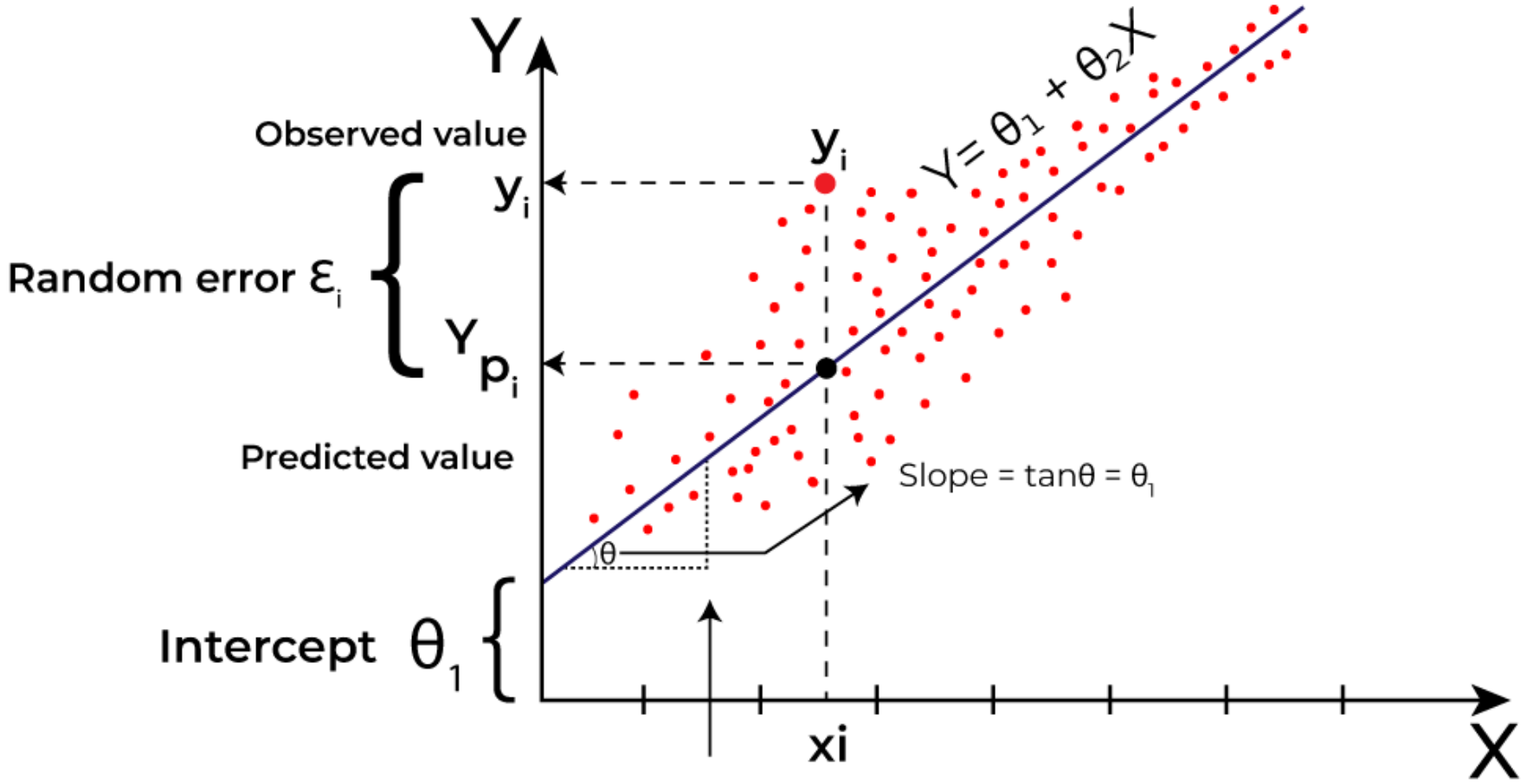
Finds the best-fitting line by minimizing prediction errors (least squares method)

It calculates coefficients for variables that minimize the error in predictions.



AI Series for Teachers

How Linear Regression Wroks



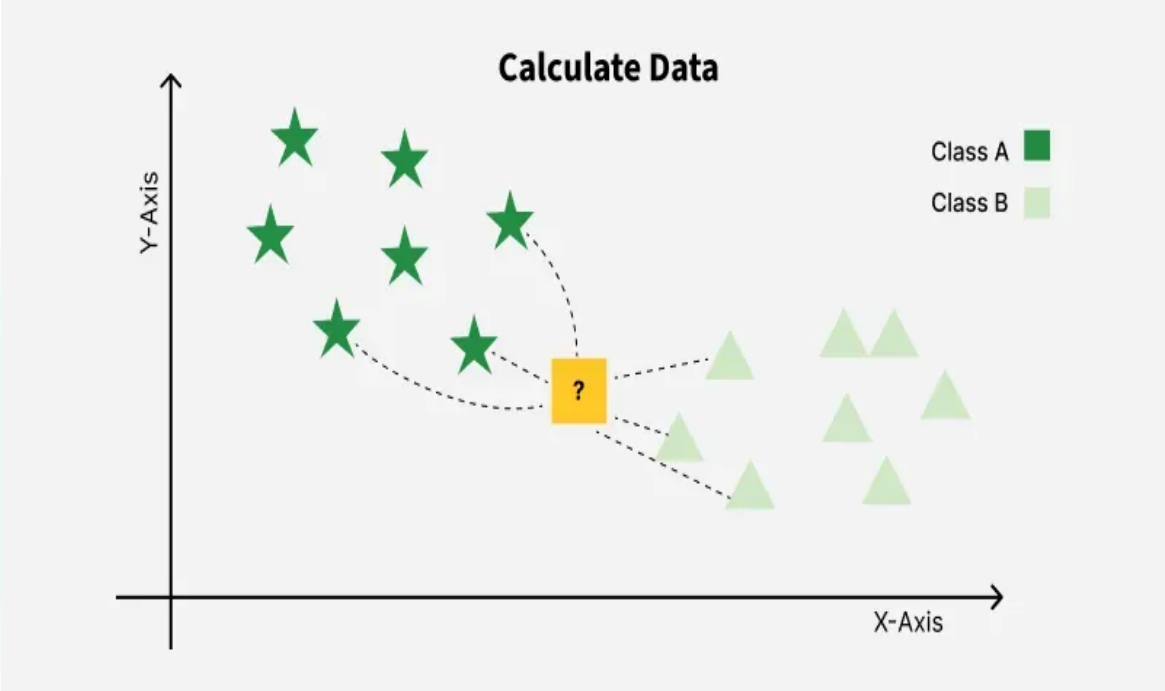
AI Series for Teachers

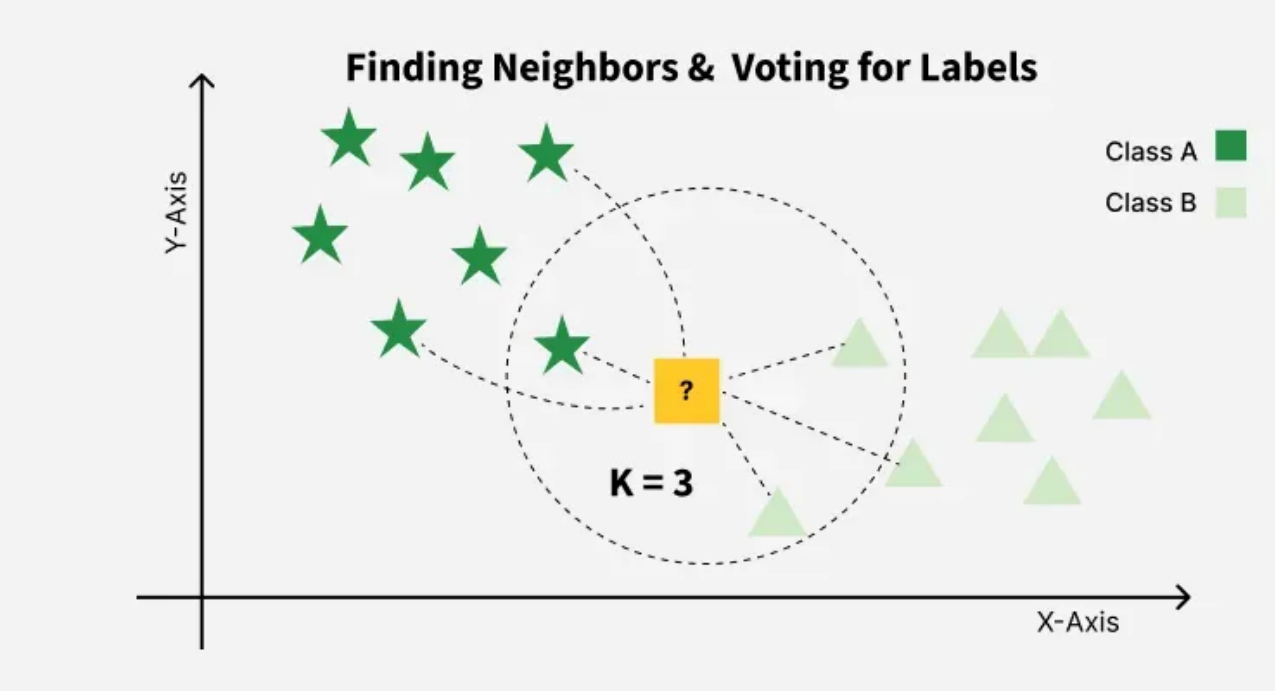
K-Nearest Neighbor(KNN) Algorithm

- K-Nearest Neighbors (KNN) is a supervised machine learning algorithm generally used for classification but can also be used for regression tasks. It works by finding the "k" closest data points (neighbors) to a given input and makes a predictions based on the majority class (for classification) or the average value (for regression).

AI Series for Teachers

K-Nearest Neighbor(KNN) Algorithm





AI Series for Teachers

Naive Bayes Classifiers

- Naive Bayes is a machine learning classification algorithm that predicts the category of a data point using probability. It assumes that all features are independent of each other. Naive Bayes performs well in many real-world applications such as spam filtering, document categorization and sentiment analysis.
- It is named as "Naive" because it assumes the presence of one feature does not affect other features. The "Bayes" part of the name refers to its basis in Bayes' Theorem.

AI Series for Teachers

Naive Bayes Classifiers

Dataset that describes the weather conditions for playing a game of Tennis

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

AI Series for Teachers

Naive Bayes Classifiers

Outlook

	Yes	No	P(yes)	P(no)
Sunny	3	2	3/10	2/4
Overcast	4	0	4/10	0/4
Rainy	3	2	3/10	2/4
Total	10	4	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

AI Series for Teachers

Naive Bayes Classifiers

Humidity

	Yes	No	P(yes)	P(no)
Hot	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

AI Series for Teachers

Naive Bayes Classifiers

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

P(Yes | today)>P(No | today)
The model predicts: ***Yes (Play Tennis)***

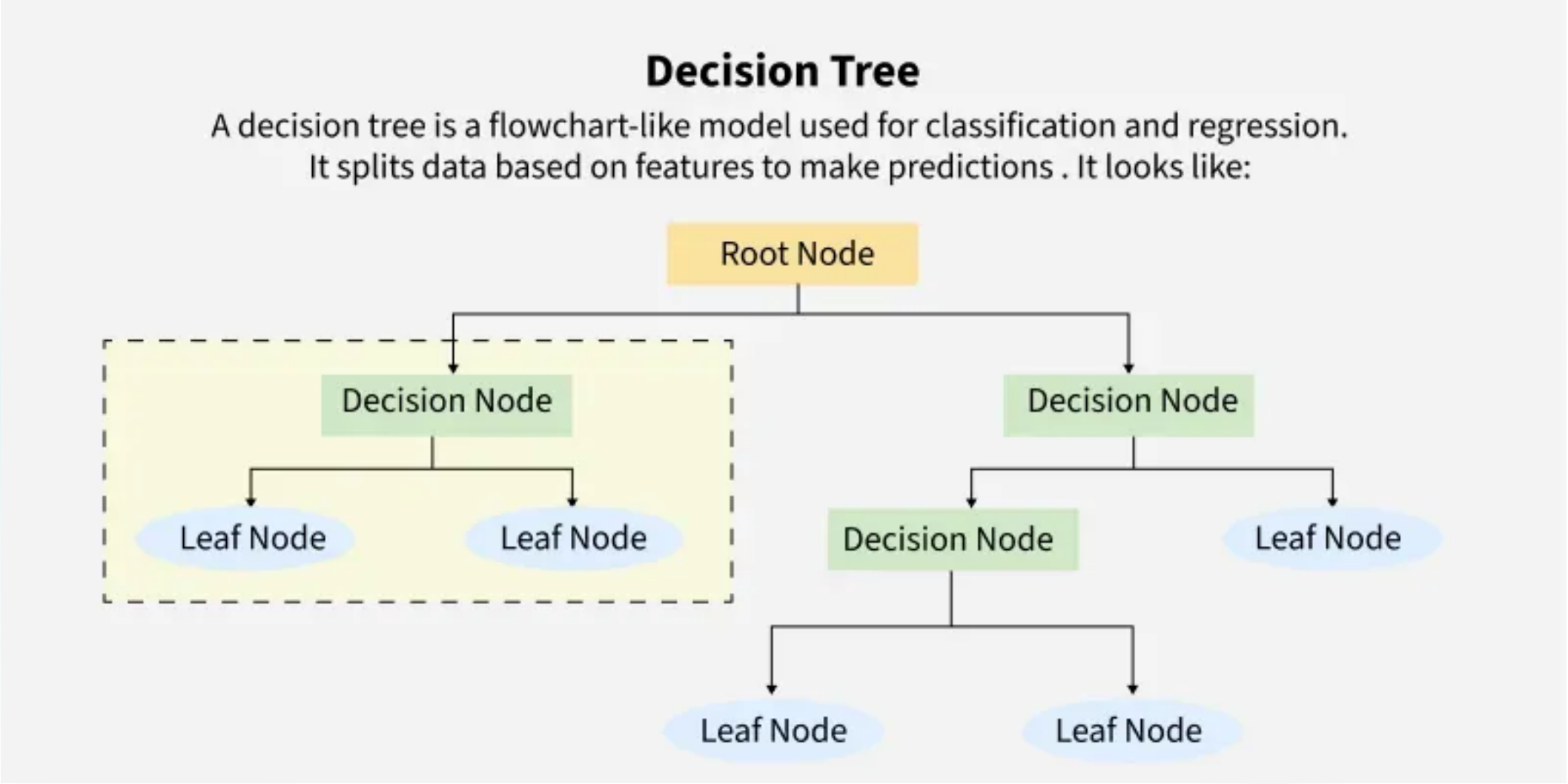
AI Series for Teachers

Decision Tree in Machine Learning

- A decision tree is a supervised learning algorithm used for both classification and regression tasks. It has a hierarchical tree structure which consists of a root node, branches, internal nodes and leaf nodes. It It works like a flowchart help to make decisions step by step where:
- Internal nodes represent attribute tests
- Branches represent attribute values
- Leaf nodes represent final decisions or predictions.

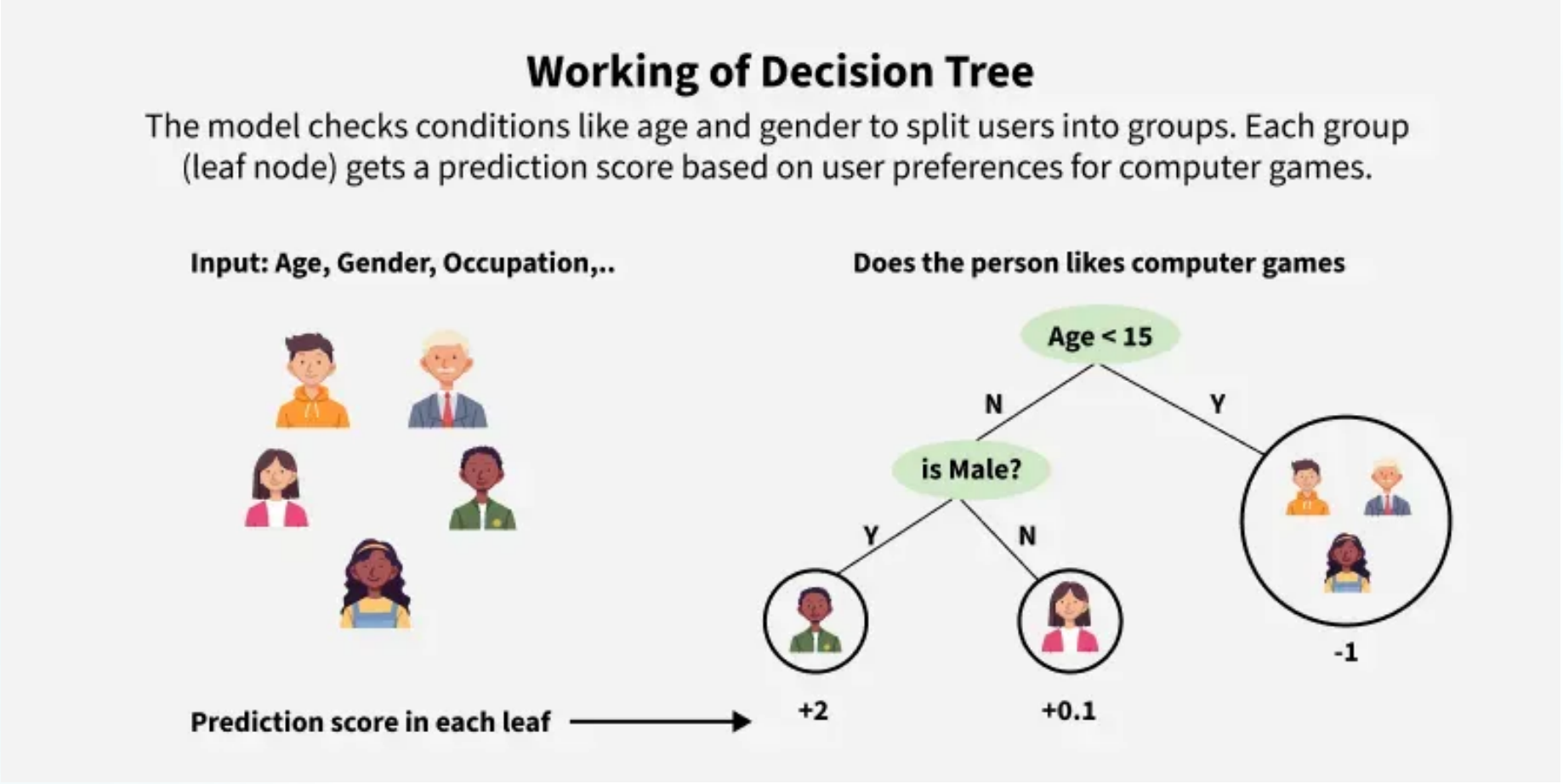
AI Series for Teachers

Decision Tree in Machine Learning



AI Series for Teachers

Decision Tree in Machine Learning



AI Series for Teachers

Decision Tree in Machine Learning

Splitting Criteria In Decision Tree

In decision trees, splitting criteria help decide which feature to split on at each node. The two most common criteria are:

Gini Index

$$I_G = 1 - \sum_{j=1}^{c} p_j^2$$

P_j : proportion of the samples that belongs to class c for a particular node

Entropy

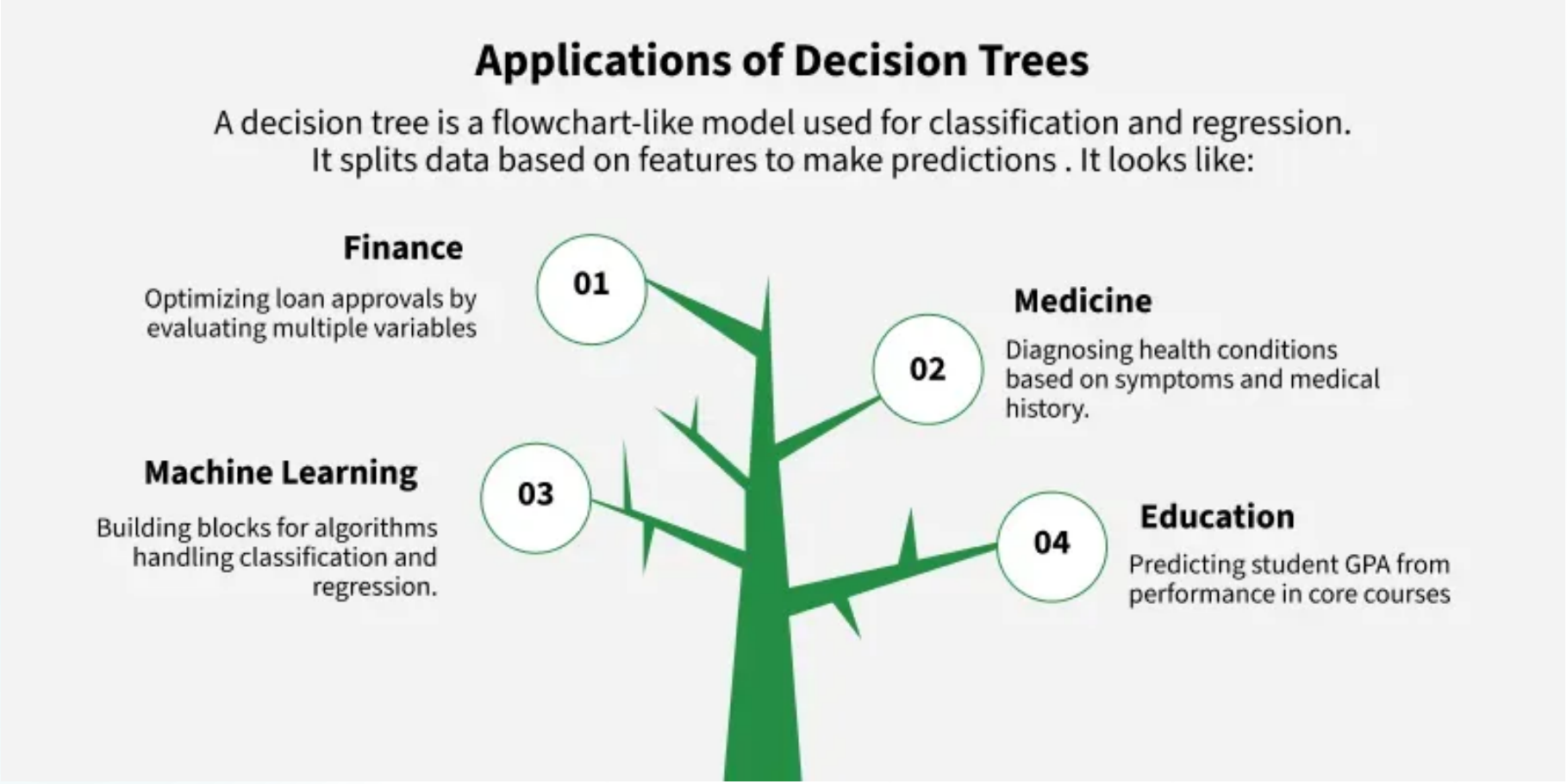
$$I_H = -\sum_{j=1}^{c} p_j log_2(p_j)$$

p_j: proportion of the samples that belongs to class c for a particular node.

*This is the the definition of entropy for all non-empty classes (p ne0) The entropy is 0 if all samples at a node belong to the same class.

AI Series for Teachers

Decision Tree in Machine Learning



AI Series for Teachers

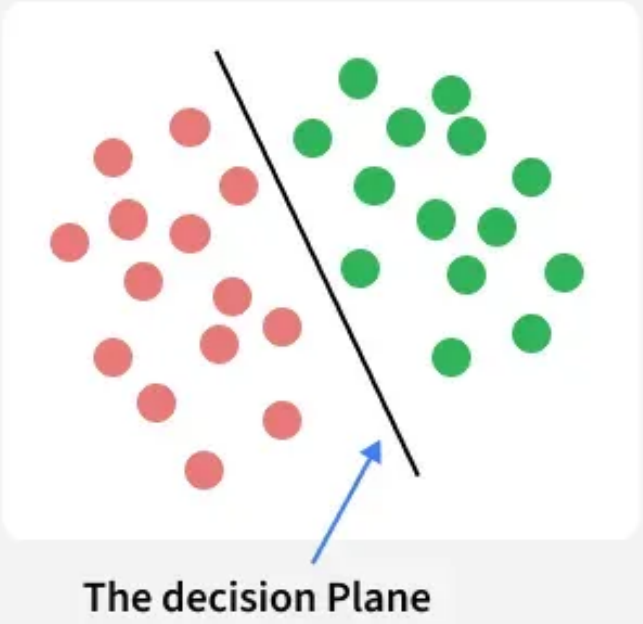
Support Vector Machine (SVM) Algorithm

- Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It tries to find the best boundary known as hyperplane that separates different classes in the data. It is useful when you want to do binary classification like spam vs. not spam or cat vs. dog.
- The main goal of SVM is to maximize the margin between the two classes. The larger the margin the better the model performs on new and unseen data.

AI Series for Teachers

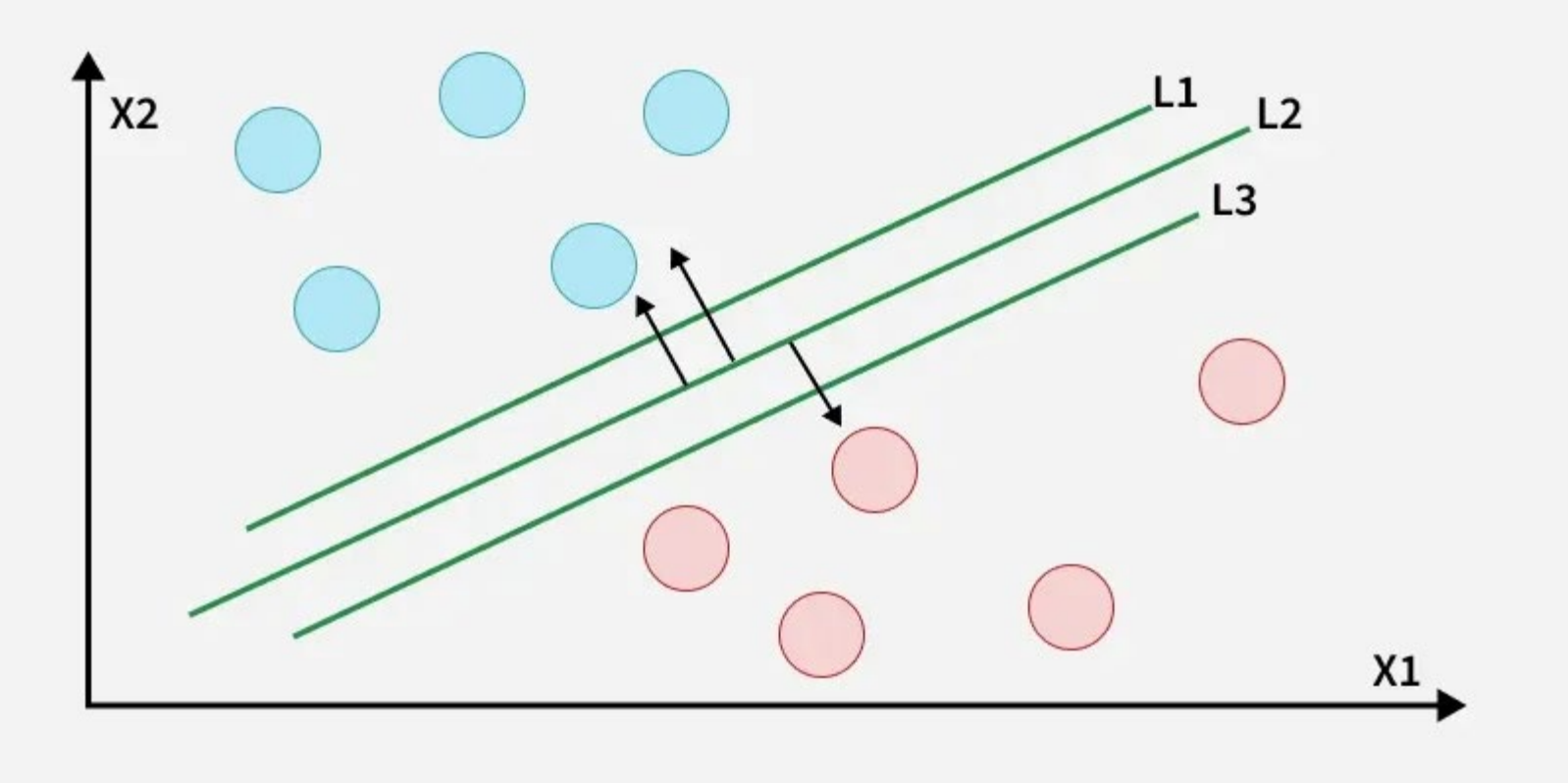
Support Vector Machine (SVM) Algorithm

- Support Vector Machine is used for classification and regression.
- It works by finding decision planes that separate data into different classes.



AI Series for Teachers

Support Vector Machine (SVM) Algorithm

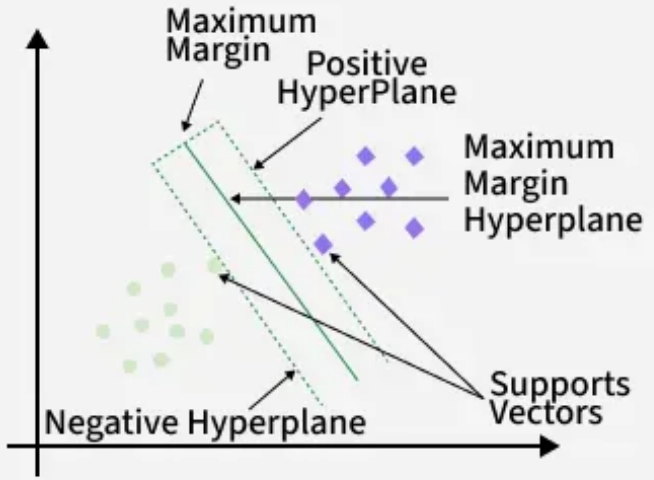


AI Series for Teachers

Support Vector Machine (SVM) Algorithm

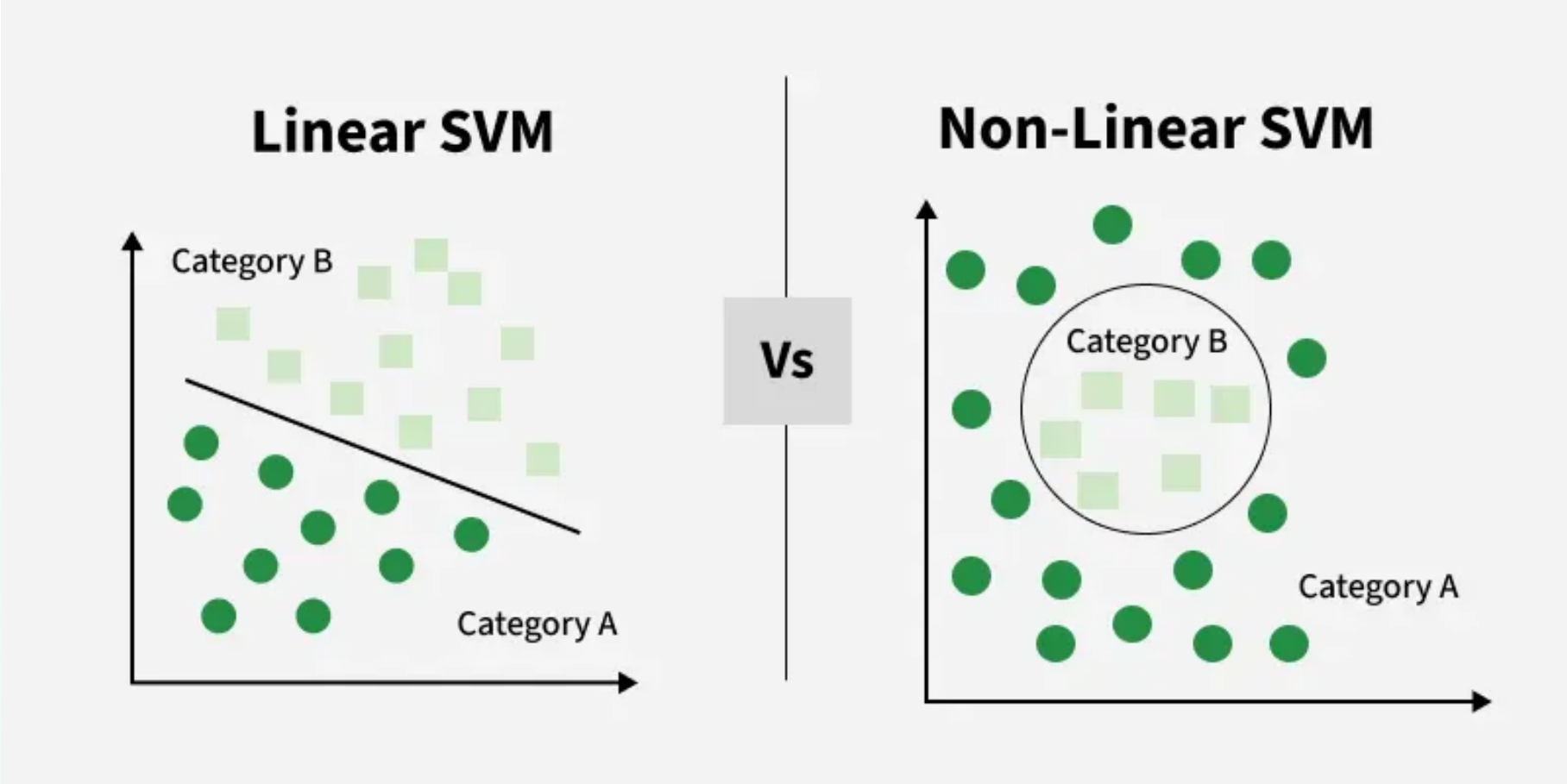
Support Vectors & Hyperplane

- Support Vectors are the closest data points to the hyperplane that define the class boundary.
- A hyperplane is a plane that separates different classes.



AI Series for Teachers

Support Vector Machine (SVM) Algorithm



AI Series for Teachers

Random Forest Algorithm in Machine Learning

Random Forest is a machine learning algorithm that uses many decision trees to make better predictions. Each tree looks at different random parts of the data and their results are combined by voting for classification or averaging for regression which makes it as ensemble learning technique. This helps in improving accuracy and reducing errors.

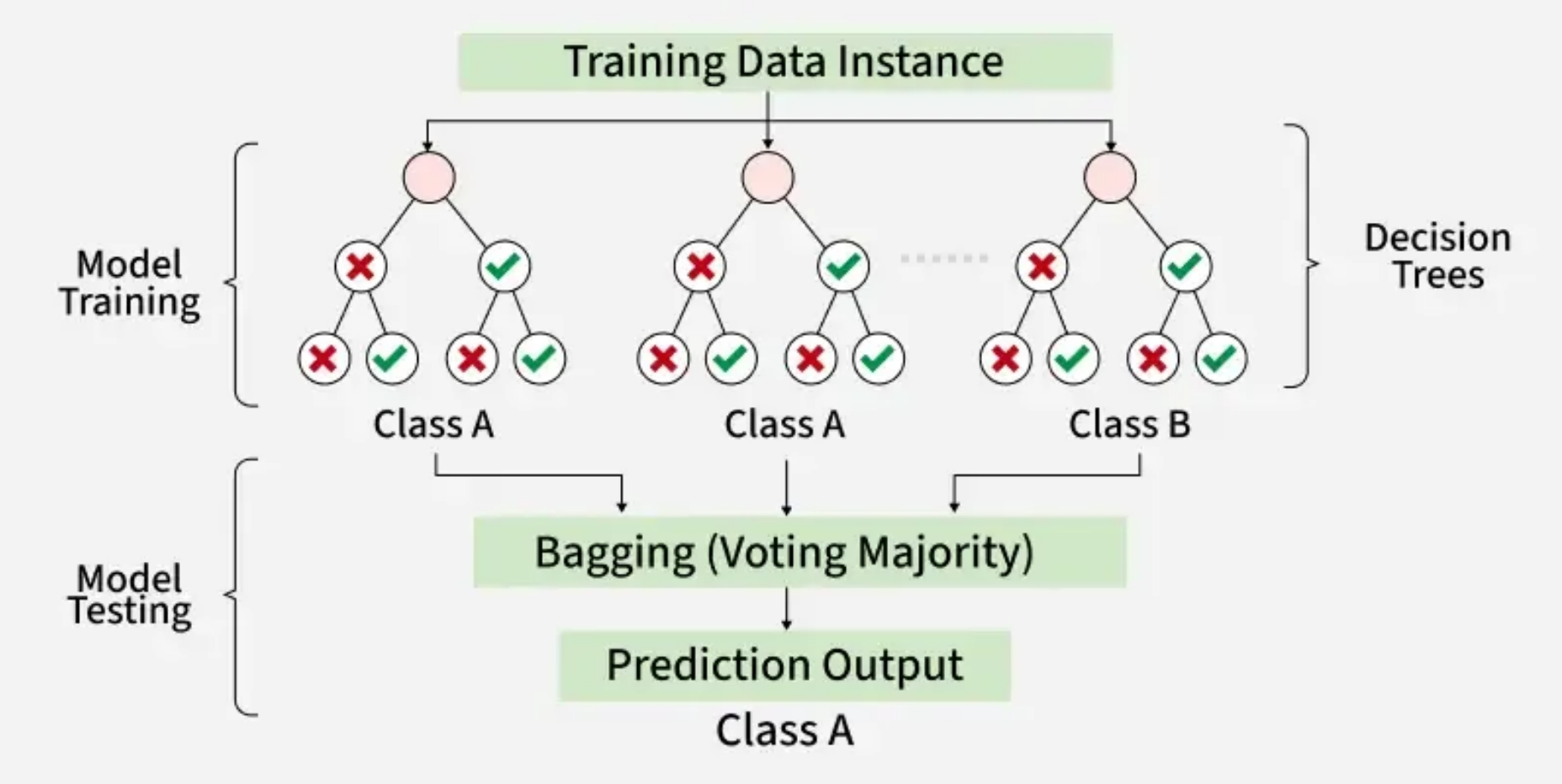
AI Series for Teachers

Working of Random Forest Algorithm

- **Create Many Decision Trees:** The algorithm makes many decision trees each using a random part of the data. So every tree is a bit different.
- **Pick Random Features:** When building each tree it doesn't look at all the features (columns) at once. It picks a few at random to decide how to split the data. This helps the trees stay different from each other.
- **Each Tree Makes a Prediction:** Every tree gives its own answer or prediction based on what it learned from its part of the data.
- **Combine the Predictions:** For **classification** we choose a category as the final answer is the one that most trees agree on i.e majority voting and for **regression** we predict a number as the final answer is the average of all the trees predictions.
- **Why It Works Well:** Using random data and features for each tree helps avoid overfitting and makes the overall prediction more accurate and trustworthy.

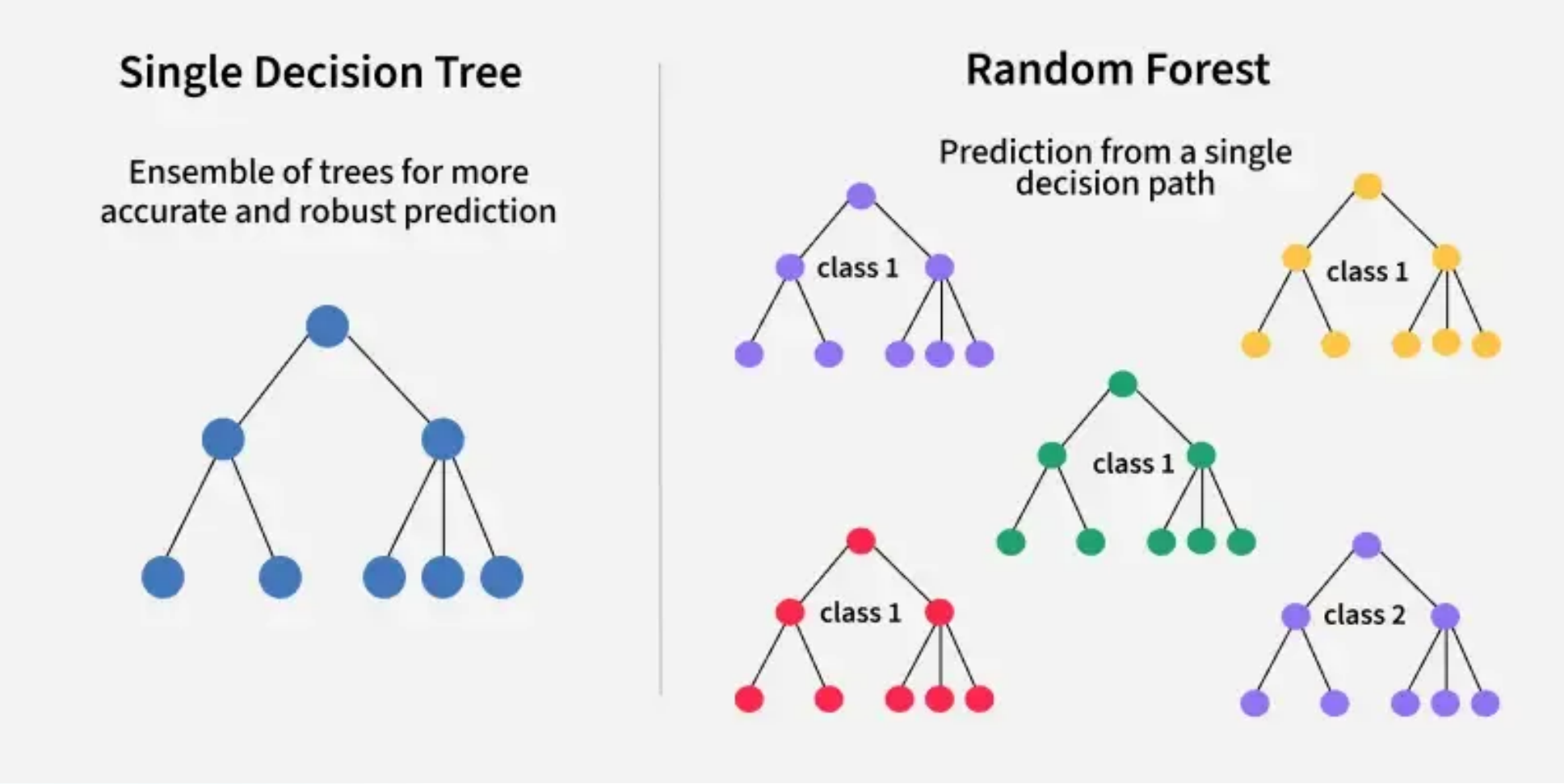
AI Series for Teachers

Working of Random Forest Algorithm



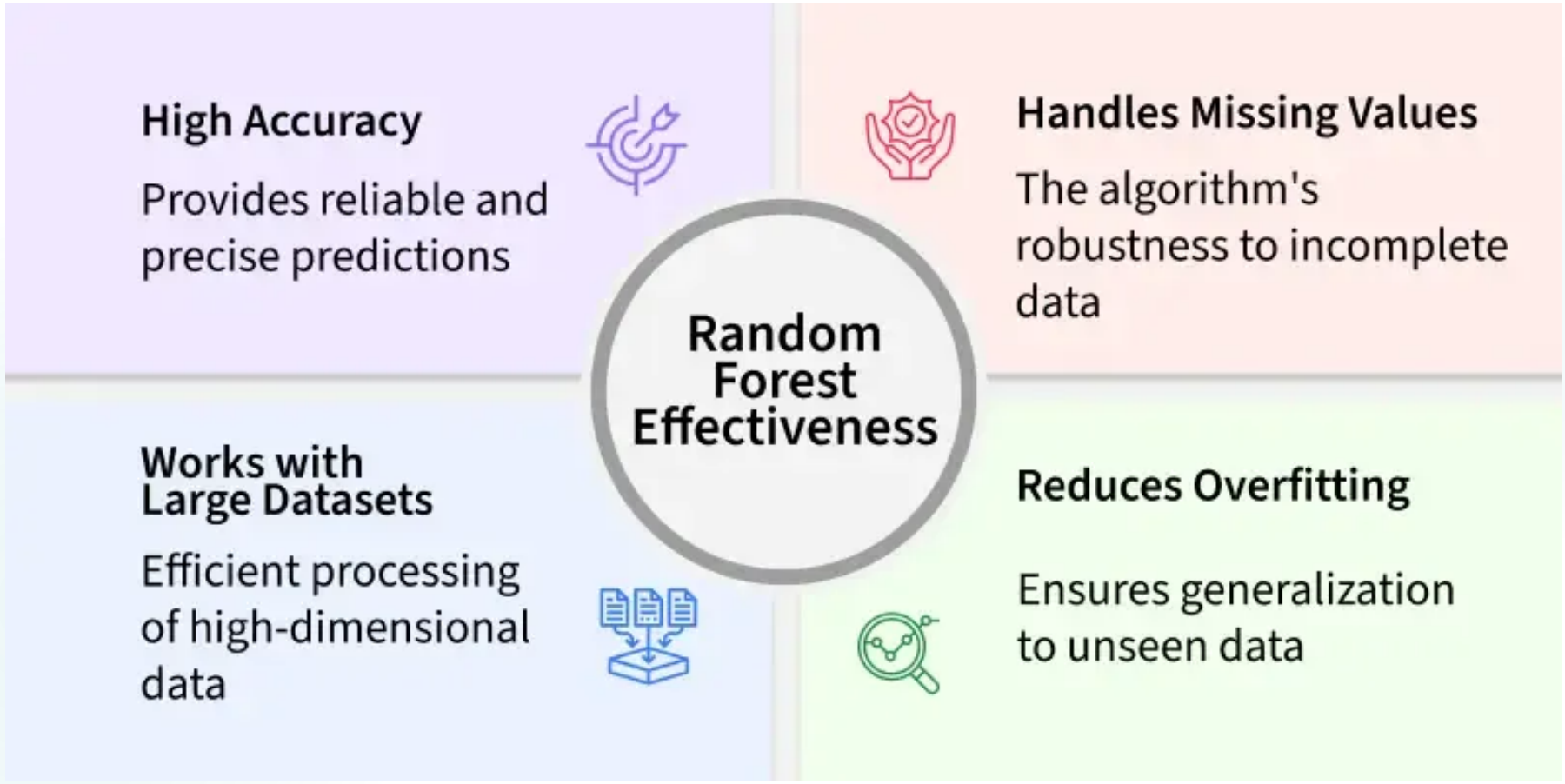
AI Series for Teachers

Working of Random Forest Algorithm



AI Series for Teachers

Advantage of Random Forest Algorithm



AI Series for Teachers

Unsupervised Learning:

Unsupervised learning is the **training** of machine using information that is **neither classified nor labeled** and allowing the algorithm to act on that information **without guidance**. Here the task of machine is to **group unsorted information** according to **similarities, patterns and differences** without any **prior training** of data. Unsupervised machine learning is more challenging than supervised learning due to the absence of labels.

Types of Unsupervised Learning:

- ❑ **Clustering**
- ❑ **Association**

AI Series for Teachers

Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Examples of unsupervised learning algorithms are:

- k-means for clustering problems.
- Apriori algorithm for association rule learning problems

AI Series for Teachers

K means Clustering – Introduction

K-Means Clustering is an unsupervised machine learning algorithm that helps group data points into clusters based on their inherent similarity. Unlike supervised learning, where we train models using labeled data, K-Means is used when we have data that is not labeled and the goal is to uncover hidden patterns or structures. For example, an online store can use K-Means to segment customers into groups like "Budget Shoppers," "Frequent Buyers," and "Big Spenders" based on their purchase histor

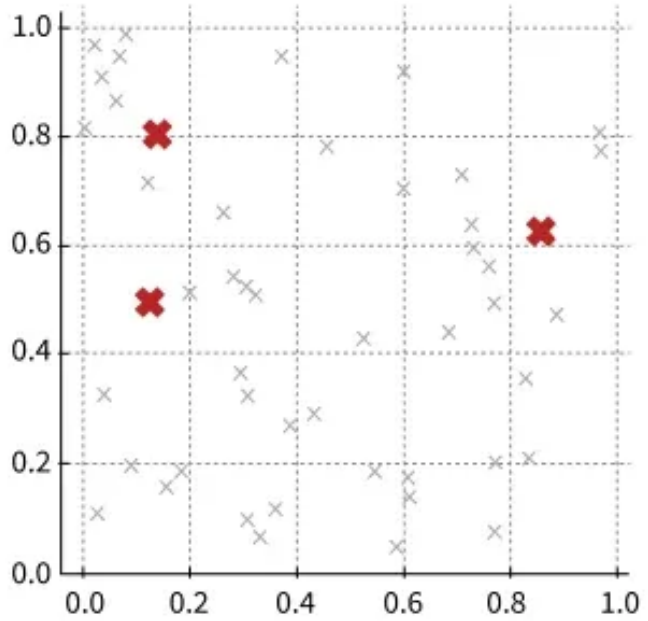
AI Series for Teachers

Working of K-means Clustering

Choose Initial Centroids

Centroids are randomly chosen from the data points. These represent the initial cluster centers.

✖ Centroids × Data Points



AI Series for Teachers

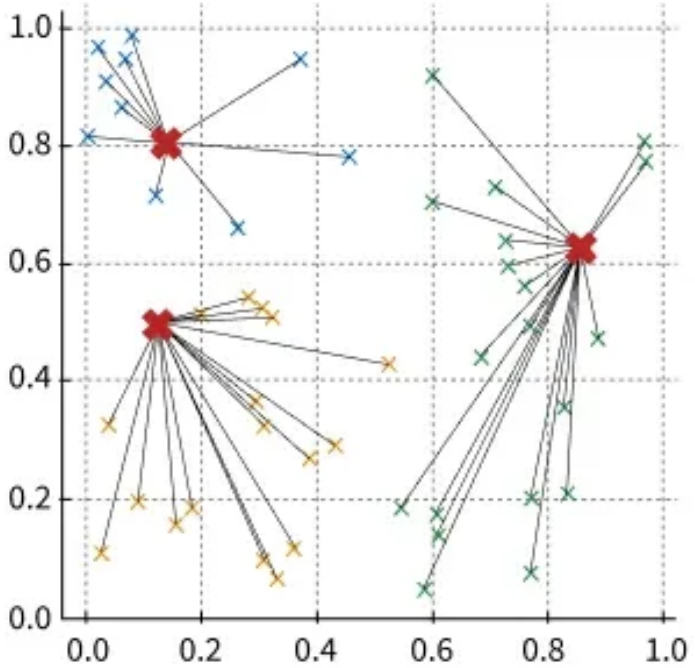
Working of K-means Clustering

Assign Points to Nearest Centroid

Each point is assigned to the nearest centroid, forming clusters

Centroids Cluster 1

Cluster 2 Cluster 3



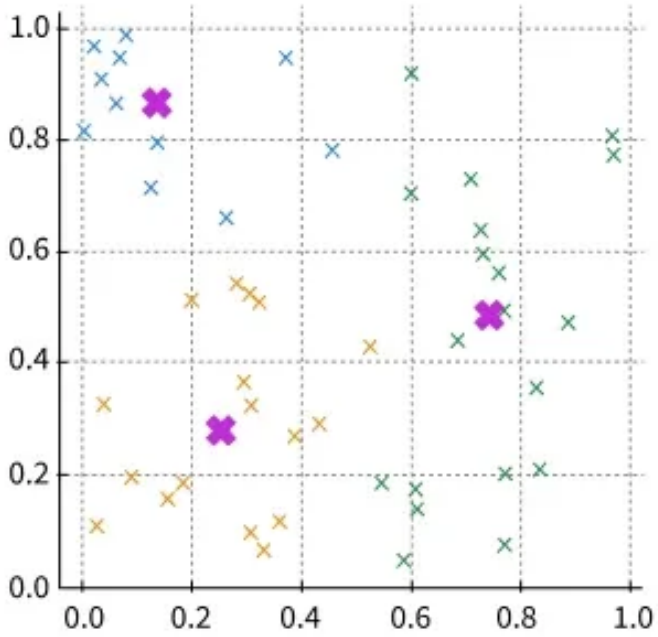
AI Series for Teachers

Working of K-means Clustering

Update Centroids

Centroids are recalculated as the mean of the points in each cluster

- New Centroids
- Cluster 1
- Cluster 2
- Cluster 3



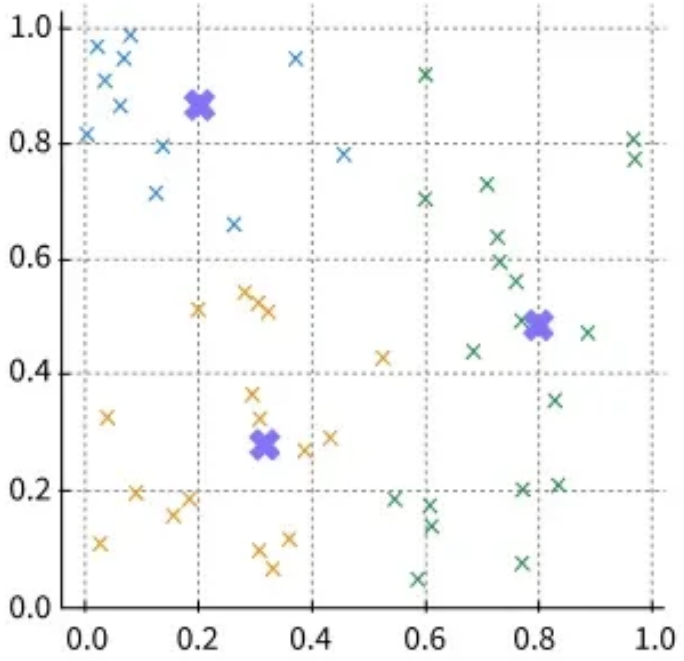
AI Series for Teachers

Working of K-means Clustering

Repeat Until Convergence

This process repeats until the centroids stabilize and do not move further.

- Final Centroids
- Cluster 1
- Cluster 2
- Cluster 3



AI Series for Teachers

Apriori algorithm for association rule learning problems

- Apriori Algorithm is a basic method used in data analysis to find groups of items that often appear together in large sets of data. It helps to discover useful patterns or rules about how items are related which is particularly valuable in market basket analysis.

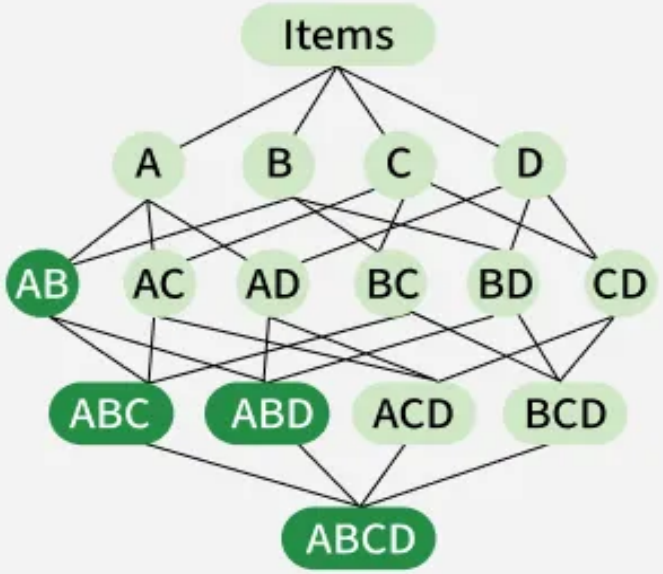
- Like in a grocery store if many customers buy bread and butter together, the store can use this information to place these items closer or create special offers. This helps the store sell more and make customers happy.

AI Series for Teachers

Apriori algorithm for association rule learning problems

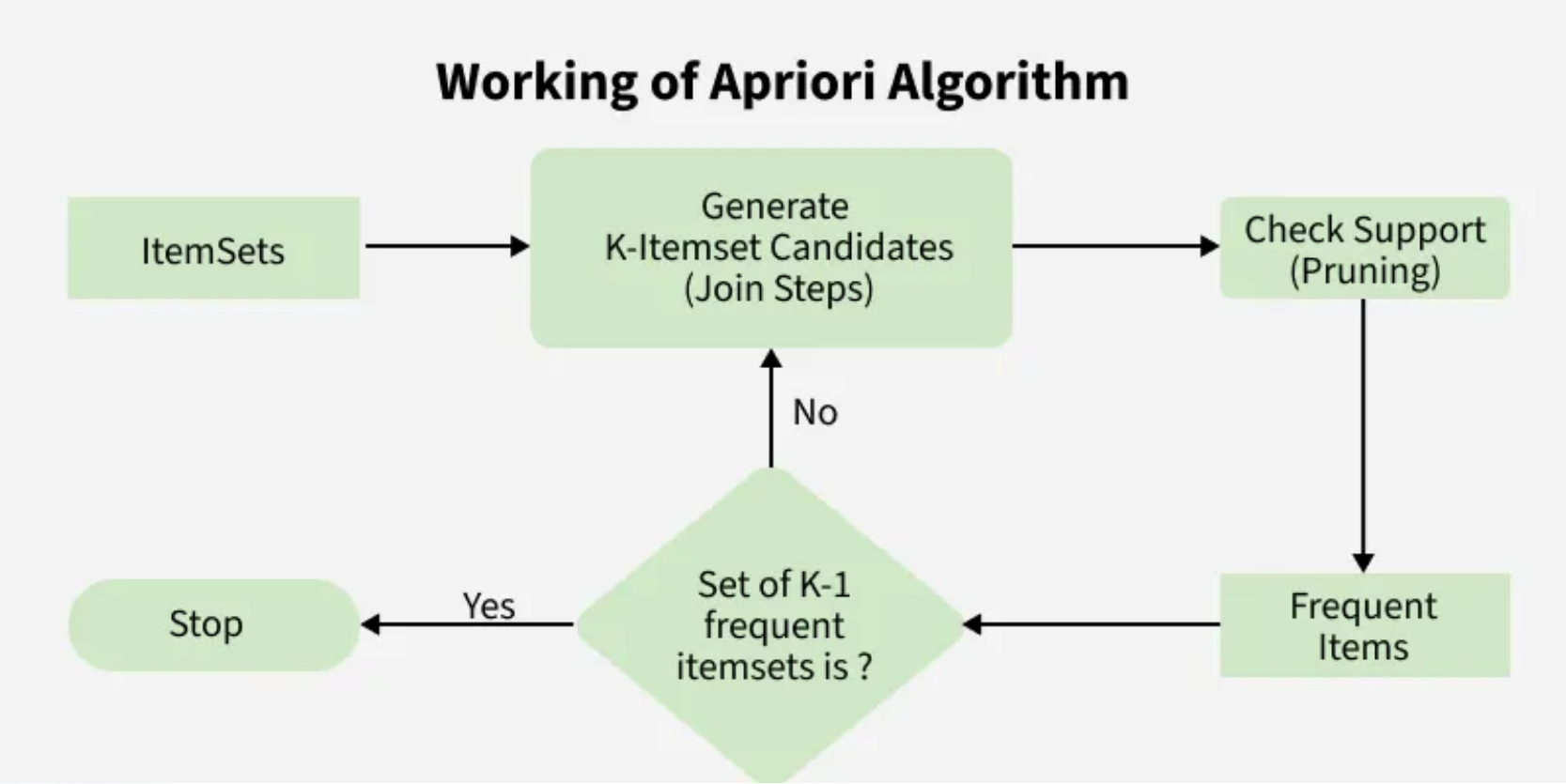
Apriori algorithm

The Apriori algorithm is a machine learning algorithm to identify relationships between items by identifying frequent itemsets.



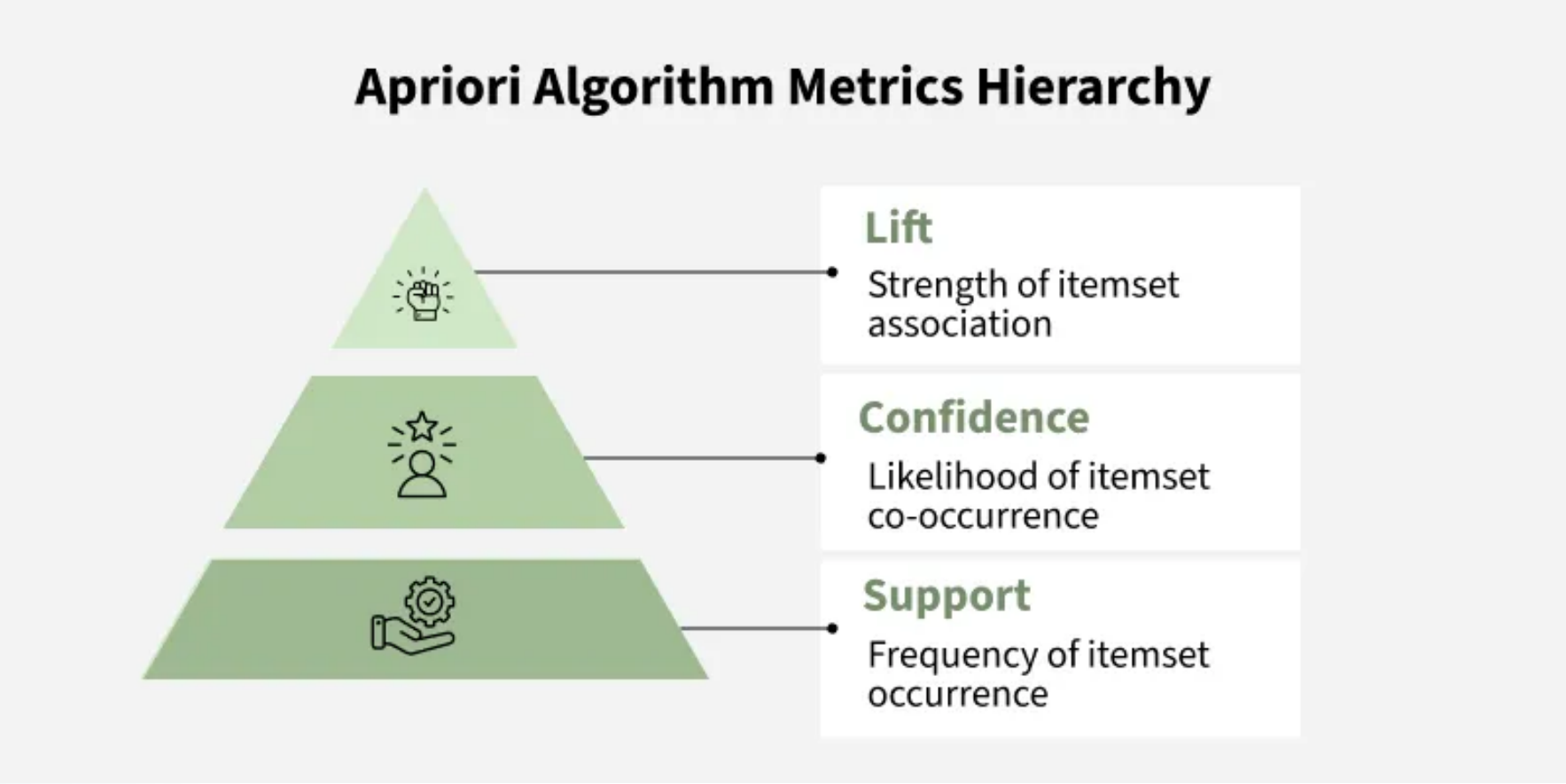
AI Series for Teachers

Apriori algorithm - Working



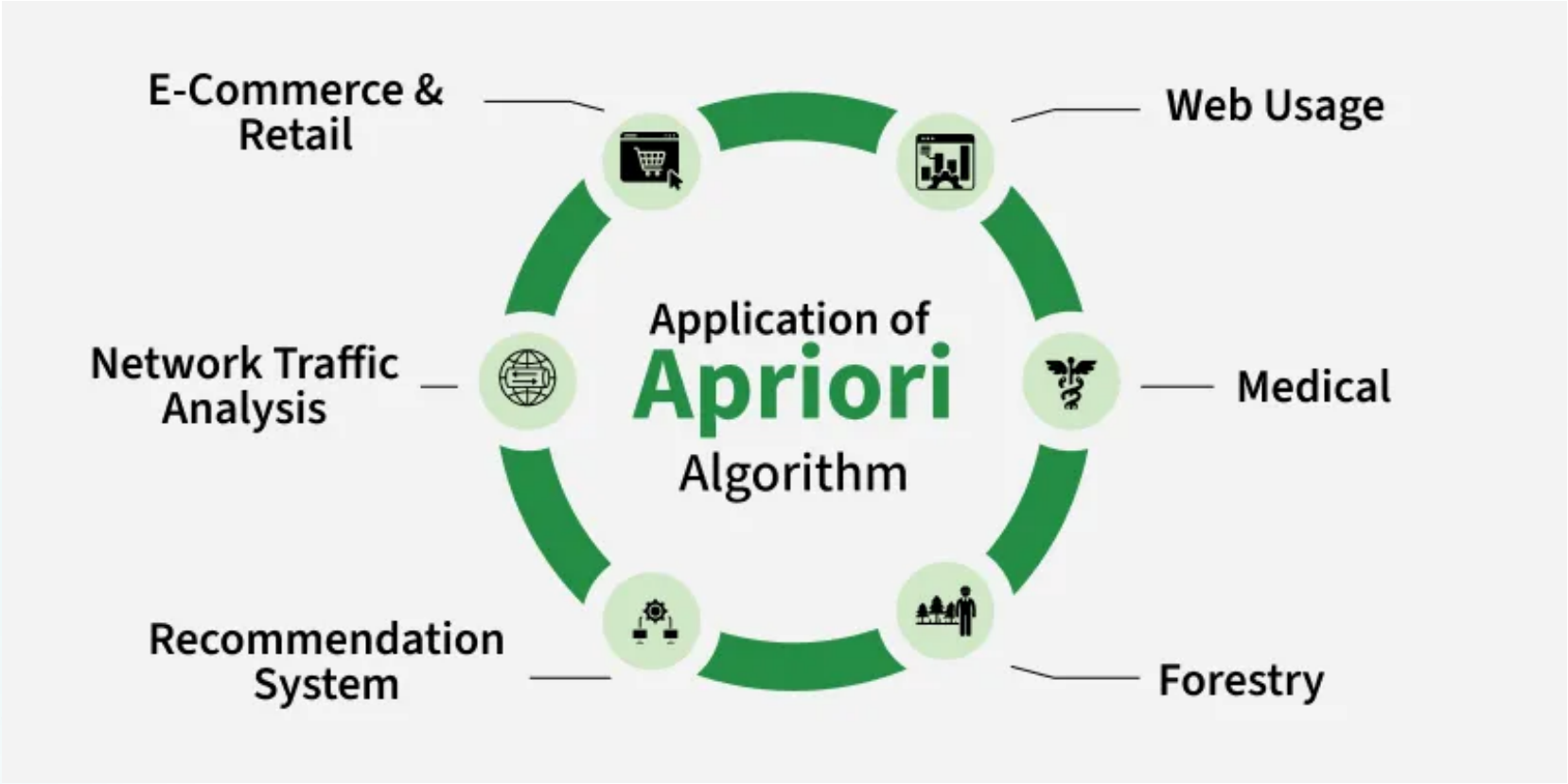
AI Series for Teachers

Apriori algorithm - Metrics



AI Series for Teachers

Apriori algorithm - Applications



AI Series for Teachers

Disadvantage of **Supervised Learning** and **Unsupervised Learning**

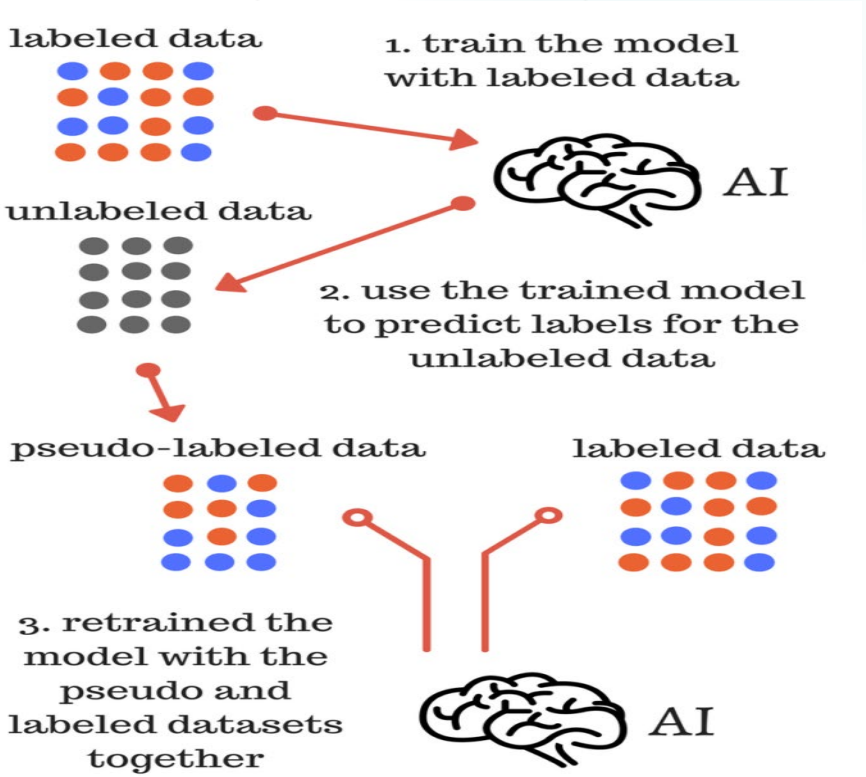
- The most basic disadvantage of any **Supervised Learning** algorithm is that the dataset has to be hand-labeled either by a Machine Learning Engineer or a Data Scientist. This is a very ***costly process***, especially when dealing with large volumes of data. The most basic disadvantage of any **Unsupervised Learning** is that it's **application spectrum is limited**.

AI Series for Teachers

Semi-supervised machine learning:

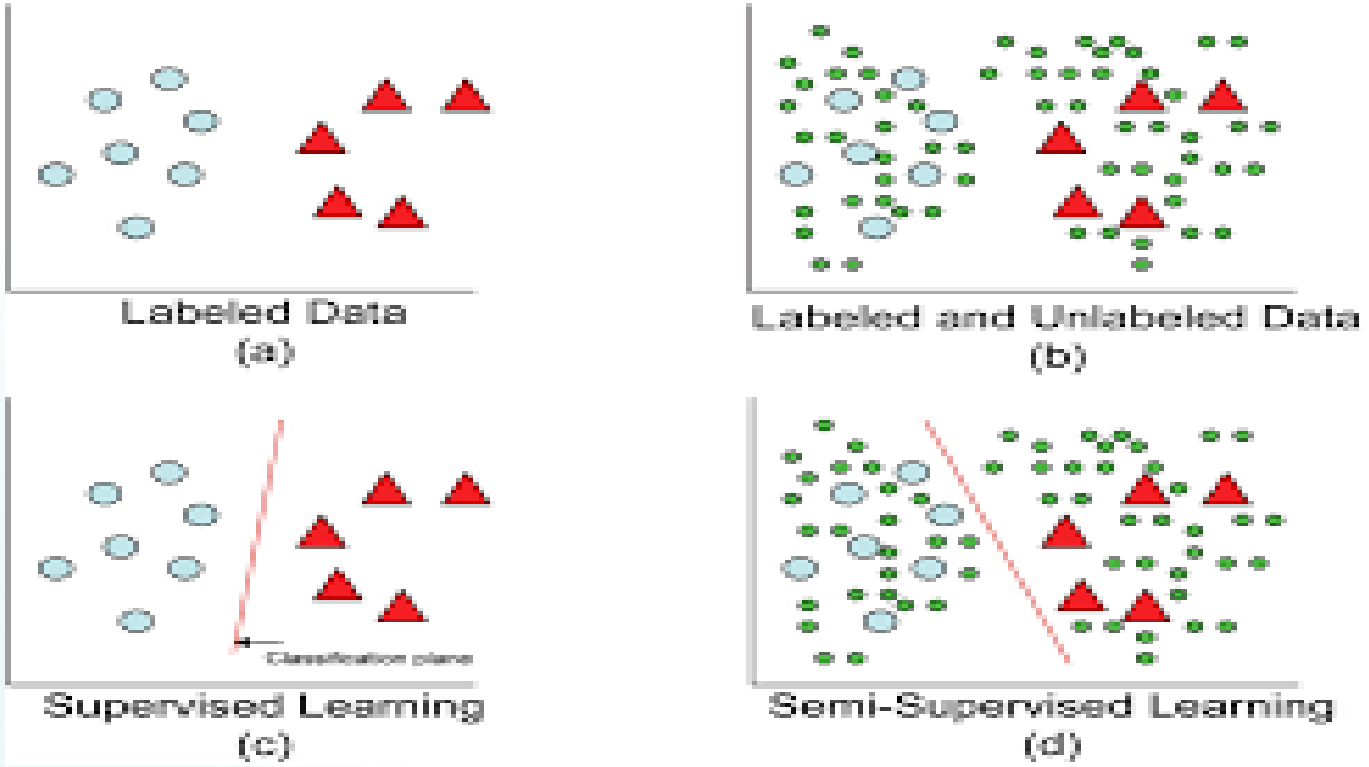
To counter these disadvantages, the concept of **Semi-Supervised Learning** was introduced. In this type of learning, the algorithm is trained upon a combination of labeled and unlabeled data. Typically, this combination will contain a very small amount of labeled data and a very large amount of unlabeled data.

- In semi supervised learning labelled data is used to learn a model and using that model unlabeled data is labelled called pseudo labelling now using whole data model is trained for further use



AI Series for Teachers

Model with labellled data and model with both labelled and unlabelled data



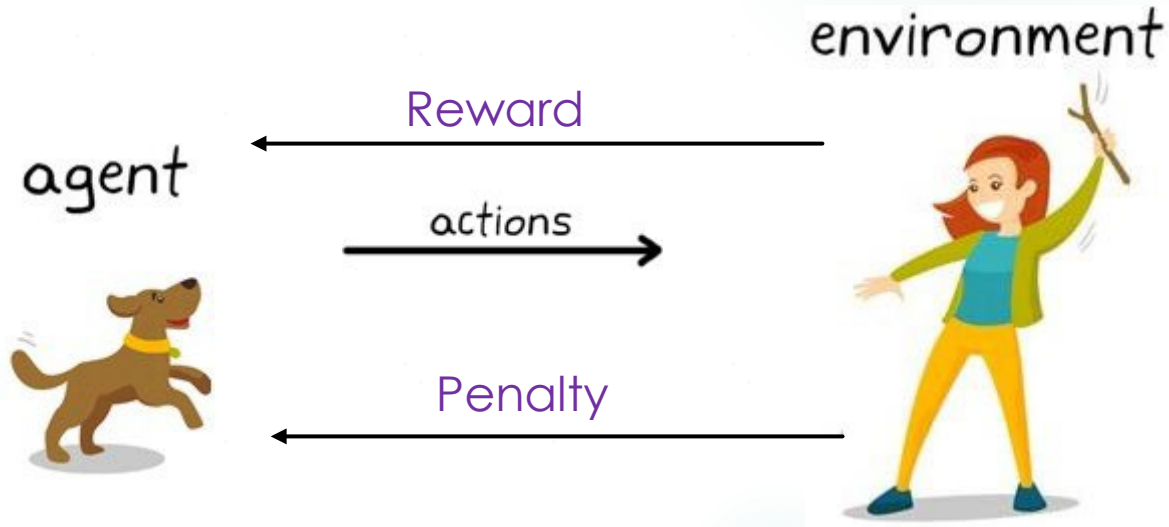
Intuitively, one may imagine the three types of learning algorithms as Supervised learning where a student is under the supervision of a teacher at both home and school, Unsupervised learning where a student has to figure out a concept himself and Semi-Supervised learning where a teacher teaches a few concepts in class and gives questions as homework which are based on similar concepts.

AI Series for Teachers

Reinforcement Learning

Reinforcement Learning (RL) is a branch of machine learning that focuses on how agents should act in an environment to maximize cumulative rewards. It is inspired by behavioural psychology, where agents learn through interaction with the environment and feedback. RL has shown promising results in robotics, game-playing AI, and autonomous vehicles.

- Examples:
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Model : **Q-Learning**



AI Series for Teachers

AI/ML in Education

Use of Computers in Education are primarily aimed towards:

- Increasing accessibility - Learning resources can be accessed from anywhere, at any time
- Facilitating personalized learning inside and outside the classroom. Learning can be tailored and adapted to each student's goals and abilities through personalized programs,
- Time-consuming, tedious tasks such as record keeping or grading multiple-choice tests can be completed through AI automation, and
- Exploring fundamental questions about how people learn

AI Series for Teachers

Explore ML applications in education: personalization, grading and prediction

Machine learning (ML) is rapidly transforming education by enabling enhanced personalization, automated grading, and powerful predictive analytics. These applications expand individualized support, improve assessment efficiency, and help forecast learner outcomes for targeted intervention.

AI Series for Teachers

Personalization in Education

ML-driven personalization allows systems to tailor educational experiences based on individual student needs, behaviors, and learning patterns. Adaptive learning platforms like DreamBox and Knewton adjust content difficulty and pacing at a granular level. Intelligent Tutoring Systems (ITS), such as Carnegie Learning and ALEKS, provide dynamic, individualized instruction and feedback. Learning Management Systems (LMS) enhanced with AI recommend resources and adapt schedules, while NLP tools like Grammarly assist with writing feedback. These customized approaches boost engagement, improve retention, and support all types of learners.

AI Series for Teachers

Automated Grading

ML algorithms in automated grading systems have significantly increased grading efficiency and consistency. By leveraging supervised learning, NLP, and neural networks, these systems analyze programming assignments, essays, and objective tests for accuracy, coherence, and relevance. They not only assign grades with high precision but also generate individualized feedback for students, identifying misconceptions and suggest improvements. AI-powered graders reduce bias, manage large volumes of submissions, and free educators for higher-level pedagogical tasks.

AI Series for Teachers

Predictive Analytics

Predictive analytics uses ML to forecast future student performance and outcomes by analyzing academic, behavioral, social, and LMS data.

These models can:

•Identify at-risk students for timely intervention.

•Personalize learning paths by recognizing strengths and weaknesses.

•Segment learners for targeted recruitment or resource allocation.

•Recommend specific support services or content to improve success rates.

Schools using these tools have seen improvements in retention, early intervention, and optimized learning strategies. Predictive analytics also enables administrators to make data-driven decisions about curriculum and policy changes.

AI Series for Teachers

Summary Table: ML Applications in Education

Application	Core Function	Example Technologies/Uses
Personalization	Tailored content and pacing	DreamBox, ALEKS, Canvas LMS
Automated Grading	Efficient, objective assessment	Automated essay graders, code evaluators
Prediction	At-risk detection, learning path recommendations	Early warning systems, academic dashboards

AI Series for Teachers

Ethical risks of ML personalization in classrooms

ML personalization in classrooms carries major ethical risks such as privacy violations, reinforcing bias and discrimination, surveillance, and diminishing student autonomy. These challenges require careful consideration as schools adopt data-driven technologies

- ❑ Privacy and Surveillance Concerns
- ❑ Bias, Fairness, and Discrimination
- ❑ Autonomy and Agency
- ❑ Mitigation Strategies

AI Series for Teachers

Privacy and Surveillance Concerns

Personalized ML tools depend on collecting and analyzing sensitive student data, including demographics, learning behaviors, and performance metrics. Risks include:

- Centralization and storage of this data make it vulnerable to breaches and misuse, exposing students' identities and private information.
- Continuous monitoring may make students feel surveilled, suppressing discussion and creativity in learning environments.
- Consent procedures are often insufficient, with students and guardians unclear about what data is collected or how it will be used

AI Series for Teachers

Bias, Fairness, and Discrimination

ML algorithms reflect the data and social context they are built upon. Key risks include:

- Historical or selection biases in training data can lead to algorithms that disadvantage marginalized or underrepresented groups, perpetuating systemic inequities.
- Opaque and complex models may reinforce unfairness without transparency or recourse for affected students.
- Targeted personalization could inadvertently prioritize certain behaviors or achievements, aggravating gaps between students

AI Series for Teachers

Privacy and Surveillance Concerns

Autonomy and Agency

- Over-reliance on ML-driven recommendations can reduce students' ability to make independent decisions, stifling autonomy and critical self-reflection.
- Mandating the use of personalized systems as a requirement for public education restricts choice for students and families who may disagree with data use practices.

AI Series for Teachers

Mitigation Strategies

- Employ privacy-preserving technologies (such as federated learning or differential privacy) to protect student data.
- Conduct regular audits for bias and incorporate fairness-aware algorithms to reduce discrimination.
- Embed transparency, explainability, and robust consent procedures to empower students and families.
- Ethical deployment of ML personalization in classrooms demands balancing educational benefits with stringent protections for privacy, fairness, and student well-being

THANK YOU

Dr. Amit Bhagat
Associate Professor
Deptt. of Mathematics, Bioinformatics & Computer Applications,
Maulana Azad National Institute of Technology, Bhopal
amitbhagat@manit.ac.in