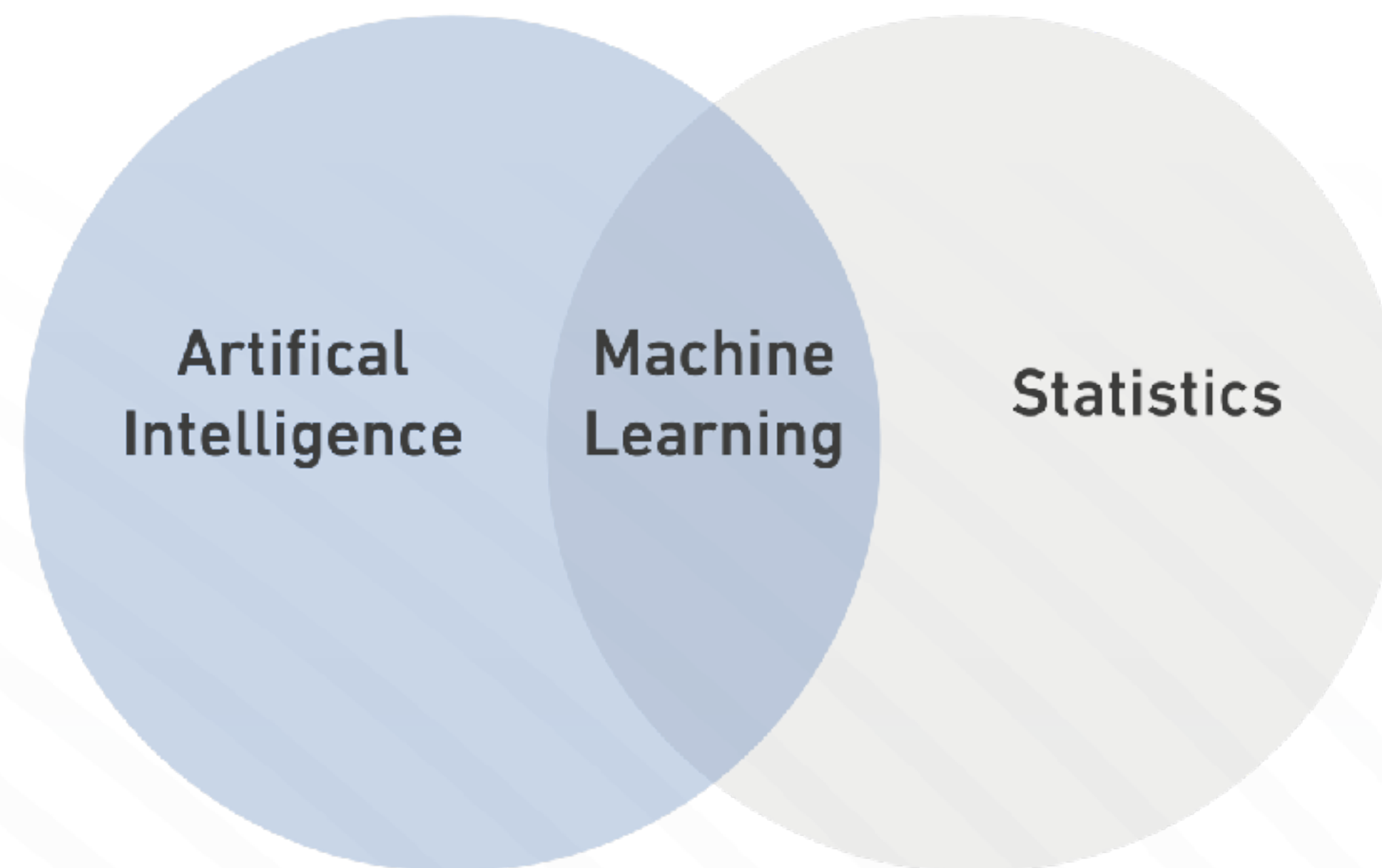




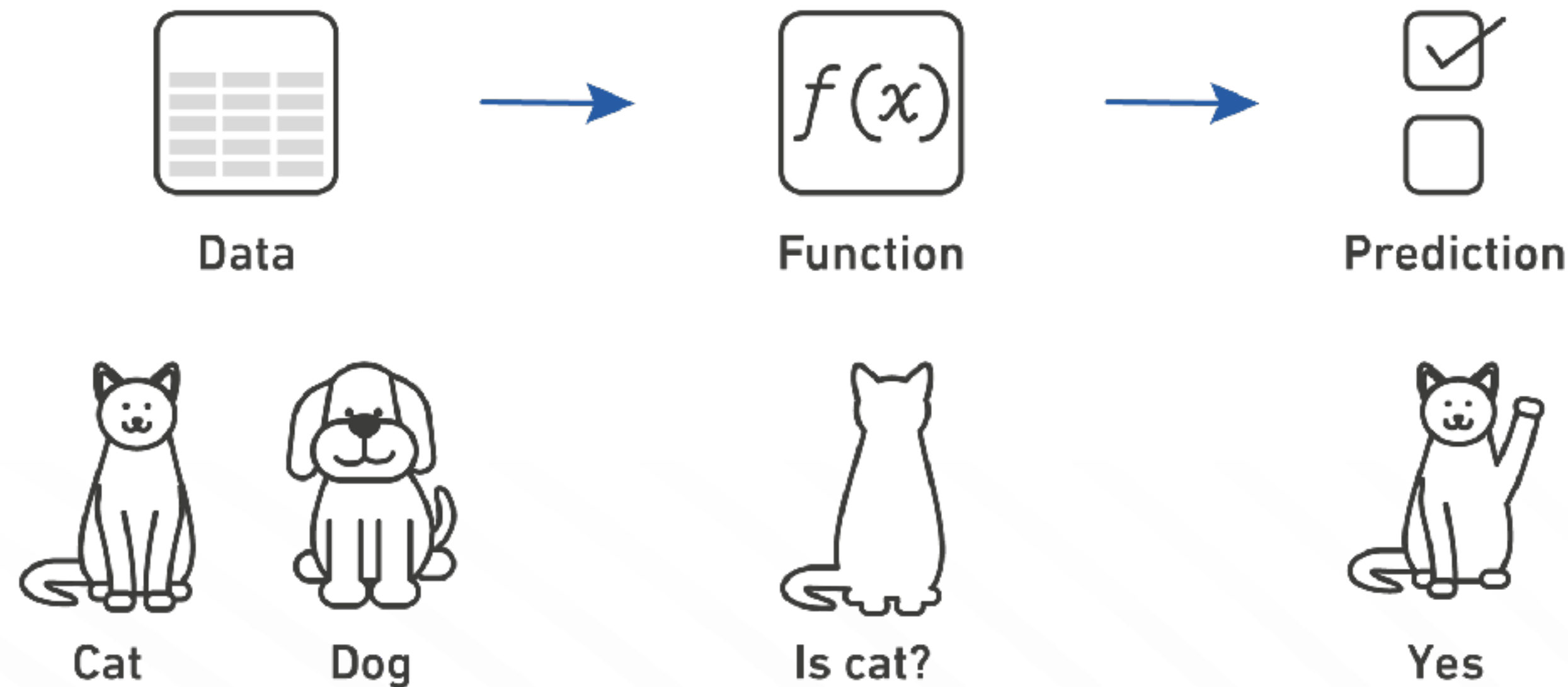
مقدمة في تعلم الآلة

تعلم الآلة (Machine Learning)

- هو عملية تعليم الآلة إنجاز مهمة معينة دون كتابة كود صريح أو أوامر صريحة لتنفيذ هذا الأمر، ويمكن التعبير عنه بأنه جزء الذكاء الاصطناعي الذي يحتوي على إحصائيات.
- أثناء عملية تعليم الآلة نقوم بإنشاء ما يسمى بالنموذج (Model) الذي نقوم بتزويده بمجموعة البيانات و الخوارزمية (algorithm) للتعلم من البيانات.

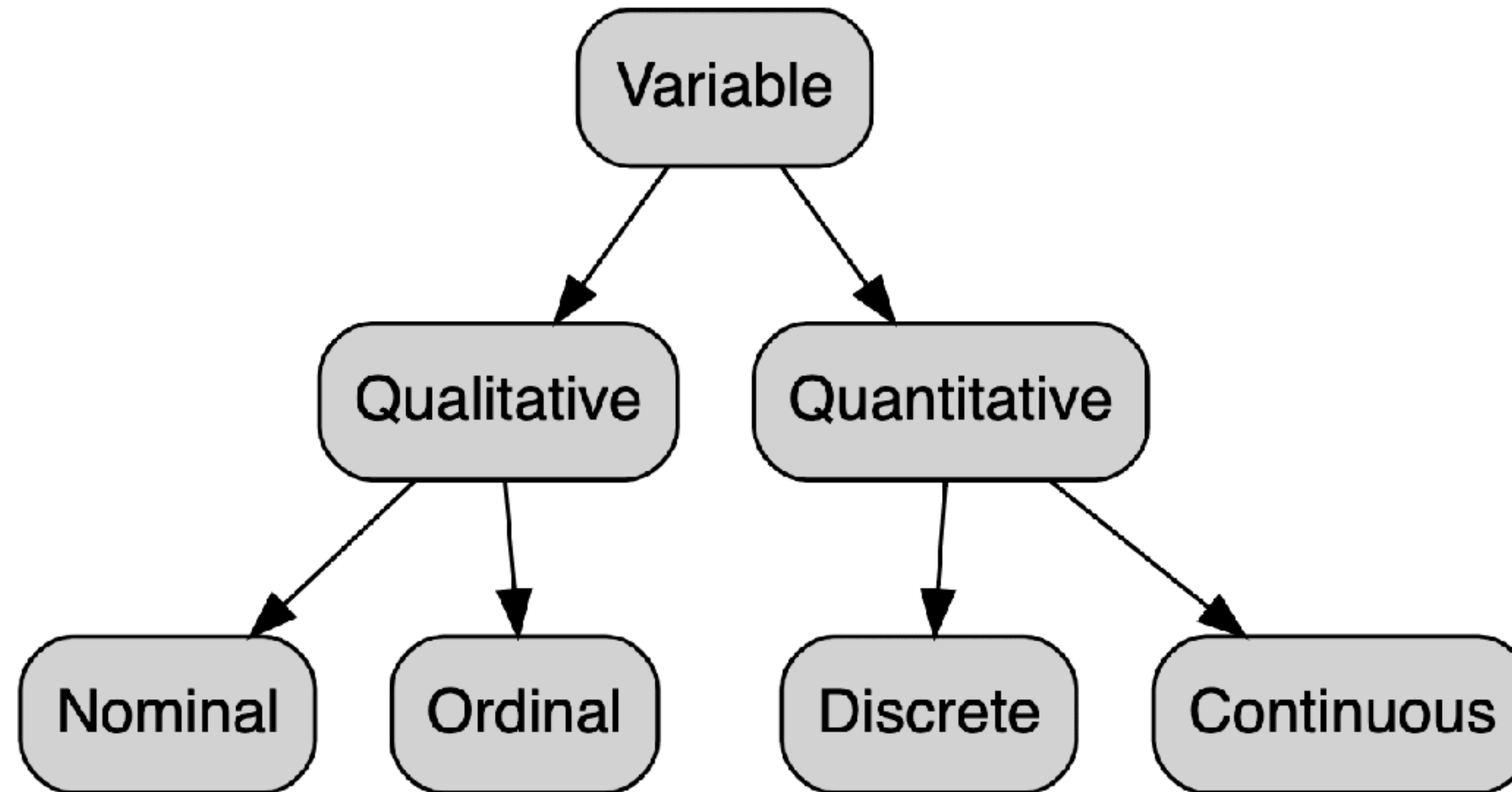


تعلم الآلة (Machine Learning)



- نقوم باستعمال البيانات لتعلم بناء دالة تكون قادرة على التنبؤ بنتيجة البيانات الجديدة على سبيل المثال لنقل أننا نريد بناء دالة تقوم بتحديد هل الصورة تحتوي على قطة أم لا؟
- في البداية سوف نقوم بإنشاء بيانات تحتوي صور للقطط وصور لاحتوي ذلك .
- ثم نقوم بتطبيق خوارزميات تعلم الآلة على مجموعة البيانات.
- تقوم هذه الخوارزميات بتعلم الدالة التي تتنبأ بالصور هل الصورة تحتوي على قطة أم لا؟

أنواع البيانات



خوارزميات تعلم الآلة (Machine Learning Algorithms)

Supervised Learning

Labeled Data
Direct Feedback
Classification and Regression

Unsupervised Learning

Unlabeled Data
No Feedback
Clustering & Dimensionality Reduction

Semi-supervised Learning

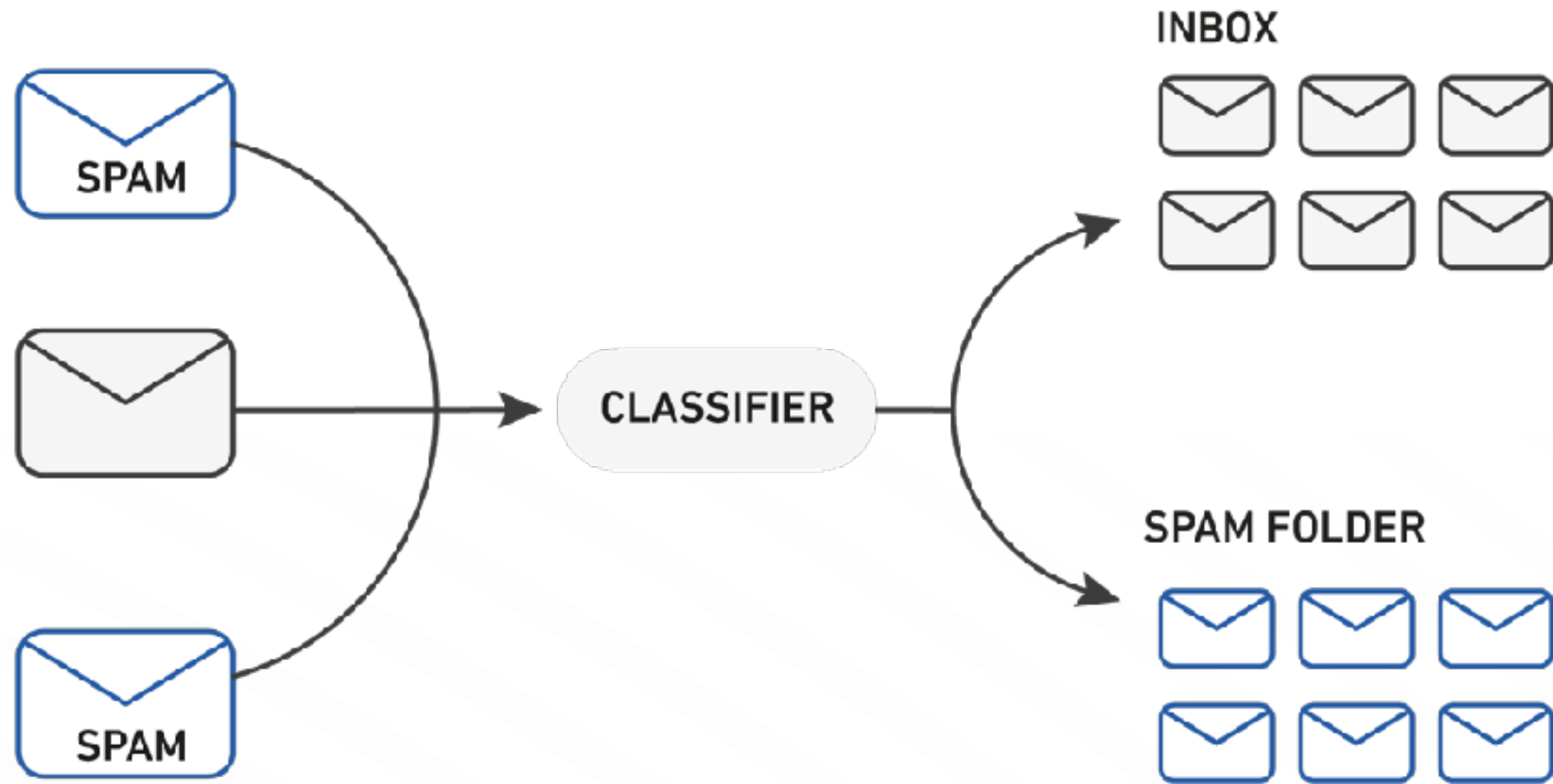
Labeled and Unlabeled Data
Some Feedback
Classification and Regression

Reinforcement Learning

Reward Based Learning
Direct Feedback
Learn series of actions

خوارزميات تعلم الآلة (Machine Learning Algorithms)

النوع الأول (Supervised Learning)



يشير هذا النوع إلى العملية التي تتعلم فيها الآلات من مجموعات البيانات المصنفة أو المعروفة (labeled data) وتنتج نموذجًا دقيقًا قادرًا على التنبؤ بملصقات البيانات غير المرئية.

مثال: عندما نقوم بتدريب الآلة على تصنيف البريد إلى بريد مزعج (spam) أو غير مزعج (spam) خلال هذه العملية تتعلم فيها الآلة من مجموعة من البيانات المصنفة إلى بريد مزعج و غير مزعج بهدف إنشاء نموذج (Model) قادر على تصنيف أي بريد جديد إلى بريد مزعج وغير مزعج.



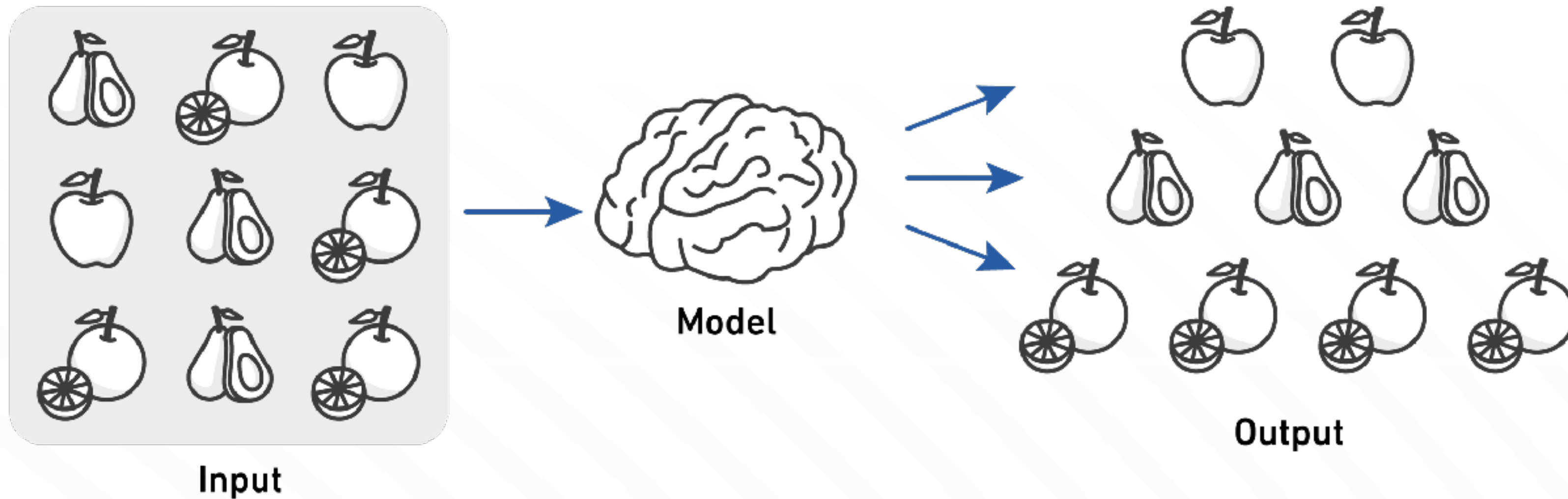
خوارزميات تعلم الآلة (Machine Learning Algorithms)

عادة ما تكون الخوارزميات المستخدمة في هذا النوع مصنفة لنوعين:
خوارزميات التصنيف (classification) مثل SVMs
خوارزميات الانحدار (Regression) مثل الانحدار الخطي (Linear Regression).

خوارزميات تعلم الآلة (Machine Learning Algorithms)

النوع الثاني (Unsupervised Learning)

يشير هذا النوع إلى العملية التي تتعلم فيها الآلات من مجموعات البيانات غير المصنفة بناءً على التشابه بين مجموعة البيانات. مثال: عندما نقوم بتدريب الآلة على تصنيف الفواكه، في هذه الحالة لا نخبر الآلة عن اسم الفاكهة، إذا كيف تقوم الآلة بالتصنيف؟ تقوم الآلة بالتصنيف بناءً على خصائص الفواكه مثل: اللون والحجم والشكل فمثلاً: إذا كان اللون أحمر يتم تصنيف الفاكهة إلى تفاح وهكذا. ومن أشهر الخوارزميات على هذا النوع: خوارزمية k-Means.





خوارزميات تعلم الآلة (Machine Learning Algorithms)

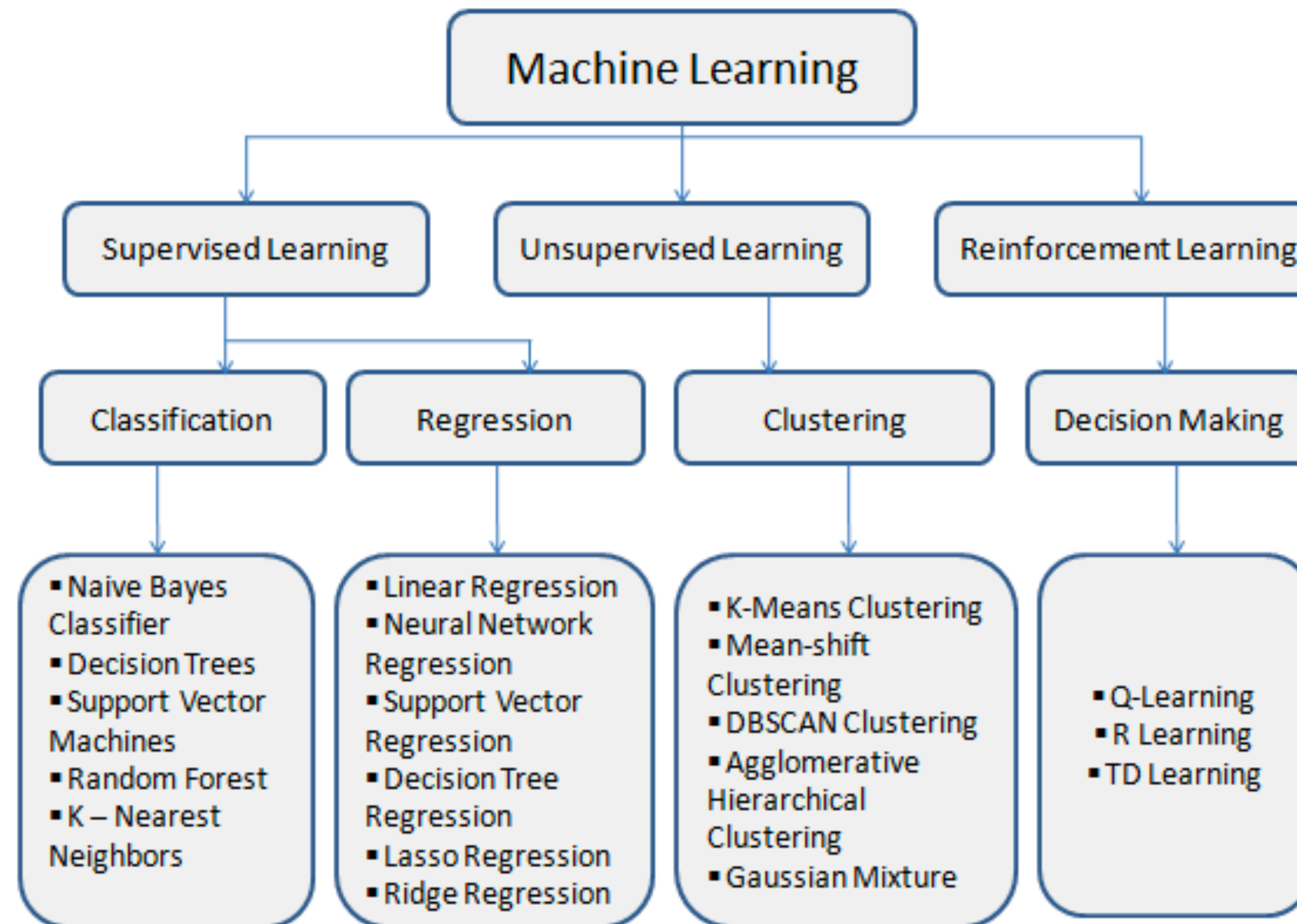
النوع الثالث (Semi-Supervised Learning)

هذا النوع من التعلم نحتاجه بالغالب عندما يكون لدينا مجموعات البيانات الكبيرة ولكن عدد قليل من هذه البيانات قد تم تصنيفها، لذا تقوم الآلة بالتعلم من كلا النوعين البيانات المصنفة والغير مصنفة.

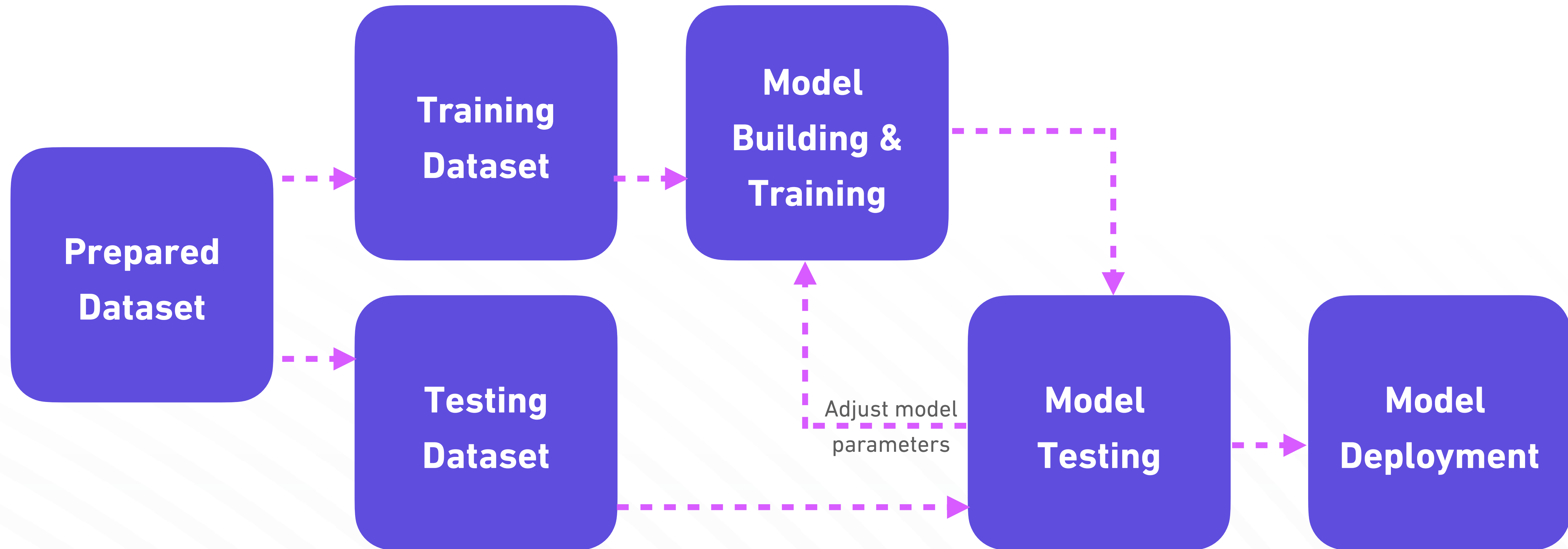
النوع الرابع (Reinforcement Learning)

يشير هذا النوع إلى العملية التي تتعلم فيها الآلات اتخاذ قرارات (actions) بناءً على البيئة الخارجية ثم مكافأة الآلة حتى نصل لتحقيق أقصى قدر من هدف معين.
مثال: بناء الرجل الآلي.

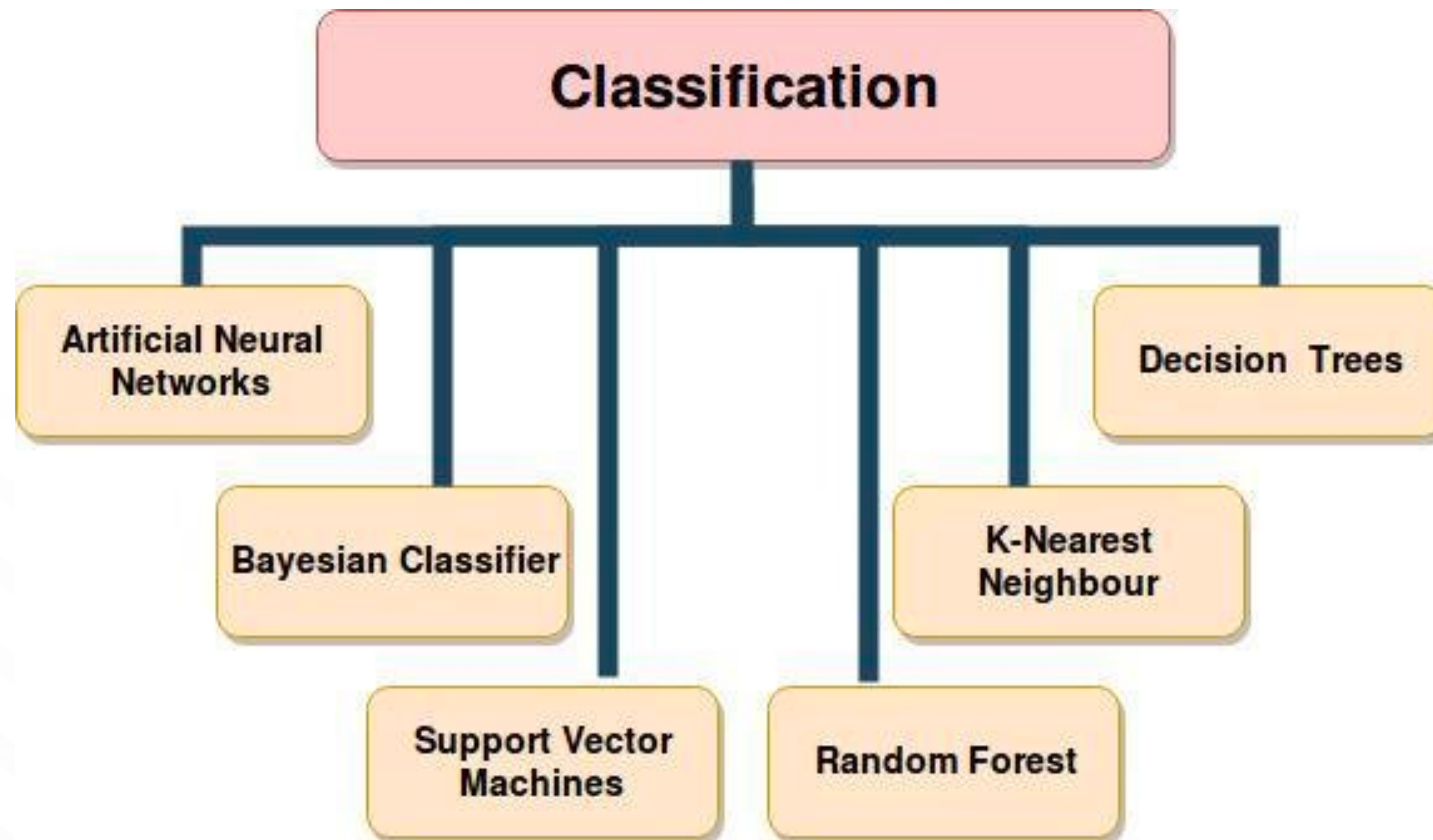
خوارزميات تعلم الآلة (Machine Learning Algorithms)



مراحل بناء نماذج تعلم الآلة (Machine Learning Models)



خوارزميات التصنيف (Classification Algorithms)



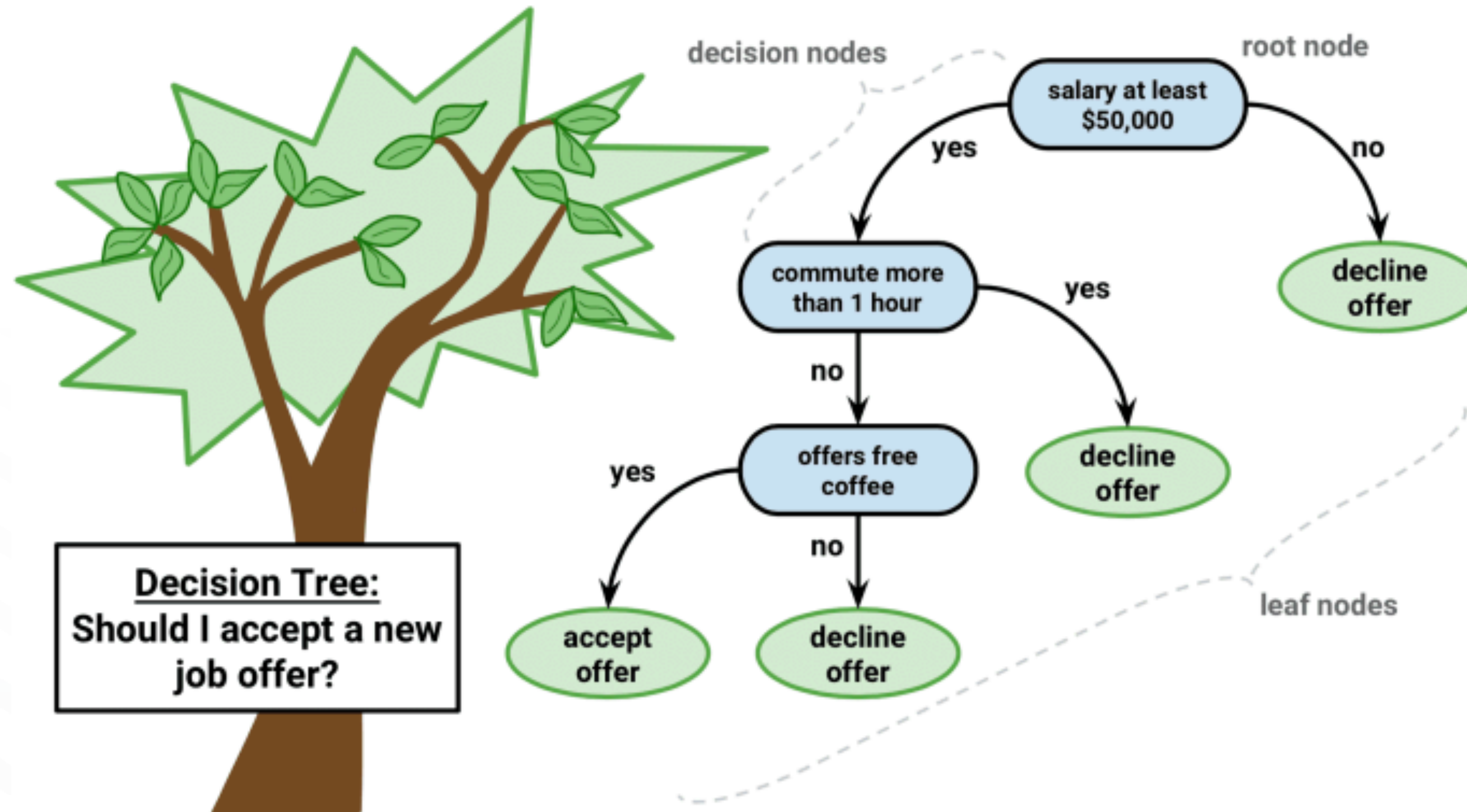


خوارزميات التصنيف (Classification Algorithms)

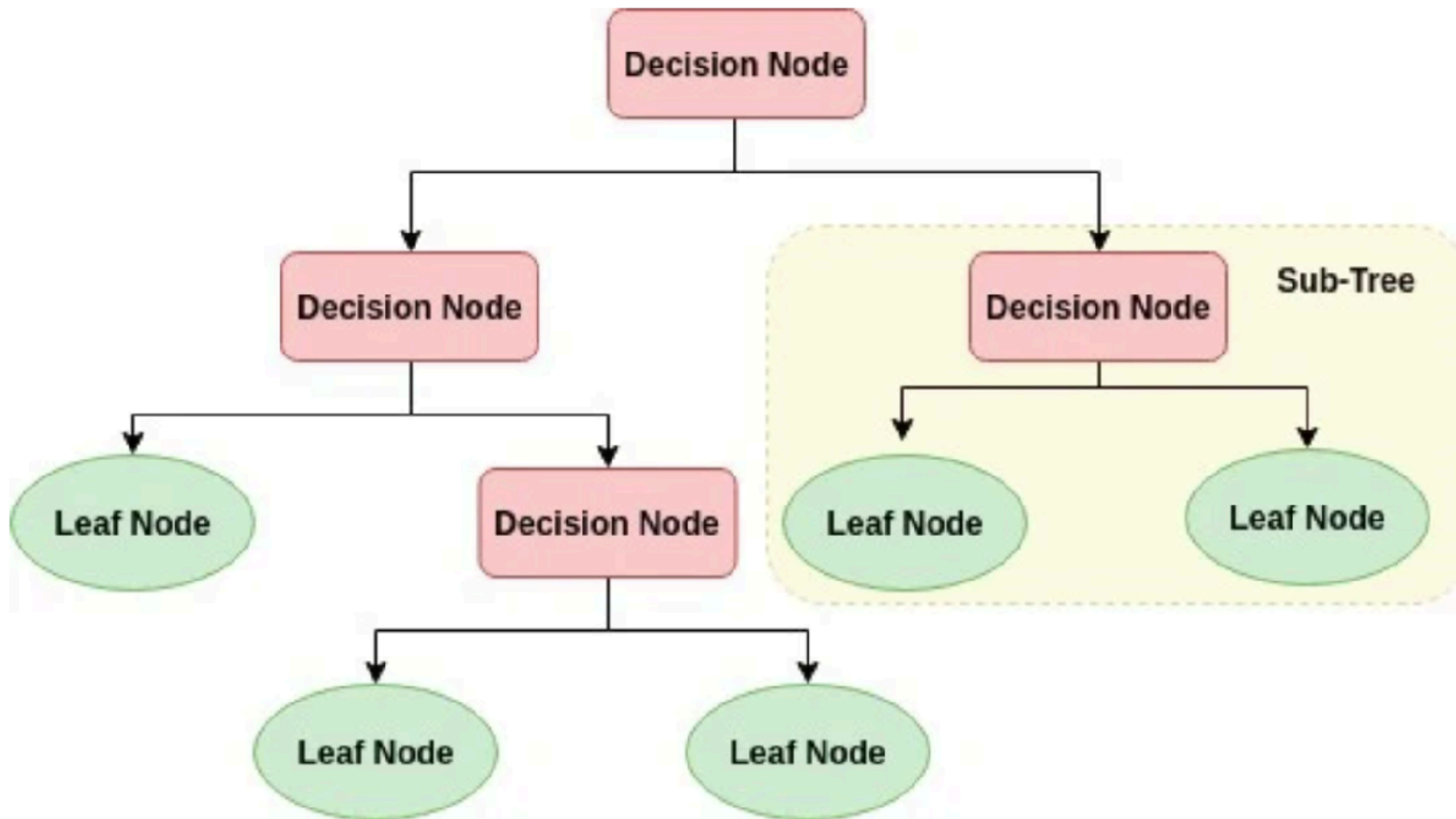
أمثلة على تطبيقات خوارزميات التصنيف

- عمليات (Credit/loan approval)
- التشخيص الطبي (if a tumor is cancerous or benign)
- كشف الاحتيال (Fraud detection: if a transaction is fraudulent)
- تصنيف صفحات الويب (Web page categorization: which category it is)

خوارزميات التصنيف (Decision Tree)



خوارزميات التصنيف (Decision Tree)



- Node في البداية تُسمى (Root).
- Decision Node تُسمى (Internal Node).
- Sub-Tree يُسمى (Branch).



خوارزميات التصنيف (Decision Tree)

X	Y	Z	Class
1	1	1	Sick
1	1	0	Sick
0	0	1	Healthy
1	0	0	Healthy

خوارزميات التصنيف (Decision Tree)

مقاييس السمات (Attribute Selection Measures)

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i),$$

• النوع الأول: Information Gain
Info(D) يسمى entropy

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}.$$

• النوع الثاني: Gain Ratio

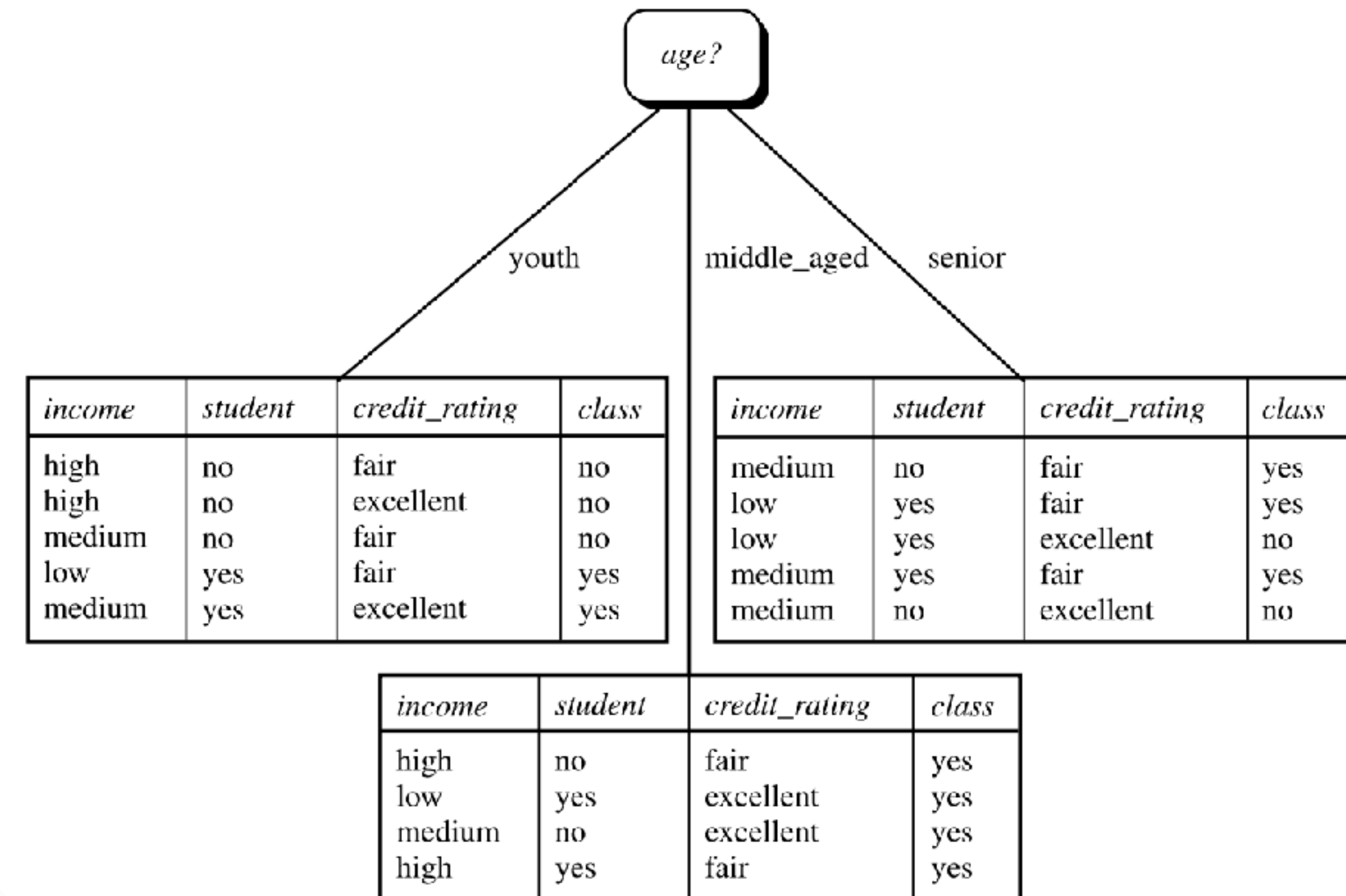
$$Gini(D) = 1 - \sum_{i=1}^m p_i^2,$$

• النوع الثالث: Gini Index

خوارزميات التصنيف (Decision Tree)

Class-Labeled Training Tuples from the *AllElectronics* Customer Database

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no



خوارزميات التصنيف (Decision Tree)

Class-Labeled Training Tuples from the *AllElectronics* Custom

RID	age	income	student	credit_rating	Class:
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

$$Info(D) = -\frac{9}{14} \log_2 \left(\frac{9}{14} \right) - \frac{5}{14} \log_2 \left(\frac{5}{14} \right) = 0.940 \text{ bits.}$$

$$\begin{aligned}
 Info_{age}(D) &= \frac{5}{14} \times \left(-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right) \\
 &\quad + \frac{4}{14} \times \left(-\frac{4}{4} \log_2 \frac{4}{4} \right) \\
 &\quad + \frac{5}{14} \times \left(-\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} \right) \\
 &= 0.694 \text{ bits.}
 \end{aligned}$$

$$\begin{aligned}
 Gain(age) &= Info(D) - Info_{age}(D) = \\
 &0.940 - 0.694 = 0.246 \text{ bits.}
 \end{aligned}$$

خوارزميات التصنيف (Decision Tree)

Class-Labeled Training Tuples from the *AllElectronics* Customer Database

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

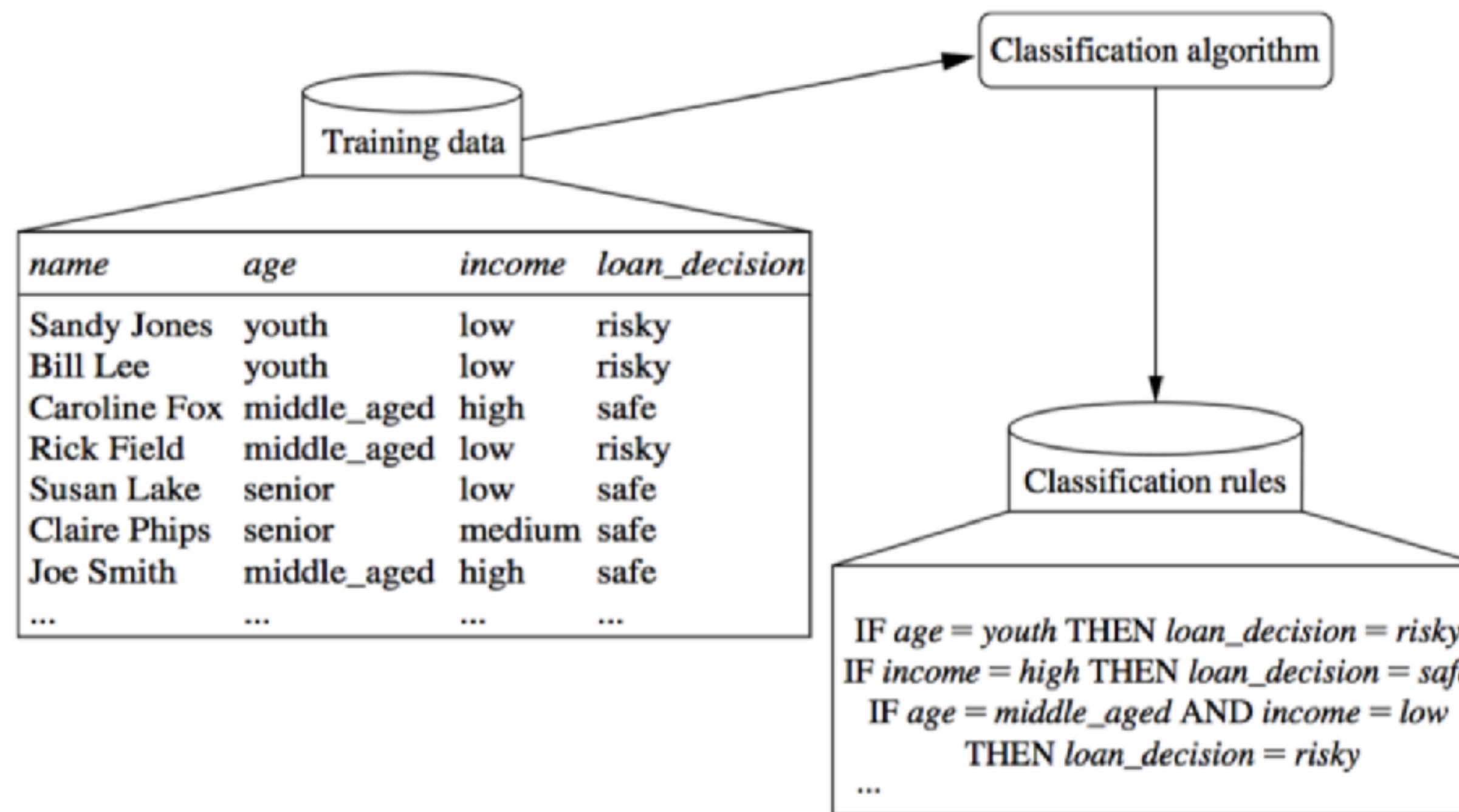
$Gain(income) = 0.029$ bits,

$Gain(student) = 0.151$ bits,

$Gain(credit_rating) = 0.048$ bits.

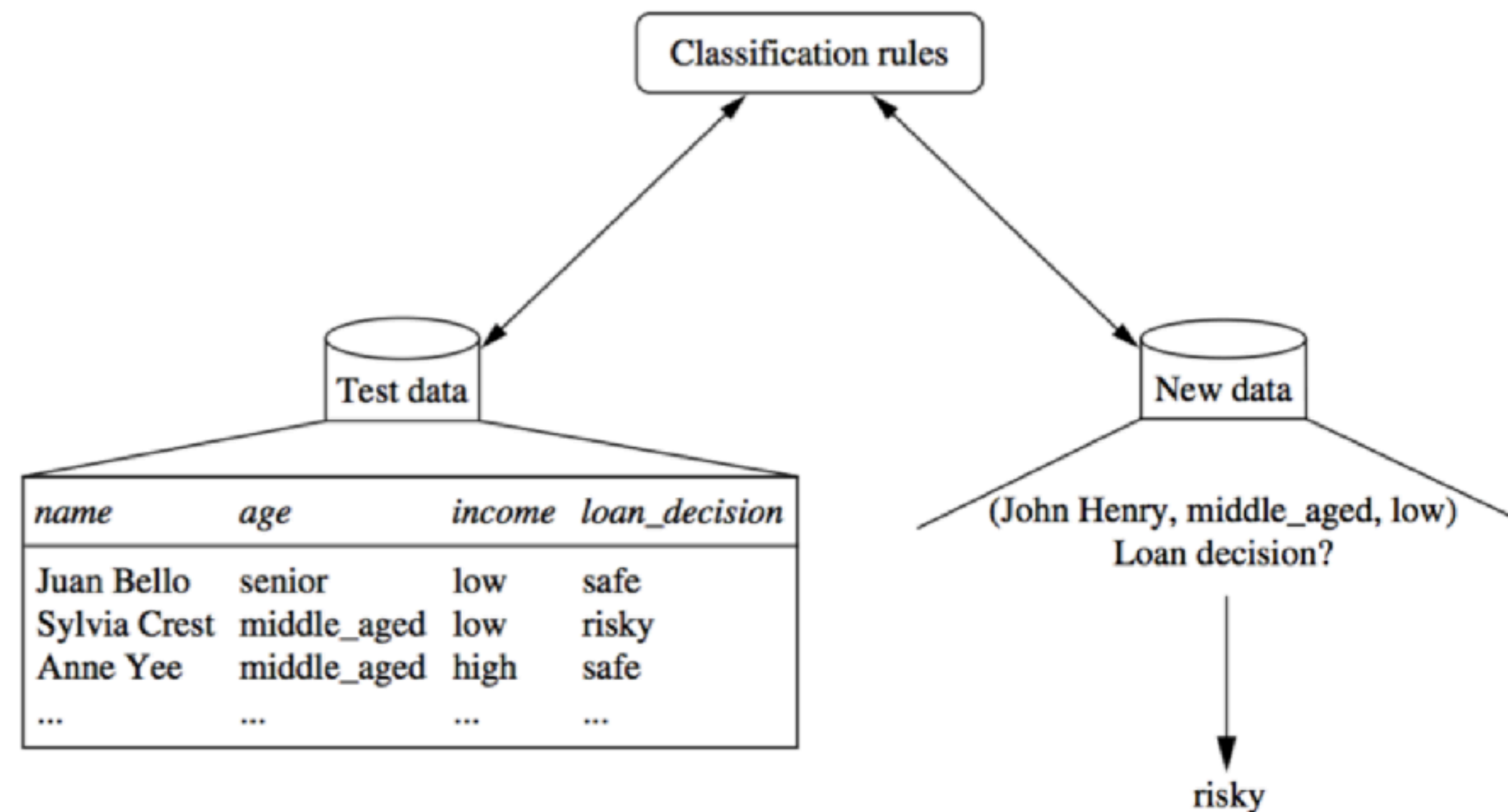
خوارزميات التصنيف (Decision Tree)

Example Step (1): Learning



خوارزميات التصنيف (Decision Tree)

Example Step (2): Classification





أمثلة لتطبيقات خوارزمية Decision Tree

مجال الرعاية الصحية

- يتم استخدام Decision Tree للتنبؤ بأمراض معينة.

التسويق

- لتحسين دقة الحملات الترويجية عن طريق مقارنة الاداء والمنتجات مع الشركات المنافسة.

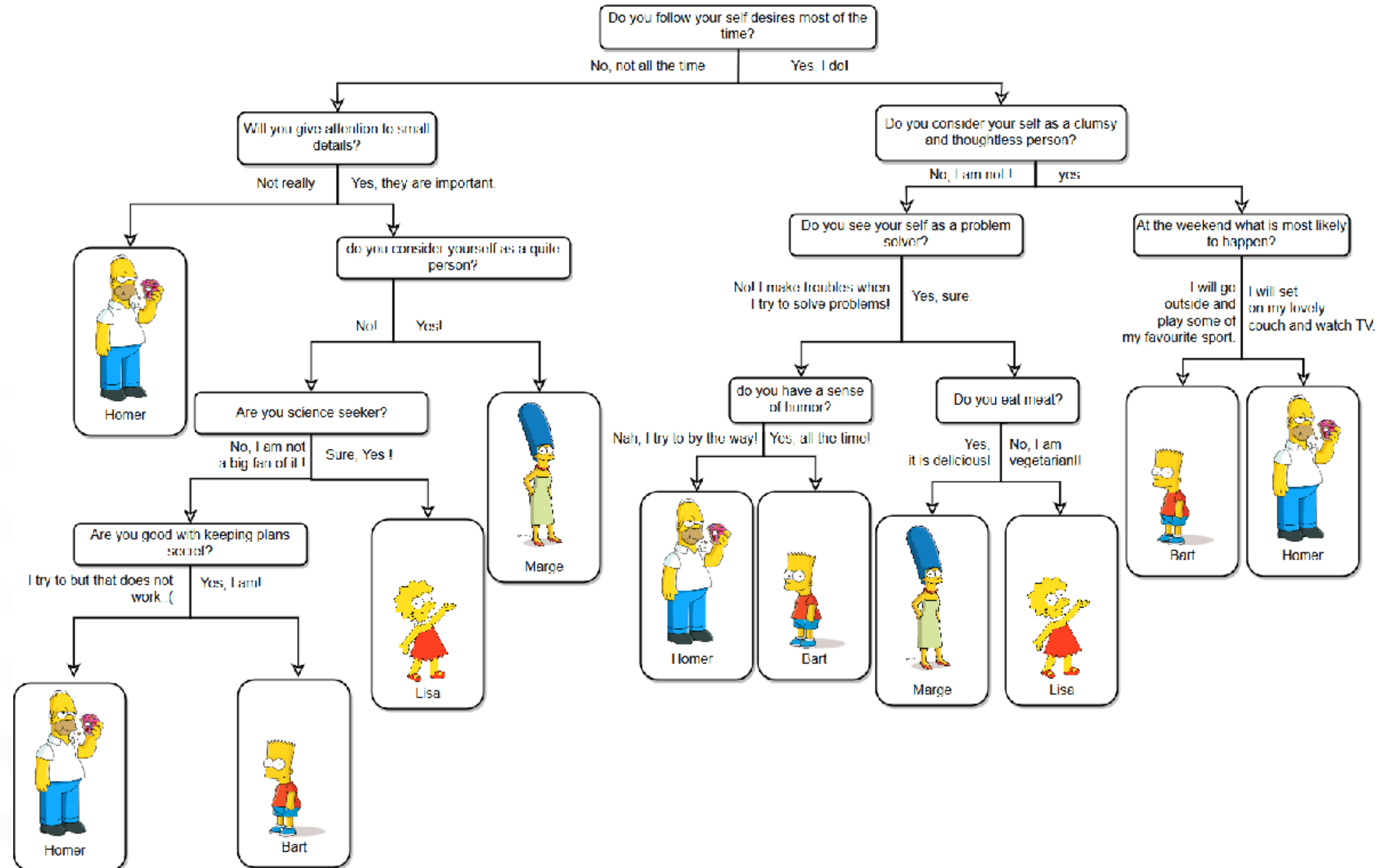
إدارة الأعمال

- استخراج البيانات من قواعد البيانات واستخدامها في تطبيقات إدارة الأعمال.

إدارة العملاء

- لقياس العلاقة بين احتياجات العملاء وتفضيلاتهم، أيضا في قياس رضا العملاء في التسوق الإلكتروني.




خوارزميات التصنيف (Decision Tree)



تقييم نماذج التصنيف (Classification Evaluation Metrics)

Regression	Classification	Recommender System
<ul style="list-style-type: none"> • Mean Absolute Error (MAE) • Root Mean Squared Error (RMSE) • R-Squared and Adjusted R-Squared 	<ul style="list-style-type: none"> • Recall • Precision • F1-Score • Accuracy • Area Under the Curve (AUC) 	<ul style="list-style-type: none"> • Mean Reciprocal Rank • Root Mean Squared Error (RMSE)

تقييم نماذج التصنيف (Confusion Matrix)

		DIAGNOSIS	
PATIENTS		Diagnosed Sick	Diagnosed Healthy
	Sick	True Positive	 False Negative
	Healthy	 False Positive	True Negative

False Positive Error is also referred to as the Type I error.

False Negative Error is also referred to as the Type II error.

تقييم نماذج التصنيف (Confusion Matrix)

PATIENTS	DIAGNOSIS	
	Diagnosed Sick	Diagnosed Healthy
Sick	True Positive	False Negative
Healthy	False Positive	True Negative

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

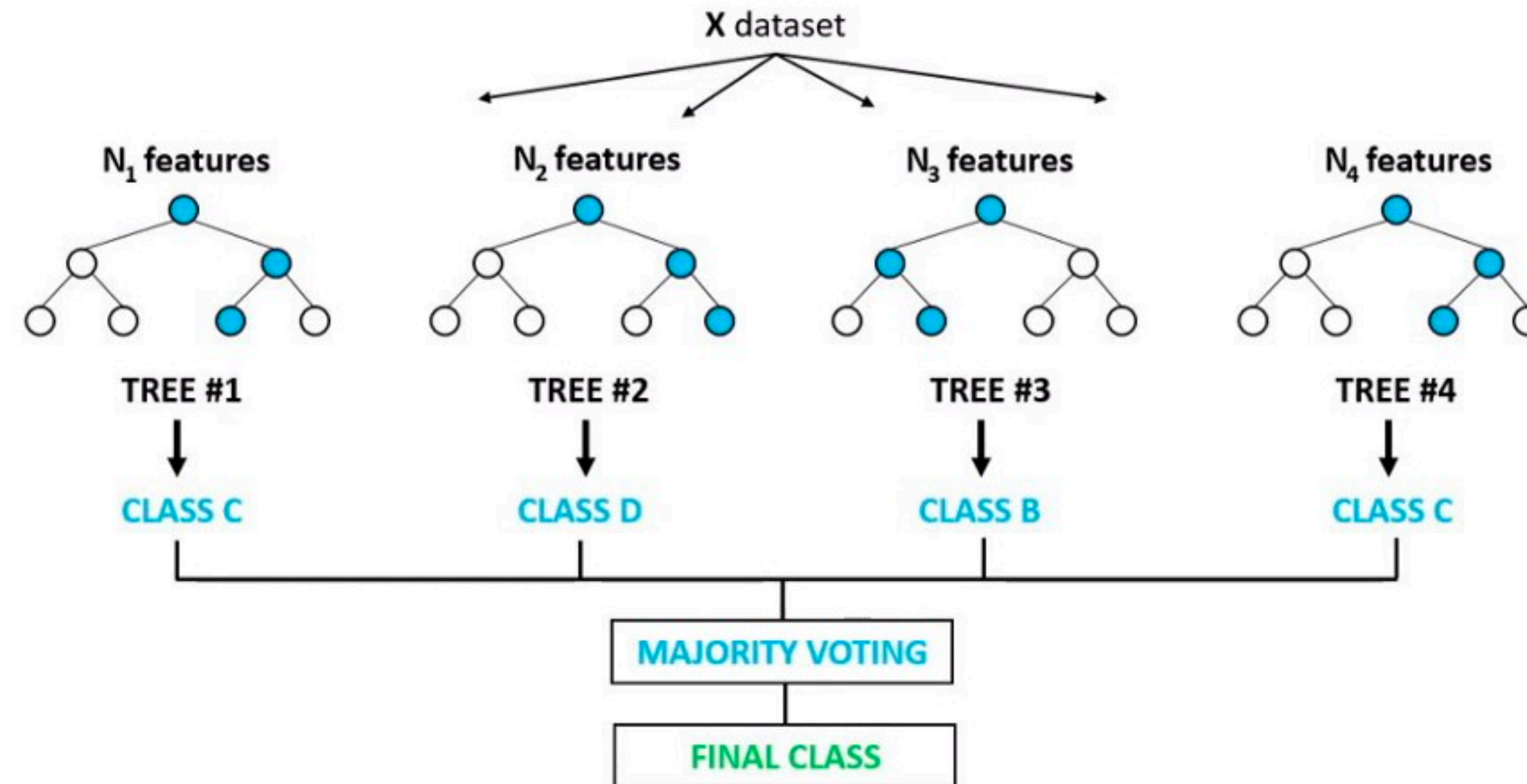
$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$

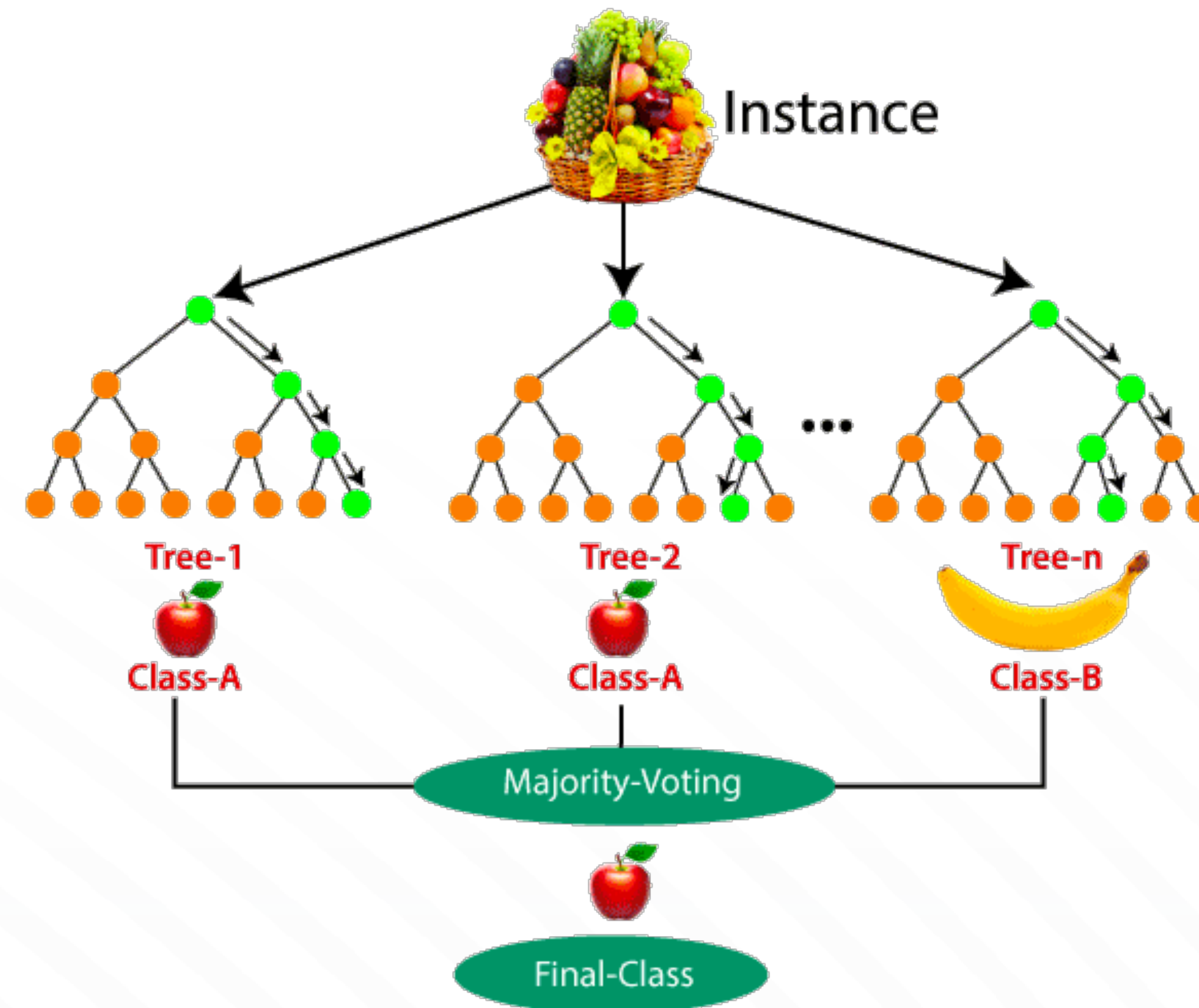
$$F1 \text{ score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

خوارزميات التصنيف (Random Forest)

Random Forest Classifier



خوارزميات التصنيف (Random Forest)



Resources

- Data Science: The Big Picture [<https://app.pluralsight.com/library/courses/data-science-big-picture/table-of-contents>].
- Introduction to Data Science [<https://link.springer.com/book/10.1007/978-3-319-50017-1>].
- Data Mining: Concepts and Techniques [<https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>].
- Deep-learning-nanodegree [<https://www.udacity.com/course/deep-learning-nanodegree--nd101>].