طرق تعلم الآلة

يوجد نوعان رئيسيان من طرق التعلم الآلي:

- **التعلم بإشراف :(Supervised Learning)** يعتمد على بيانات ذات مخرجات معروفة مسبقًا(Labels) ، والهدف هو بناء دالة تتنبأ بالمخرجات انطلاقًا من المدخلات. يطبّق أساسًا في **التصنيف** (Classification)لتحديد الفئات، و**الانحدار** (Regression)للتنبؤ بقيم مستمرة.

التصنيف:(Classification)

- **منح القروض البنكية:**

التنبؤ بما إذا كان العميل سيسدد القرض أو سيتعثر (فئتان: "سيسدد" / "لن يسدد").

- **تحليل الجدارة الائتمانية:(Credit Scoring)**

تصنيف العملاء إلى "جيد" أو "سيئ" من حيث المخاطر.

- **التسويق الموجه:**

تحديد ما إذا كان الزبون سيستجيب لحملة تسويقية معينة (نعم / لا).

2.الانحدار:(Regression)

- **التنبؤ بالناتج المحلي الإجمالي:(GDP)**

اعتمادًا على متغيرات مثل الاستثمار، الاستهلاك، والإنفاق الحكومي.

- **توقع أسعار الأسهم أو السلع:**

مثل التنبؤ بسعر النفط أو الذهب بناءً على بيانات العرض والطلب والعوامل الجيوسياسية.

- **تقدير الطلب على منتج:**

مثلاً: تقدير حجم المبيعات الشهرية لشركة وفقًا للإنفاق الإعلاني والدخل الفردي.

◆ الخلاصة:

- إذا كان الهدف **التنبؤ بفئة/قرار** ⟵نستخدم **التصنيف**.
- إذا كان الهدف **التنبؤ بقيمة عددية** ⟵نستخدم **الانحدار**.

- **التعلم بدون إشراف :(Unsupervised Learning)** لا يتوفر على مخرجات معنونة، بل يسعى لاكتشاف الأنماط والهياكل الخفية في البيانات. من أبرز تطبيقاته **التجميع (Clustering)** وتقليل الأبعاد (Dimensionality Reduction).

1.التجميع:(Clustering)

- **تقسيم العملاء إلى شرائح:(Customer Segmentation)**

تجميع العملاء وفقًا لسلوك الشراء أو القدرة الشرائية (مثل: عملاء ذوي إنفاق مرتفع، متوسط، منخفض).

- **تصنيف الأسواق:**

تجميع الأسواق العالمية أو المحلية حسب خصائصها الاقتصادية (مستقرة / ناشئة / متقلبة).

- **تحليل سلوك المستهلك:**

تقسيم المستهلكين حسب أنماط الشراء (شراء موسمي، شراء متكرر، شراء عشوائي).

2.تقليل الأبعاد:(Dimensionality Reduction)

- **تحليل المخاطر المالية:**

تبسيط مجموعة ضخمة من المؤشرات المالية (أسعار الفائدة، التضخم، سعر الصرف، إلخ) إلى أبعاد أقل لفهم الاتجاهات الرئيسية.

- **تصور البيانات الاقتصادية:**

عرض بيانات اقتصادية عالية التعقيد (مثل بيانات التجارة الدولية) في بعدين أو ثلاثة لتسهيل التفسير.

3.اكتشاف الأنماط والتطرفات:(Pattern & Anomaly Detection)

- **كشف الغش المالي أو المحاسبي:**

اكتشاف العمليات المشبوهة في المعاملات البنكية أو البيانات المالية.

- **تحليل الاستهلاك الطاقوي:**

رصد أنماط غير طبيعية في استهلاك الكهرباء أو الوقود التي قد تشير إلى تهريب أو سوء استخدام.

◆ الخلاصة:

- **التجميع** مفيد لتقسيم العملاء أو الأسواق.
- **تقليل الأبعاد** لتبسيط البيانات الاقتصادية المعقدة.
- **كشف الأنماط** لاكتشاف الغش أو السلوكيات غير المعتادة.

أولا: التعلم بإشراف:

التعلم الخاضع للإشراف (Supervised Learning) يعتمد على بيانات مصنفة (مدخلات x مع مخرجات y)، حيث يتعلم النموذج دالة $f(x) = y$ ليربط بينهما، مما يمكّنه من التنبؤ بمخرجات دقيقة لمدخلات جديدة استنادًا إلى تجارب سابقة.

1. التصنيف:

التصنيف **(Classification)**هو نوع من التعلم الخاضع للإشراف، يهدف إلى إيجاد علاقة بين البيانات المدخلة والتسميات (Labels) استنادًا إلى بيانات تدريب مصنفة، بحيث يمكن للنموذج التنبؤ بالفئة الصحيحة لبيانات جديدة.

ينقسم التصنيف إلى:

1.حسب عدد التسميات:

أ. التصنيف أحادي العلامة:(Single-Label Classification)

كل عينة ترتبط بفئة واحدة فقط.

◆ **مثال اقتصادي** :تصنيف شركة إلى قطاع اقتصادي واحد (زراعي، صناعي، خدمي).

ب. التصنيف الثنائي:(Binary Classification)

حالتان فقط (نعم/لا، إيجابي/سلبي).

◆ **مثال اقتصادي** :التنبؤ بما إذا كان العميل **سيوافق على قرض أم سيرفض** طلبه البنك (قبول / رفض).

ج. التصنيف متعدد الفئات:(Multi-Class Classification)

أكثر من فئتين.

- **مثال اقتصادي** :تصنيف المشاريع الاستثمارية إلى **منخفضة، متوسطة، عالية** المخاطر.

د. التصنيف متعدد العلامات:(Multi-Label Classification)

العينة قد تنتمي لأكثر من فئة في نفس الوقت.

- **مثال اقتصادي** :مقال اقتصادي قد يندرج تحت أكثر من تصنيف في آن واحد مثل:
  - "تجارة دولية"
  - "مالية"
  - "استثمار"

أو مثال آخر :**شركة ناشئة** يمكن أن تُصنف في وقت واحد ضمن "التكنولوجيا المالية "(FinTech) و"التجارة الإلكترونية."

1. **حسب عدد التسميات:**
   - تصنيف أحادي العلامة :(Single-Label) كل عينة مرتبطة بتسمية واحدة.
   - تصنيف متعدد العلامات :(Multi-Label) العينة قد تنتمي لعدة تسميات في نفس الوقت.

التصنيف أحادي العلامة:(Single-Label Classification)

- كل **عينة بيانات** يمكن أن تأخذ **تسمية واحدة فقط** من مجموعة الفئات.
- النموذج يتنبأ **بفئة وحيدة** لكل حالة.
  - مثال اقتصادي:

عميل بنك يتم تصنيفه إما "جيد" أو "سيئ" من حيث السداد. (لا يمكن أن يكون الاثنين معًا).

2 **التصنيف متعدد العلامات:(Multi-Label Classification)**

- كل **عينة بيانات** يمكن أن تنتمي إلى **عدة فئات في نفس الوقت.**
- النموذج يتنبأ بمجموعة تسميات للعينة.
  - مثال اقتصادي:

مقال اقتصادي قد ينتمي في الوقت نفسه إلى أكثر من مجال: "مالية"، "تجارة دولية"، و"استثمار."

أو شركة ناشئة تُصنف في آن واحد ضمن "**التكنولوجيا المالية "(FinTech)** و**"التجارة الإلكترونية.**"

- **الفرق الأساسي:**

- **Single-Label:** كل عينة ← تسمية واحدة فقط.
- **Multi-Label:** كل عينة ← قد تحتوي عدة تسميات معًا.

- التصنيف أحادي العلامة(Single-Label)
- التصنيف متعدد العلامات(Multi-Label)

◆ تخيل لدينا نفس العينة (عميل بنك):

) Single-Label .1أحادي العلامة:(

كل عميل يُعطى فئة واحدة فقط:

- العميل" ⟶ A جيد"
- العميل" ⟶ B سيئ"

⟵النتيجة دائمًا تصنيف واحد لكل عينة.

) Multi-Label .2متعدد العلامات:(

العميل نفسه قد ينتمي إلى أكثر من فئة في نفس الوقت:

- العميل" ⟶ A جيد في السداد" + "مستثمر نشط"
- العميل" ⟶ B سيئ في السداد" + "مستهلك كثيف"

⟵النتيجة عدة تسميات للعينة الواحدة.

2. **حسب عدد الفئات:**

- تصنيف ثنائي :(Binary) فئتان فقط (نعم/لا، ذكر/أنثى).
- تصنيف متعدد الفئات :(Multi-Class) أكثر من فئتين (تصنيف أنواع الزهور أو المنتجات)

أ. التصنيف أحادي العلامة:

التصنيف أحادي العلامة **(Single-Label Classification)**يعني أن كل عينة ترتبط بتسمية واحدة فقط، ويتنبأ النموذج بعلامة واحدة لكل حالة. وينقسم إلى:

✓ **تصنيف ثنائي :(Binary)** حالتان فقط.

✓ **تصنيف متعدد الفئات :(Multi-Class)** أكثر من فئتين.

ب. التصنيف الثنائي **(Binary Classification)**: هو أبسط أنواع التصنيف، حيث تنتمي كل عينة إلى إحدى فئتين فقط (إيجابية أو سلبية). مثال: اختبار الانتخابات (انتخب/امتنع).

ت. التصنيف متعدد الفئات:

التصنيف متعدد الفئات **(Multi-Class Classification)**يسمح بتمييز العينات إلى أكثر من فئتين دون حد محدد، مثل: تصنيف الأخبار حسب الموضوع، أو الكتب حسب التخصص، أو الحيوانات في الصور.

ث. التصنيف متعدد العلامات:

التصنيف متعدد العلامات **(Multi-Label Classification)**يعني أن العينة الواحدة قد ترتبط بعدة تسميات في آن واحد، مثل خبر تقني عن iPhone يرتبط بفئة "تكنولوجيا" و"علامة تجارية". يستخدم بكثرة في **معالجة اللغة الطبيعية** (تصنيف النصوص)، وكذلك في **استرجاع المعلومات، تشخيص الأمراض، والمعلوماتية الحيوية** .وهو تعميم للتصنيف أحادي العلامة، حيث يمكن ربط كل حالة بمجموعة من العلامات بدلاً من واحدة فقط.

أ. التصنيف أحادي العلامة:(Single-Label Classification)

كل عينة لها **تسمية واحدة فقط**.

◆ **مثال اقتصادي** :تصنيف مؤسسة إلى قطاع واحد فقط (زراعي، صناعي، خدمي).

ب. التصنيف الثنائي:(Binary Classification)

أبسط أنواع التصنيف :**حالتان فقط (نعم/لا أو إيجابي/سلبي)**.

- **مثال اقتصادي** :التنبؤ إذا كان العميل **سيوافق** على قرض أم **سيرفض** طلبه البنك (قبول / رفض).
- **مثال آخر** :التنبؤ إذا كان سهم شركة سيرتفع أو ينخفض غدًا.

ج. التصنيف متعدد الفئات:(Multi-Class Classification)

أكثر من فئتين، لكن العينة تأخذ **فئة واحدة فقط**.

- **مثال اقتصادي** :تصنيف المشاريع الاستثمارية حسب درجة المخاطرة :**منخفضة – متوسطة – عالية**.

2. الانحدار – regression

الانحدار **(Regression)**يختلف عن التصنيف في كونه يتنبأ بقيم عددية مستمرة، بينما التصنيف يتعامل مع فئات منفصلة. هو أسلوب إحصائي يحدد العلاقة بين **المتغيرات المستقلة** و **المتغير التابع**.

أمثلة تطبيقية:

- التنبؤ بدرجات الطلاب بناءً على عدد ساعات الدراسة.
- تقدير سعر سيارة مستعملة اعتمادًا على خصائصها (العلامة التجارية، السنة، المسافة المقطوعة).
- في السيارات ذاتية القيادة: التنبؤ بالزاوية المناسبة لعجلة القيادة استنادًا إلى بيانات أجهزة الاستشعار.

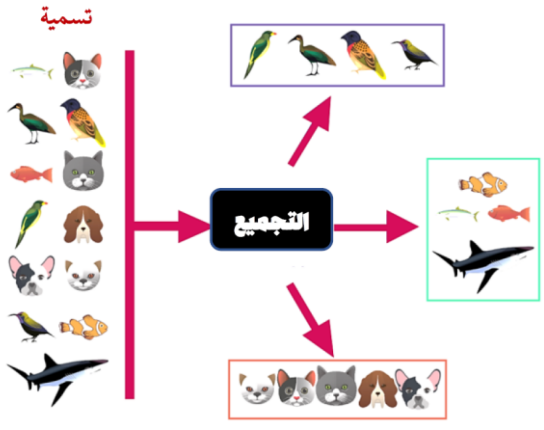
ثانيا/ التعلم بإشراف:

التعلم بدون إشراف **(Unsupervised Learning)**يُستخدم عند غياب البيانات المصنفة، حيث تقوم الخوارزمية بالبحث ذاتيًا عن **أنماط وهياكل خفية** داخل البيانات.

- ✓ لا يعرف النموذج مسبقًا ما هو "الصواب أو الخطأ."
- ✓ يعتمد على **تشابه البيانات واختلافها** لتجميعها أو تبسيطها.
- ✓ مفيد في **التحليل الاستكشافي**، مثل تقسيم المستهلكين إلى شرائح دون وجود تسميات سابقة

1. التجميع: Clustering

التجميع **(Clustering)**هو تقسيم البيانات إلى مجموعات (عناقيد) بحيث تتشابه العناصر داخل كل مجموعة وتختلف عن المجموعات الأخرى، مع تقليل المسافة بين نقاط البيانات داخل نفس المجموعة قدر الإمكان.



شكل 2_1. التجميع.

نظرًا لأن التجميع يتم بواسطة خوارزمية، فمن المحتمل أنك ستكون قادرًا على اكتشاف الارتباطات غير المعروفة سابقًا في البيانات التي يمكن أن تساعدك في مواجهة تحدي الأعمال من منظور جديد

2. تقليل الأبعاد:

تقليل الأبعاد **(Dimensionality Reduction)**هو تبسيط البيانات عبر خفض عدد المتغيرات (السمات) مع الحفاظ قدر الإمكان على المعلومات الأساسية، وذلك بتحويلها من فضاء عالي الأبعاد إلى فضاء أصغر دون فقدان الخصائص المهمة.

3. المقارنة بين التعلم بإشراف مع التعلم بدون إشراف

المقارنة بين **التعلم بإشراف** و**التعلم بدون إشراف** يمكن تبسيطها عبر مثال تعلم لغة جديدة:

أ. **التعلم بإشراف** :يشبه إعطاء الطفل قاموسًا يترجم الكلمات مباشرة، فيتعلم بسرعة لكنه يظل محدودًا في الفهم العميق (يعتمد على بيانات مصنفة جاهزة).

ب. **التعلم بدون إشراف** :يشبه إعطاء الطفل كتبًا باللغة الأجنبية ليكتشف القواعد والأنماط بنفسه. العملية أصعب وتستغرق وقتًا أطول، لكنها تمنحه فهماً أعمق وبنية لغوية أوسع (يستخرج الأنماط خفية من البيانات غير المصنفة).

التعلم بإشراف(Supervised Learning)

- الفكرة: لدينا بيانات مُسماة (Labels) نعرف فيها النتيجة مسبقًا، ونُعلم النموذج كيف يتنبأ بها.

أمثلة اقتصادية وإدارية:

الاقتصاد :التنبؤ بالناتج المحلي الإجمالي (GDP) لدولة اعتمادًا على بيانات مثل: الاستثمار، الصادرات، والاستهلاك .)انحدار .(Regression

التسيير المالي :التنبؤ بمدى احتمال **إفلاس الشركات** بالاعتماد على نسب مالية (ديون، سيولة، ربحية)) .تصنيف ثنائي .(Binary

إدارة الموارد البشرية :التنبؤ بما إذا كان موظف ما سيترك الشركة أم لا (Yes/No) بالاعتماد على بيانات مثل: الراتب، عدد ساعات العمل، الرضا الوظيفي.

التسويق :التنبؤ بقيمة مبيعات منتج معين اعتمادًا على ميزانية الإعلانات والموسمية (انحدار).

ثانياً :التعلم بدون إشراف(Unsupervised Learning)

- الفكرة: لا نملك بيانات مُسماة مسبقًا، النموذج يكتشف الأنماط والعلاقات المخفية بنفسه.

أمثلة اقتصادية وإدارية:

الاقتصاد :تجميع الدول إلى مجموعات (Clusters) بناءً على مؤشرات اقتصادية مثل التضخم، البطالة، والنمو، دون تحديد الفئات مسبقًا.

التسيير المالي :تجميع العملاء في البنوك حسب سلوكهم المالي (مدخرون، مقترضون، مستثمرون) دون وجود تسميات جاهزة.

إدارة الموارد البشرية :تحليل بيانات الموظفين لاكتشاف أنماط غير مرئية، مثل تقسيمهم إلى مجموعات حسب الأداء والتحفيز، من غير أن نحدد الفئات مسبقًا.

التسويق :تقسيم السوق (Market Segmentation) إلى مجموعات من المستهلكين (حسب الدخل، التفضيلات الشرائية، العمر) دون تسميات جاهزة مسبقًا.

- الفرق الأساسي:

بإشراف :عندنا **نتيجة معروفة مسبقًا**)مثل: ربح/خسارة، قبول/رفض.(

بدون إشراف :نحاول اكتشاف **أنماط وعلاقات جديدة** في البيانات (مثل: تقسيم العملاء، تجميع الدول).

4. لماذا التعلم بدون إشراف؟

فعال جدًا عند توفر بيانات مصنفة، حيث يقيس النموذج أداؤه بمقارنة المخرجات المتوقعة بالحقيقية، ويسعى لتقليل الخطأ. لكنه يواجه قيودًا:

- تكلفة عالية لوضع العلامات يدويًا.
- محدودية التعميم، إذ يتعلم فقط من البيانات التي تم تدريبه عليها (مناسب للذكاء الاصطناعي الضيق *Narrow AI*).

التعلم غير الخاضع للإشراف:(Unsupervised Learning)

لا يعتمد على بيانات معنونة، بل يتعلم البنية الأساسية للبيانات ويكشف أنماطًا خفية.

- مثال: تجميع الصور المتشابهة (كراسي معًا، قطط معًا).
- يقلل عبء وضع العلامات يدويًا.
- أكثر مرونة للتعامل مع بيانات غير معروفة أو متغيرة باستمرار.
- أقل كفاءة أحيانًا من التعلم بإشراف، لكنه مناسب للمشكلات المفتوحة ولتعميم المعرفة بشكل أوسع.

ثالثا: التعليم المعزز

التعلم المعزز **(Reinforcement Learning)**مستوحى من طريقة تعلم الحيوانات، حيث يتعلم "الوكيل" السلوك الأمثل من خلال **التفاعل مع البيئة** عبر التجربة والخطأ، فيُكافأ على الأفعال الصحيحة ويُعاقب على الخاطئة. الهدف هو تطوير **سياسة (Policy)** لاتخاذ قرارات تزيد من إجمالي المكافآت المتوقعة. يتميز بقدرته على التعلم في بيئات ديناميكية غير معروفة مسبقًا دون إشراف مباشر.

اختيار النموذج وتقييمه:

اختيار النموذج وتقييمه في التعلم الآلي يتم عبر خطوتين أساسيتين:

1. **اختيار النموذج:(Model Selection)**
   - يشمل تحديد **المعاملات الفائقة (Hyperparameters)** مثل عدد الطبقات في الشبكة العصبية أو عدد الأشجار في.Random Forest
   - أو اختيار الخوارزمية الأنسب من بين عدة خوارزميات.
2. **تقييم النموذج:(Model Evaluation)**
   - الهدف هو تقدير خطأ النموذج على بيانات جديدة غير مرئية (خطأ التعميم).
   - لا يكفي قياس الأداء على بيانات التدريب فقط لتفادي مشكلة الحفظ.(Overfitting)

طرق التقييم:

- عند توفر بيانات كافية: تقسيمها إلى **تدريب / تحقق / (Validation) اختبار**.

- عند قلة البيانات: تقسيمها إلى **تدريب / اختبار** مع استخدام **التحقق المتقاطع.(Cross-Validation)**

القاعدة الأساسية: يجب أن تكون بيانات التدريب والتحقق والاختبار مأخوذة من نفس التوزيع لضمان دقة التقدير.

طرق التقييم:

في حين أن تدريب النموذج هو خطوة أساسية، فإن كيفية تعميم النموذج على البيانات غير المرئية هي جانب مهم بنفس القدر يجب مراعاته بعد تصميم أي نموذج للتعلم الآلي. يجب التأكد من أن النموذج فعال حقًا وأن نتائج تنبؤاته يمكن الوثوق بها.

يمكن تدريب خوارزمية التصنيف على مجموعة بيانات محددة مع مجموعة فريدة من المعاملات التي يمكن أن تخلق حدود قرار تتناسب مع البيانات. نتيجة تلك الخوارزمية المعينة لا تعتمد فقط على المعاملات المتوفرة لتدريب النموذج، ولكن أيضًا على نوع بيانات التدريب. إذا كانت بيانات التدريب تحتوي على تباين بسيط أو كانت البيانات موحدة، فقد يؤدي النموذج إلى overfitting وإنتاج نتائج متحيزة على البيانات غير المرئية. لذلك، يتم استخدام طرق مثل التحقق المتبادل لتقليل overfittingفي التحقق من الصحة. التحقق من الصحة المتبادل هو أسلوب يقسم مجموعة التدريب الرئيسية إلى مجموعتين من التدريب والبيانات التجريبية (التحقق من الصحة). الطريقة الأكثر شيوعًا للتحقق المتبادل هي التحقق متعدد الأجزاء K- fold Validation، والذي يقسم مجموعة البيانات الرئيسية إلى أجزاء k متساوية الحجم. k هو رقم يحدده المستخدم ، وعادة ما يتم تحديد 5 أو 10.في هذه الطريقة،في كل مرة يتم استخدام واحدة من المجموعات الفرعية k كمجموعة تحقق (اختبار) ويتم تجميع المجموعة الفرعية k-1 معًا لتشكيل مجموعة تدريبية. للحصول على الكفاءة الكلية للنموذج، يتم حساب متوسط تقدير الخطأفي جميع التجارب.

معايير التقييم:

لحساب معايير التقييم لنموذج التصنيف، نحتاج إلى أربع مجموعات من الفئة الحقيقية وفئة التنبؤ مع العناوين، الموجبة الحقيقية، الموجبة الخاطئة، السلبية الحقيقية والسلبية الخاطئة، والتي يمكن تمثيلها في مصفوفة الارتباك Confusion ( ) Matrixجدول 2 – 1(:

•موجب حقيقي :) TP (على سبيل المثال ، عندما كانت القيمة الفعلية للفئة "نعم" ، توقع النموذج أيضًا "نعم")أي توقع صحيح.)

•موجب خاط ئ : (FP) على سبيل المثال ، عندما كانت القيمة الفعلية للفئة "لا" لكن النموذج توقع "نعم")أي توقع خاطئ.)

•منفى خاطئ : (FN) على سبيل المثال ، عندما تكون القيمة الفعلية للفئة " Yes " ، لكن النموذج توقع "لا")أي توقع خاطئ.)

•منفى حقيقي : (TN) على سبيل المثال ، عندما تكون القيمة الفعلية للفئة "لا" وتوقع النموذج "لا")أي ، كان التوقع صحيحًا.)

جدول 2_1. مصفوفة الارتباك

		الفئة المتوقعة	
		مثبت	منفى
الفئة الحقيقي	مثبت	موجب حقيقي (TP)	منفى خاطئ (FN)
	منفى	موجب خاطئ (FP)	منفى حقيقي (TN)

المعيار الأكثر شيوعًا الذي يتم الحصول عليه من مصفوفة الارتباك هو دقتها) **accuracy** (أو عكسها: خطأ التنبؤ (prediction error :

$$\text{الدقة} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$\text{خطأ التنبؤ} = 1 - \text{الدقة}$$

الدقة هي نسبة عدد التنبؤات الصحيحة إلى العدد الإجمالي لعينات الإدخال. عندما يتعين علينا تقييم نموذج ما، فإننا غالبًا ما نستخدم معدلات الخطأ والدقة، ولكن ما نركز عليه بشكل أساسي هو مدى موثوقية نموذجنا، وكيف يعمل على مجموعة بيانات مختلفة (قابلية التعميم(ومدى مرونته. لا شك أن الدقة معيار مهم للغاية يجب أخذه الاعتبار، لكنها لا تقدم دائمًا صورة كاملة لأداء النموذج.