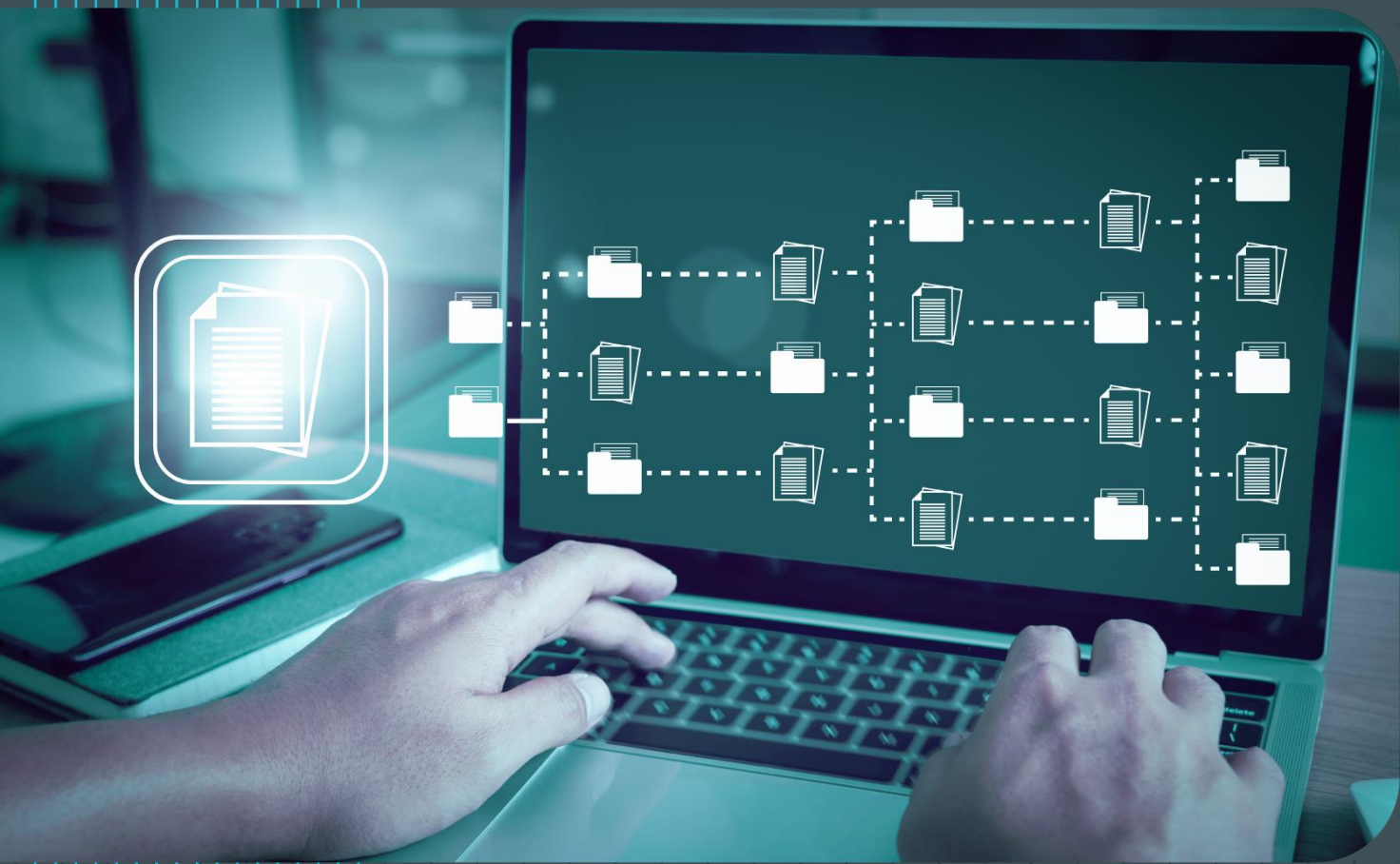


المركز الوطني
للأمن السيبراني
National Cyber
Security Center

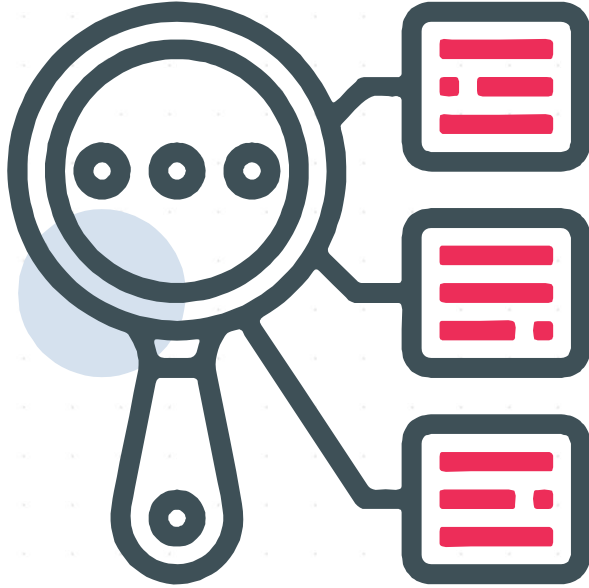


تصنيف البيانات

المحاور

- تعريف تصنيف البيانات
- أهمية تصنيف البيانات
- تقنيات تصنيف البيانات
- التعامل مع البيانات غير المتوازنة
- تطبيقات عملية
- الخاتمة

تعريف تصنيف البيانات



مفهوم تصنيف البيانات:

- تصنيف البيانات هي عملية تنظيم وترتيب البيانات في مجموعات أو فئات بناءً على خصائصها المشتركة.

هدف تصنيف البيانات:

- الهدف الرئيسي من تصنيف البيانات هو تطبيق نموذج أو قاعدة لتحديد الفئة التي تنتمي إليها كل نقطة في مجموعة البيانات.
- يتيح تصنيف البيانات للحاسب والأنظمة الذكية التعرف على الأنماط واتخاذ قرارات استنادًا إلى البيانات المتاحة.

أهمية تصنيف البيانات

تصنيف البيانات له أهمية كبيرة نظرًا لتأثيره الإيجابي على فهم البيانات واتخاذ قرارات أفضل. وترجع أهميته إلى:

- **تنظيم البيانات:** تنظيم وترتيب البيانات في هياكل منهجية، مما يجعلها أكثر فهمًا سهولة في الوصول و الإسترجاع.
- **استخدام البيانات لاتخاذ القرارات:** يمكن أصحاب القرار من فهم البيانات بشكل أفضل واتخاذ قرارات استراتيجية أكثر دقة.
- **تدريب النماذج والذكاء الاصطناعي:** حيث يتعلم النموذج من البيانات المصنفة مسبقًا ليتمكن من التنبؤ بتصنيف البيانات الجديدة.
- **تحسين دقة التحليل:** يتيح التصنيف فصل البيانات إلى فئات مختلفة، مما يسهم في تحديد الأنماط والاتجاهات بشكل أفضل.
- **تطوير تطبيقات متقدمة:** يُستخدم تصنيف البيانات في بناء تطبيقات متقدمة مثل أنظمة التوصية، وتصنيف الصور، وتحليل النصوص، والتشخيص الطبي، مما يجعل البيانات قابلة للاستفادة في مجموعة واسعة من التطبيقات التكنولوجية.

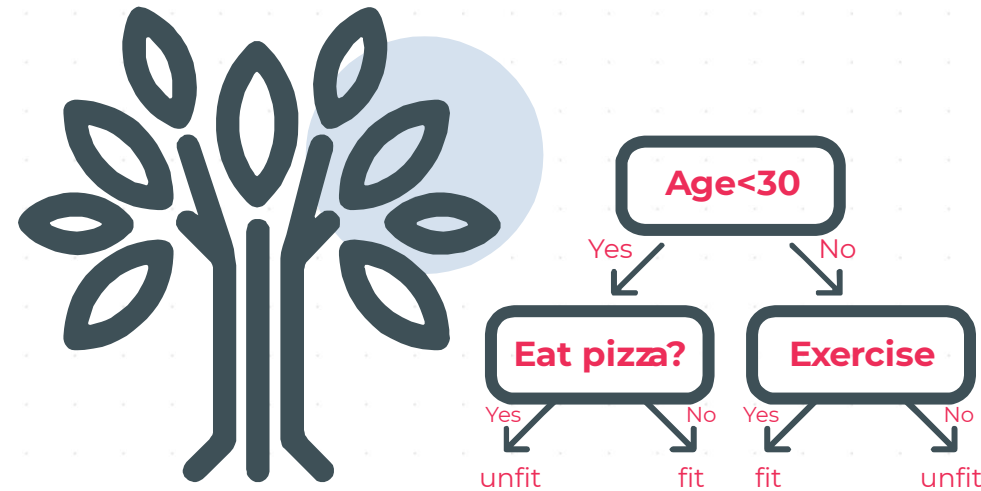
تقنيات تصنيف البيانات

تقنيات تصنيف البيانات تشمل مجموعة متنوعة من الأساليب والخوارزميات التي تستخدم لتحديد التصنيف الصحيح للبيانات. ومن بعض التقنيات الشائعة في تصنيف البيانات:

شجرة القرار (Decision Trees):

تعتمد على هيكل شجري يمثل قرارات تتخذ بناءً على سمات معينة. يتم تقسيم البيانات تدريجيًا بناءً على مجموعة من الأسئلة القرارية.

Is person fit or unfit ?



تقنيات تصنيف البيانات

شجرة القرار (Decision Trees) :

لنفترض أن لدينا مجموعة بيانات صغيرة تتعلق بتصنيف الفواكه إلى "تفاح" أو "برتقال". السمات المتاحة هي "اللون" و "القطر".

1- البيانات:

لون	قطر	تصنيف
أحمر	7	تفاح
أحمر	6	تفاح
أخضر	8	برتقال
أخضر	7	برتقال
أحمر	6.5	تفاح

2- بناء شجرة القرار:

الهدف: تصنيف الفاكهة (تفاح أو برتقال) استنادًا إلى اللون والقطر.

الخطوة 1: اختيار السمة الأفضل للتقسيم.

في هذا المثال، يمكن أن يكون اللون السمة الأفضل.

الخطوة 2: تقسيم البيانات استنادًا إلى اللون.

إذا كان اللون أحمر، فهي تفاح.

إذا كان اللون أخضر، ننظر إلى القطر.

إذا كان القطر أكبر من 7.5 سم، فهو برتقال.

إذا كان القطر أقل من 7.5 سم، فهو تفاح.

3- التصنيف:

يمكن استخدام شجرة القرار لتصنيف فاكهة جديدة بناءً على اللون والقطر.

على سبيل المثال، فاكهة حمراء بقطر 6.8 سم ستُصنف كـ "تفاح".

تقنيات تصنيف البيانات

التصنيف اللوجستي (Logistic Regression) :

يستخدم لتصنيف البيانات بين فئتين. يعتمد على اللوغاريتم لتقدير الاحتمالات واتخاذ القرار بناءً على القيم المعيارية. مثال: لنفترض أن لدينا مجموعة بيانات صغيرة لتصنيف طلاب إما "ناجحين" أو "راسبين" استنادًا إلى عدد ساعات الدراسة.

1- البيانات:

ساعة دراسية	نتيجة
2	راسب
4	راسب
6	ناجح
8	ناجح
10	ناجح

2- تدريب النموذج:

يستخدم نموذج التصنيف اللوجستي لتحديد فرص النجاح بناءً على عدد ساعات الدراسة.

3- التصنيف:

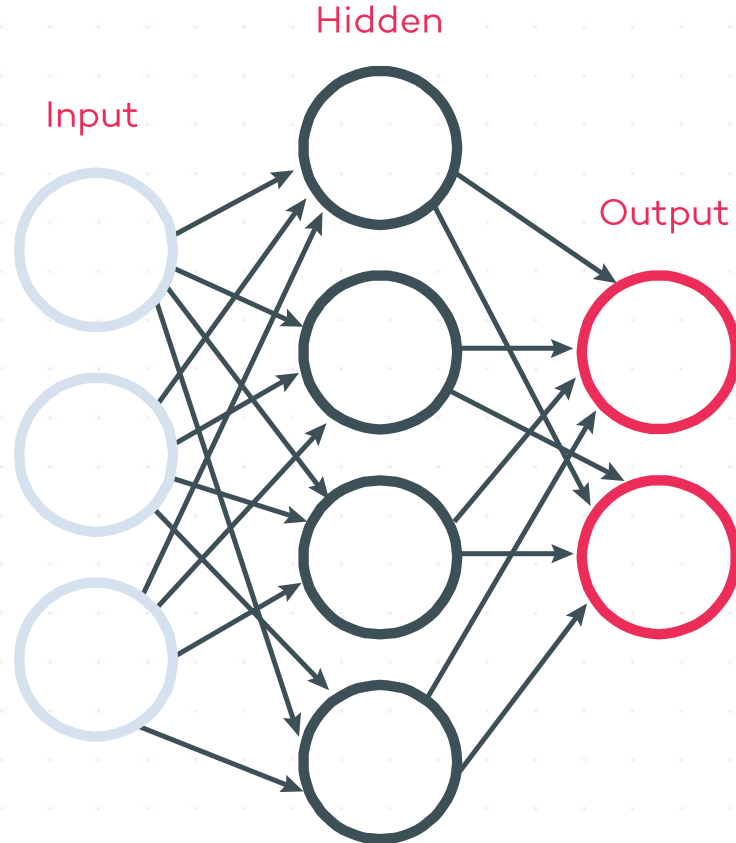
إذا كانت ساعات الدراسة أقل من 5، يكون التصنيف "راسب".

إذا كانت ساعات الدراسة 5 أو أكثر، يكون التصنيف "ناجح".

4- تصنيف متقدم للبيانات:

إذا كان الطالب قضى 7 ساعات في الدراسة، يتوقع أن يتم تصنيفه كـ "ناجح".

تقنيات تصنيف البيانات



الشبكات العصبية (Neural Networks):

تستند إلى تقليد الشبكة العصبية في الدماغ البشري. تتألف من طبقات من الوحدات الصناعية (العقد) تعمل على تعلم وفهم الأنماط المعقدة.

التعامل مع البيانات غير المتوازنة



التعامل مع البيانات غير المتوازنة يشير إلى الحالة التي تكون فيها فئات البيانات غير متوازنة بشكل كبير، أي أن هناك اختلافاً كبيراً في عدد السجلات بين الفئات المختلفة. عندما تكون هناك فجوة كبيرة في توازن الفئات، يمكن أن تؤثر هذه الحالة على أداء نماذج التصنيف. و من طرق التعامل مع هذا النوع من البيانات:

استخدام أساليب التيار الزائد (Oversampling) :

زيادة عدد السجلات في الفئة القليلة عن طريق تكرار السجلات أو إنشاء سجلات جديدة مشابهة. يهدف ذلك إلى تحسين توازن الفئات.

التعامل مع البيانات غير المتوازنة

- استخدام أساليب التيار الناقص (Undersampling):

تقليل عدد السجلات في الفئة الكثيرة لتحسين التوازن. ولكن يجب مراعاة عدم فقدان المعلومات التي قد يتسبب فيها هذا النهج.



- تقديم إشارات ذات أهمية أكبر:

يمكن تقديم معلومات إضافية أو تعزيز أهمية الفئة القليلة لتحسين الأداء.



- استخدام أساليب الإفراج الذاتي (Self-paced Learning)

تعتمد على مفهوم الإفراج الذاتي، حيث يتم تصنيف النماذج السهلة أولاً قبل التركيز على النماذج الأكثر تعقيداً.



تطبيقات عملية



يستخدم تصنيف البيانات في مجموعة واسعة من التطبيقات، ومن بينها:

- تصنيف البريد الإلكتروني (Spam Detection):
يتم استخدام تقنيات التصنيف لتحديد ما إذا كانت رسائل البريد الإلكتروني تُعتبر "غير مرغوبة". كما يمكن للنماذج أن تتعلم من السلوكيات والمحتوى لتحديد الرسائل المشبوهة.

تطبيقات عملية

- **تصنيف المستندات:**

يُستخدم تصنيف البيانات في فهم محتوى المستندات وتصنيفها إلى فئات مختلفة، مثل تصنيف المقالات الإخبارية أو تصنيف الوثائق القانونية.



- **التصنيف الطبي (Medical Diagnosis):**

يمكن استخدام تقنيات التصنيف لتحديد تشخيص الأمراض بناءً على البيانات الطبية، مثل صور الأشعة أو نتائج الاختبارات.



تطبيقات عملية

- تصنيف المشتريات عبر الإنترنت (E-commerce Product Classification):

يُستخدم التصنيف لتصنيف المنتجات في المتاجر عبر الإنترنت إلى فئات مثل الإلكترونيات، الملابس، أو المنتجات الغذائية.



- تصنيف الصور (Image Classification):

يمكن استخدام تقنيات التصنيف لتحديد محتوى الصور وتصنيفها، مثل تصنيف الحيوانات في الصور أو التمييز بين وجوه الأشخاص.



تطبيقات عملية

• تصنيف المستخدمين عبر الإنترنت (User Behavior Classification):

يمكن استخدام تقنيات التصنيف لفهم سلوك المستخدمين عبر الإنترنت وتصنيفهم إلى فئات مثل "مهتم بالرياضة" أو "مهتم بالتكنولوجيا".



• تصنيف الرسائل الاجتماعية (Social Media Content Classification):

يمكن استخدام تقنيات التصنيف لتصنيف مشاركات وتغريدات وسائل التواصل الاجتماعي إلى فئات مثل "إيجابي" أو "سلبي" أو "إعلان"، وذلك لتحليل ردود الفعل والاتجاهات.



الخاتمة



- عملية تصنيف البيانات هي عملية ذات أهمية كبيرة لأنها تمثل الأساس الرئيسي للاستفادة من البيانات بشكل فعال في مجالات عديدة.
- تقنيات تصنيف البيانات تشمل مجموعة متنوعة من الأساليب والخوارزميات مثل شجرة القرار و الشبكات العصبية و غيرها.
- هناك العديد من التطبيقات الحية لتصنيف البيانات مثل تصنيف رسائل البريد الإلكتروني و محتوى مواقع التواصل الاجتماعي.

نتمنى لكم السلامة في

العالم
الرقمي.