



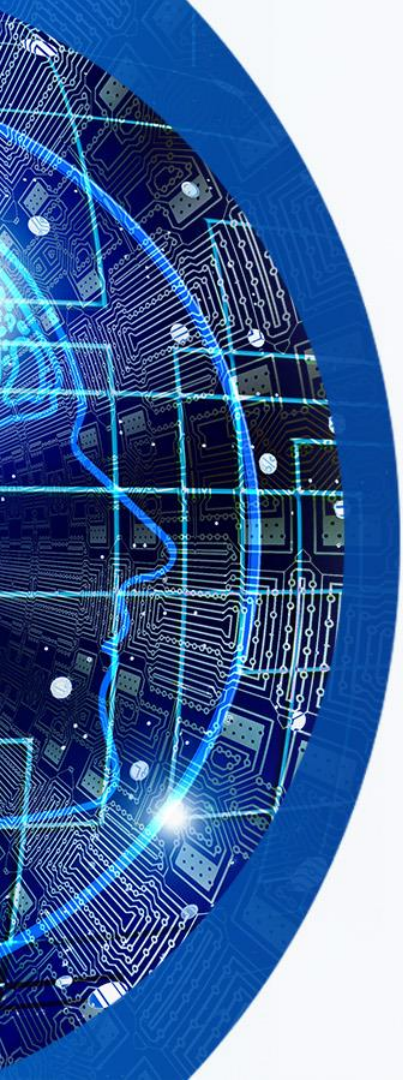
علم البيانات و تعلم الالة
Data science and Machine learning

علم البيانات



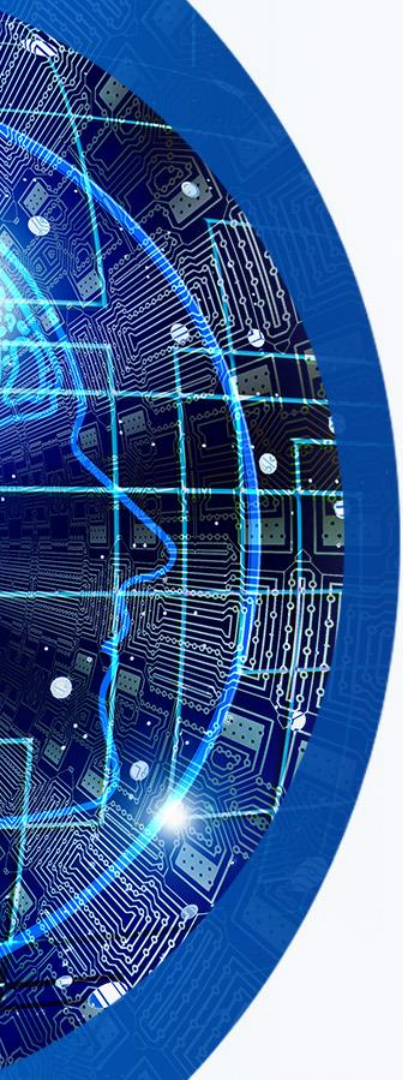
مقدمة

• تعد البيانات اليوم أداة ووقودًا للباحثين واصحاب القرار و للشركات لاكتساب رؤى مهمة وتحسين الأداء . و قد سيطر علم البيانات على كل المجالات تقريبًا في العالم.



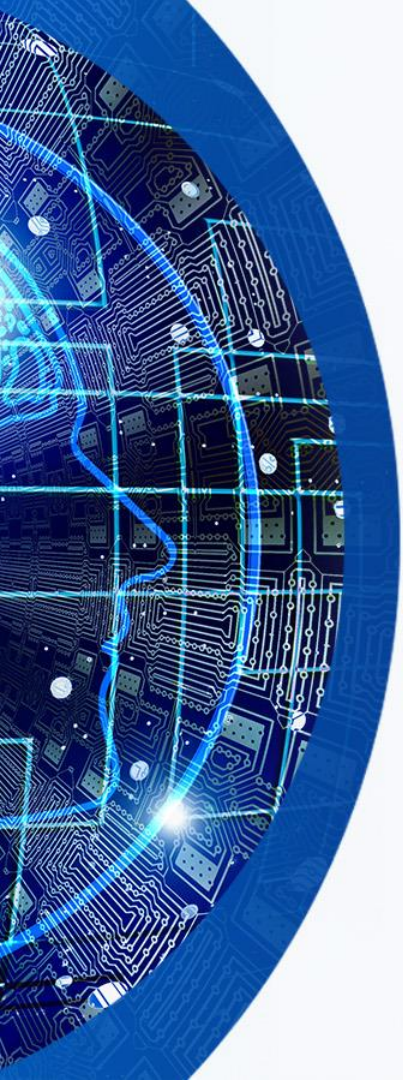
ما هو علم البيانات؟

- علم البيانات هو مجال بحث يكتشف من خلال مجموعة من الأدوات والخوارزميات ومبادئ التعلم الآلي المختلفة و كذلك التنقيب عن النصوص باستخدام الرياضيات والإحصاءات لمحاولة ايجاد الأنماط المخفية للبيانات الخام.



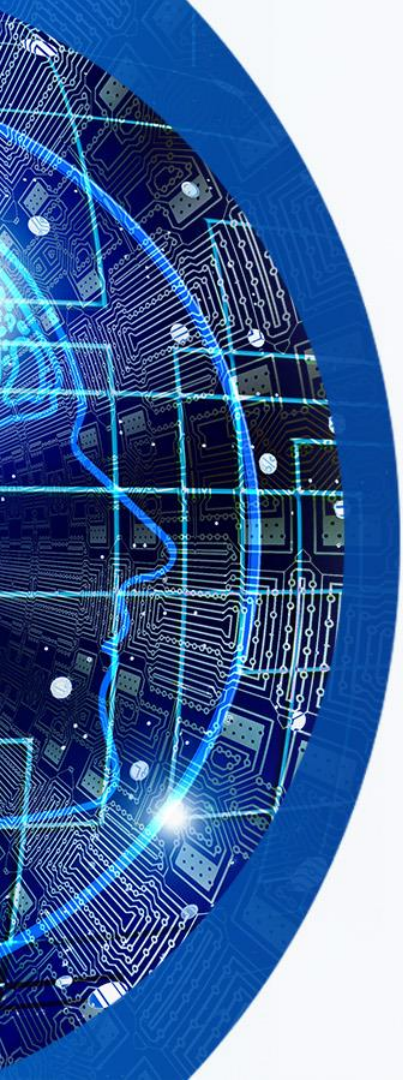
لماذا نتعلم علم البيانات

- إن كمية البيانات التي يتم إنتاجها عالميا غير مسبوقة ومن المتوقع أن تستمر مع نمو سكان العالم وزيادة إمكانية الوصول إلى الإنترنت، كانت هناك تغييرات واسعة النطاق في نمط سلوك العملاء في التسوق عبر الإنترنت، والاستثمار في سوق الأوراق المالية، وما إلى ذلك، وكل منها يتطلب الحاجة إلى تحليل عميق للبيانات. هذا هو المكان الذي يلعب فيه علم البيانات ويطلبه الشركات والمؤسسات وما إلى ذلك.



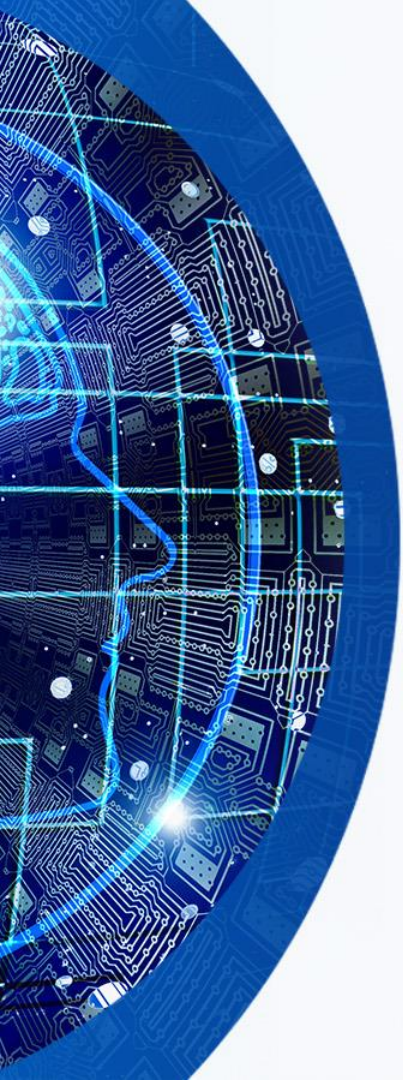
تحليل البيانات

• تحليل البيانات هو عملية جمع ونمذجة وفحص البيانات لاستخراج الأفكار المفيدة في صنع القرار. يُطلق على الشخص المسؤول عن هذه العملية محلل البيانات. يستخرج محلل البيانات هاته البيانات من خلال عدة طرق، مثل تنظيف وتحويل ونمذجة البيانات.



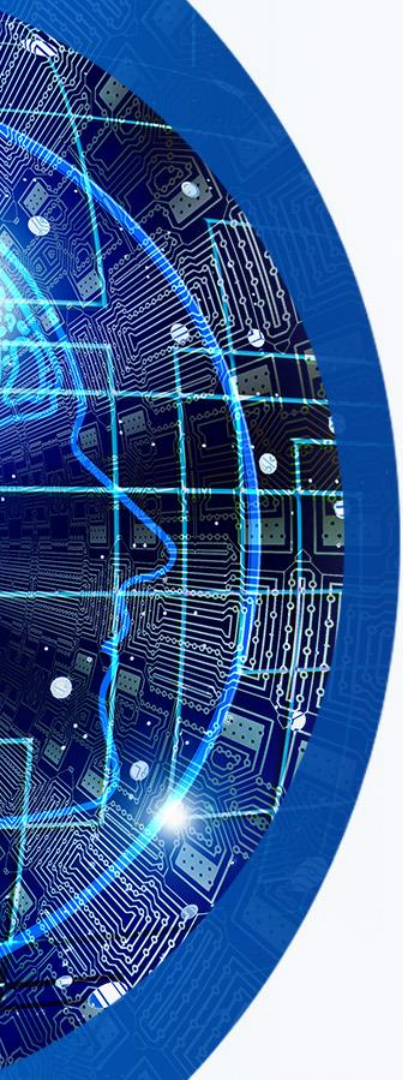
الفرق بين علم البيانات و تحليل البيانات

- على الرغم من أن علم البيانات وتحليل البيانات قد يأتي من مجال إحصائي مشترك ، إلا أن خرائطهم وجذورهم مختلفة.
- يتمثل الاختلاف الرئيسي الأول بين عالم البيانات ومحلل البيانات في أنه بينما يتعامل محلل البيانات مع حل المشكلات ، يحدد عالم البيانات المشكلات ثم يحلها. يتم توظيف محلي البيانات من قبل الشركات لحل مشاكلهم الصناعية ، الادارية و التجارية.
- لا يحتاج محللو البيانات إلى مهارات الاتصال أو الفطنة التجارية. فمحلل البيانات يقتصر على حدود تحليل البيانات.
- لا يتعامل محلل البيانات مع تطوير النمذجة التنبؤية أو الأدوات الإحصائية للتنبؤ بالبيانات.عكس عالم البيانات.
- يحتاج علماء البيانات إلى تعديل نماذج البيانات لإنشاء منتجات بيانات بشكل أفضل.



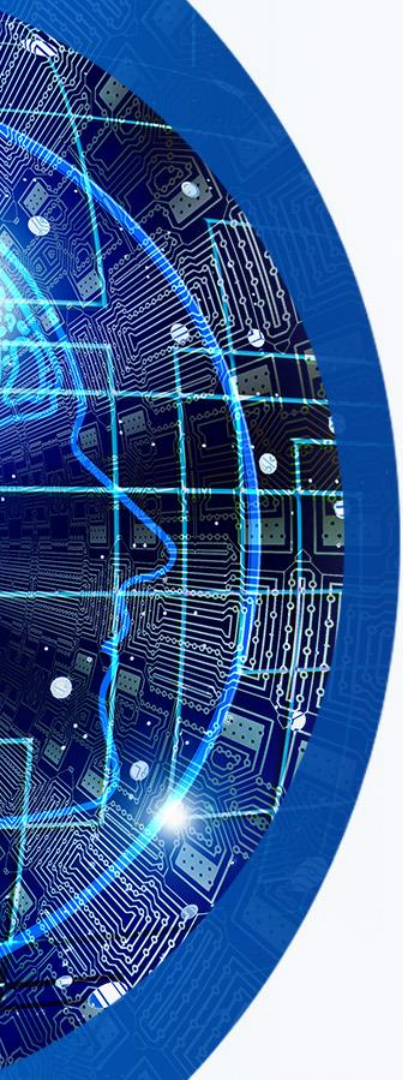
انواع تحليل البيانات

- هناك أربعة أنواع من تحليل البيانات تنطبق على جميع البيانات. ولكن ما هو أفضل نوع من نماذج تحليل البيانات لأي مسألة او قضية؟



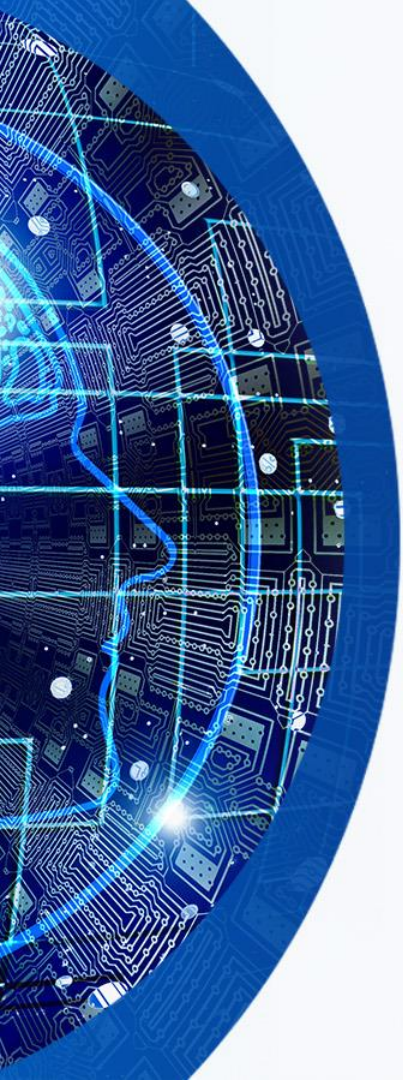
التحليل الوصفي - ماذا حدث Descriptive Analysis

- التحليل الوصفي هو الخطوة الأولى والأكثر شيوعاً والأبسط في إجراء أي عملية تحليل إحصائي، والغرض منه هو الإجابة عن السؤال ، ماذا حدث؟ بمعنى آخر ، فإنه يفحص ويصف ما حدث في الماضي: مثل الإيرادات الشهرية ، ومبيعات الأشهر الستة الماضية ، والزيارات السنوية لموقع الويب، وما إلى ذلك. يتم ذلك عن طريق تلخيص البيانات السابقة وبناء وتفسير البيانات الأولية من مصادر مختلفة لتحويلها إلى رؤى قيمة.



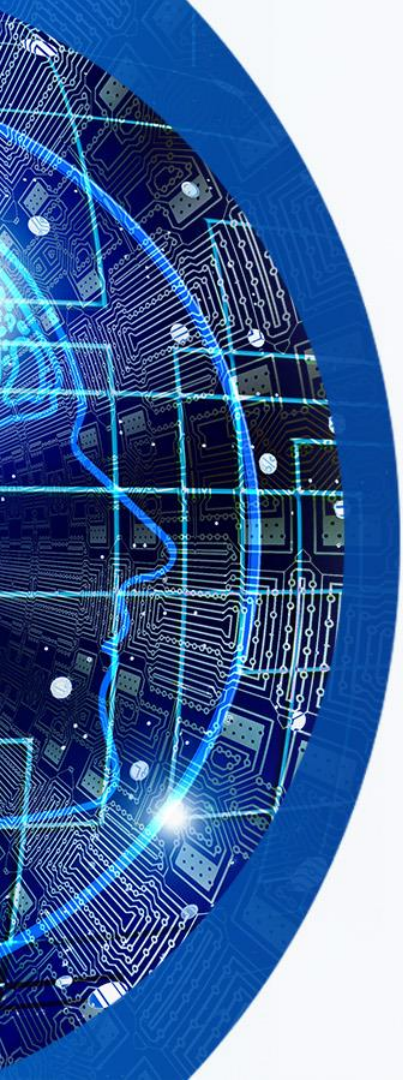
التحليل التشخيصي - لماذا حدث هذا

- يتميز التحليل التشخيصي بأساليب مثل التنقيب عن البيانات والارتباط، ويلقي نظرة أعمق على البيانات لفهم أسباب الأحداث والسلوكيات، مما يسمح لك بفهم معلوماتك بسرعة أكبر والإجابة على الأسئلة المهمة، حيث يفحص التحليل التشخيصي البيانات ويجعل العلاقات المتبادلة.



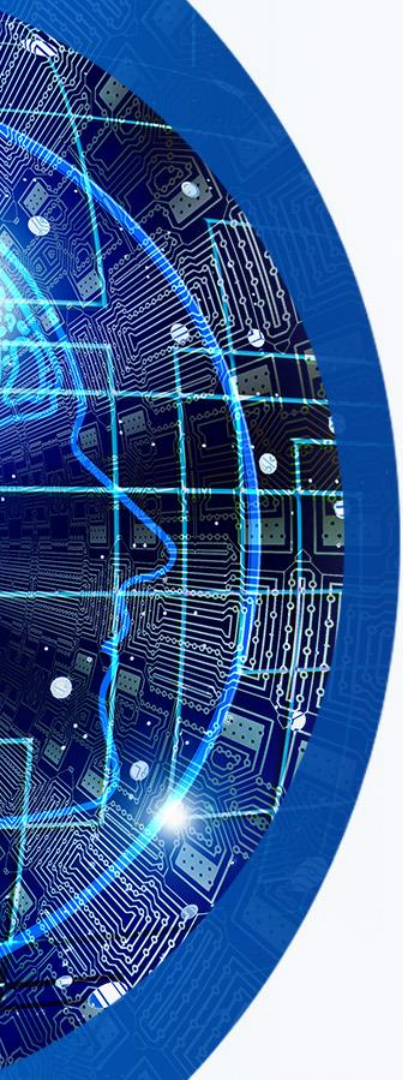
التحليل التنبئي - ماذا سيحدث *Predictive Analysis*

- يحدد التحليل التنبئي النتائج المحتملة من خلال تحديد الميول في التحليلات التشخيصية والوصفية. يأخذ التحليل التنبئي البيانات السابقة ويغديه في نموذج التعلم الآلي الذي يأخذ في الاعتبار الأنماط الرئيسية. ثم يتم تطبيق النموذج على البيانات الحالية للتنبؤ بما سيحدث. هذا يسمح للمنظمة لاتخاذ إجراءات وقائية.



التحليل التوجيهي - ما هي الإجراءات التي يجب اتخاذها؟

• التحليل التوجيهي هي عملية تحلل البيانات وتقدم نصائح فورية حول كيفية تحسين ممارسات الأعمال لتناسب العديد من النتائج المتوقعة في الواقع، التحليل الوصفي "يأخذ ما نعرفه(البيانات) "، ويفهم البيانات للتنبؤ بما يحدث، ويقترح أفضل الخطوات التطلعية بناء على المحاكاة من خلال سيناريوهات محتملة، ويحدد العواقب المحتملة لكل منها. الغرض من التحليل التوجيهي، هو اقتراح طريقة عملية لتجنب المشاكل المستقبلية أو لتحقيق أقصى استفادة من عملية التحليل.



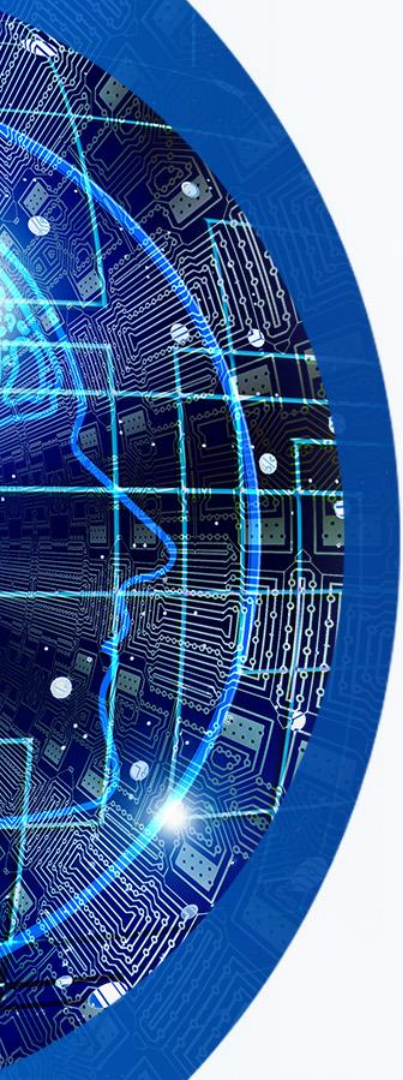
دورة حياة علم البيانات

• المرحلة الاولى: جمع البيانات

• ان عملية جمع البيانات ليست سهلة. يجب تحديد البيانات ذات الصلة بالمشروع. هناك أيضا مهمة تحديد موقع جمع البيانات. تتوفر أنواع مختلفة من مصادر البيانات عبر الإنترنت من خلال المواقع الرسمية و الإخبارية واستطلاعات الرأي وما إلى ذلك على الموقع و وسائل التواصل الاجتماعي.

• المرحلة الثانية: تنظيف البيانات

• من المهم معرفة أن البيانات التي تم جمعها في المرحلة الأولى غير منظمة. يجب تنقية البيانات الخام وتصنيفها. هذا يعني البحث عن أي (بيانات مكررة، بيانات شاذة، إلخ) في البيانات من أجل تجنب أي أخطاء في الخطوات التالية. بسبب المهام التي يتم تنفيذها ، عادة ما تكون المرحلة الثانية هي الجزء الذي يستغرق وقتا طويلا في مشروع علمي.

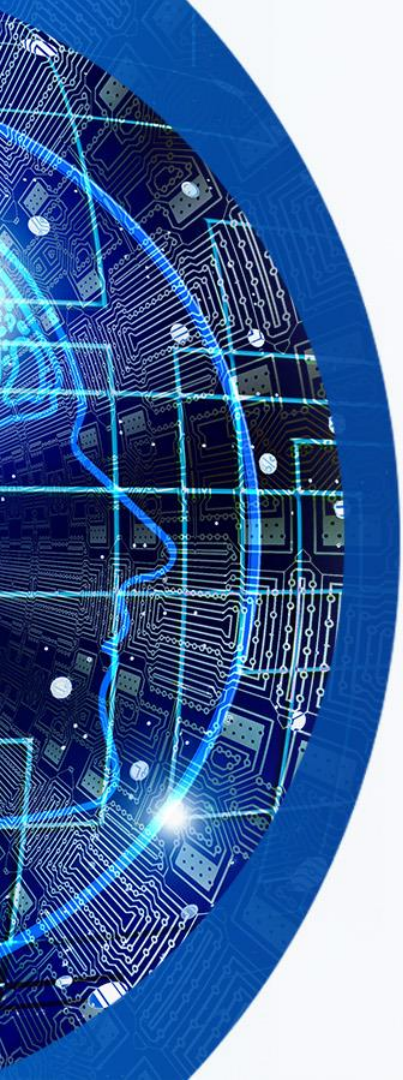


• المرحلة الثالثة: تنقيب البيانات

- يقوم علماء البيانات بفحص البيانات للحصول على فكرة أوسع عن الأنماط والاتجاهات الرئيسية في مجموعة البيانات. يسيطر التمثيل المرئي والتحليل الإحصائي على هذه المرحلة. يسلط تنقيب البيانات الضوء على النقاط التي تحتاج إلى مزيد من التحليل. تسمح أدوات التمثيل المرئي للبيانات بتسجيل واستكشاف العناصر المهمة بشكل أكبر

• المرحلة الرابعة: نمذجة البيانات

- النمذجة تعتبر فهم العلاقة بين عناصر البيانات. و من خلال نمذجة البيانات، يرى عالم البيانات كيف تتفاعل العناصر الأكثر أهمية وتتجمع معا. هناك عدة طرق لبناء نموذج. يمكن أن تكون هذه النمذجة من خلال أساليب التعلم الآلي أو النمذجة الإحصائية.

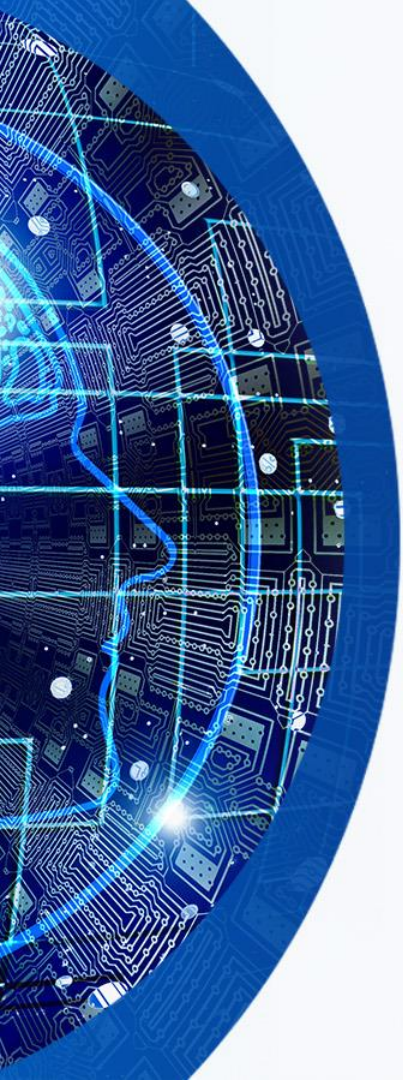


١٠ المرحلة الخامسة: تفسير البيانات

- هي الخطوة الأخيرة في دورة حياة علم البيانات، وهي تفسير البيانات. إذا كنت تريد تنفيذ اكتشافاتك القيمة، فيجب أن تكون قادرا على تقديم هذه الأفكار بطريقة جذابة ومفهومة حتى يتمكن أصحاب المصلحة في المشروع من فهمها بسهولة.

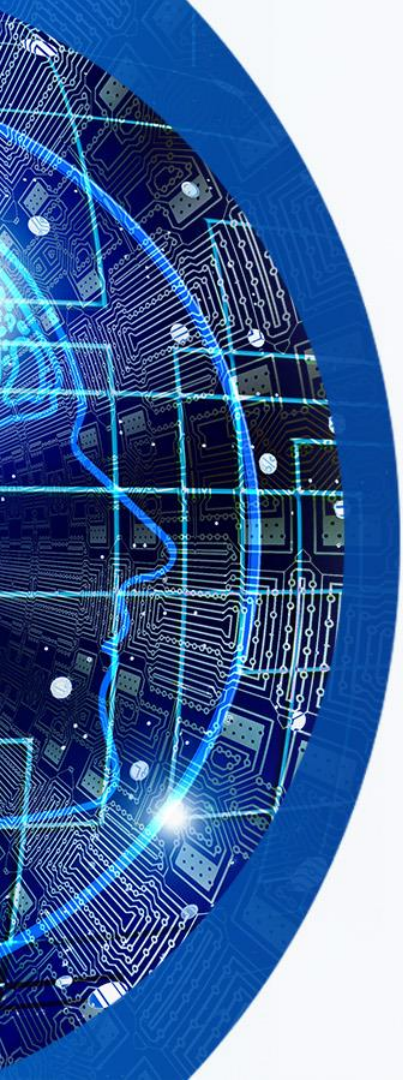
تعلم الآلة

• تعريف التعلم الآلي ، يعرف على أنه استخدام وتطوير أنظمة الكمبيوتر القادرة على التعلم والتكيف دون تخطيط صريح. تستخدم هذه الأنظمة الخوارزميات، الرياضيات والنماذج الإحصائية لتحليل واستنتاج الأنماط المختلفة في البيانات.



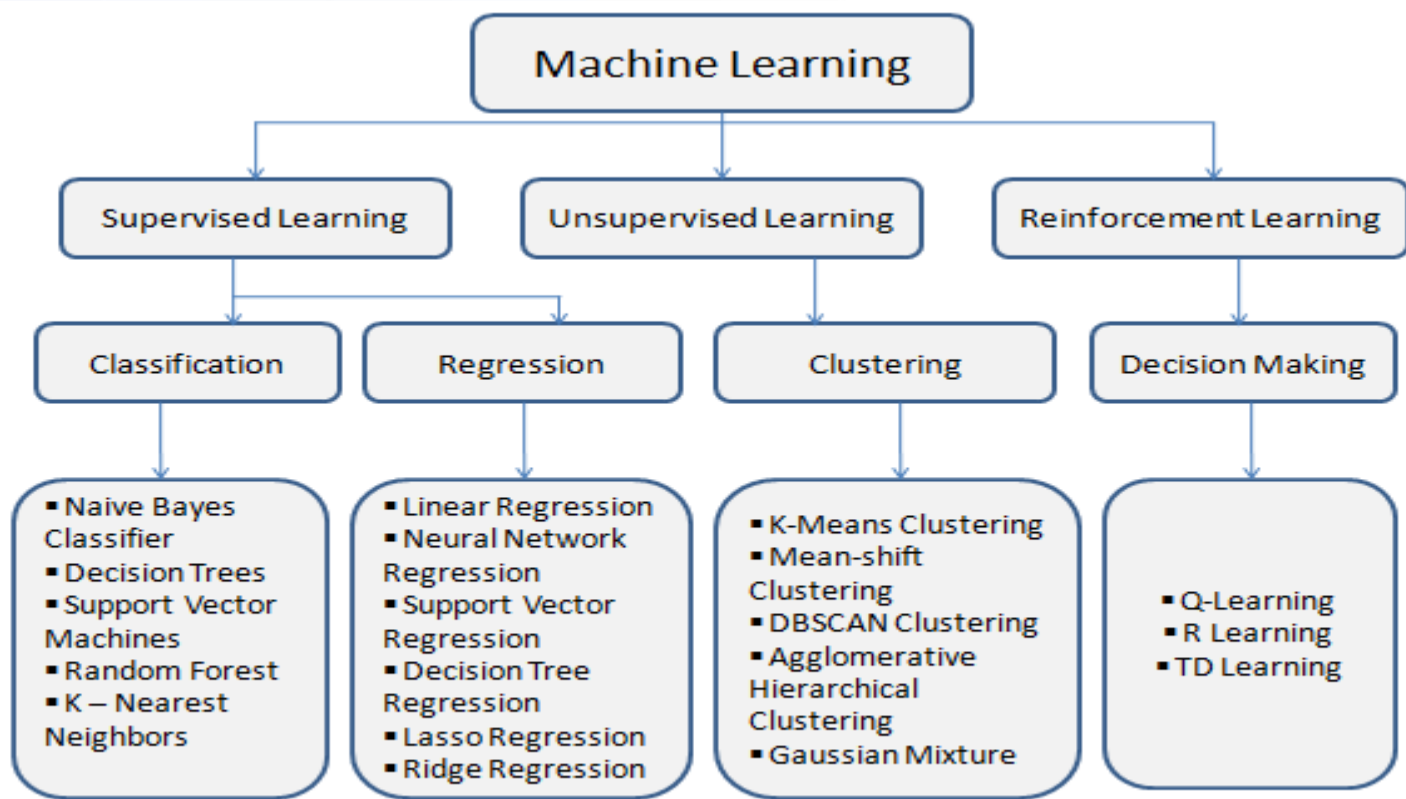
لماذا تعلم الالة؟

- التعلم الالي هو تكامل علوم الكمبيوتر والتخصصات الرياضية المختلفة التي تُستخدم فيها مفاهيم علوم الكمبيوتر لبناء نماذج رياضية قوية يمكنها حل مجموعة من المشكلات المتشابهة وذات الصلة.
- كمية البيانات المتاحة لنا تزايد باستمرار. تستخدم الالات هذه البيانات للتعلم وتحسين النتائج و تزويدنا بها.
- يتم استخدام التعرف على الصور وتوليد النص وتصنيف النص وتشخيص الامراض وغيرها الكثير في العالم الحقيقي.



طرق التعلم الآلي:

- يمكن أن تحقق خوارزميات التعلم الآلي القدرة على التعلم من خلال ثلاث طرق مختلفة:
- التعلم مع الإشراف: supervised learning التعلم الخاضع للإشراف في التعلم الآلي هو طريقة تنشئ نموذجاً للتنبؤ بالنتائج بناء على البيانات المصنفة، مثل السعر والكمية.
- التعلم دون إشراف: unsupervised learning التعلم غير الإشرافي في التعلم الآلي يكون عندما نقدم أمثلة إلى الخوارزمية دون أي توجيه ونترك إنشاء التسمية للخوارزمية. مثل نص كتابي، الشبكات الاجتماعية، ورسائل البريد الإلكتروني، والمدونات، والتغريدات، والصور الرقمية، وبيانات الجوال، وصفحات الويب، إلخ.
- التعلم المعزز: Reinforcement learning التعلم المعزز هو أسلوب يوفر تغذية راجعة تعليمية باستخدام آلية المكافأة. و تحدث عملية التعلم كعامل يتفاعل مع البيئة ويحاول إيجاد طرقاً مختلفة لتحقيق نتيجة. يتلقى الوكيل مكافأة أو عقوبة وذلك عندما يصل إلى الهدف أو لا. يمكننا استخدام التعلم المعزز عندما نحتاج إلى برنامج للعمل بنجاح في بيئة تنافسية، مثل لعبة الفيديو أو سوق الأوراق المالية.



المجالات الأساسية للدراسة في علم البيانات

• تعلم الآلة

• بالنسبة لعالم البيانات ، يعد تعلم الآلة مهارة أساسية. تتمثل الفكرة الأساسية للتعلم الآلي في السماح للألات بالتعلم بشكل مستقل باستخدام كتلة البيانات التي يتم تغذيتها بالجهاز كمدخلات. مع تقدم التكنولوجيا ، يتم تدريب الآلات على التصرف مثل البشر في القدرة على اتخاذ القرار.

• التعلم العميق

• غالبا ما يستخدم التعلم العميق في علم البيانات. لأنها تعمل بشكل أفضل بكثير من طرق التعلم الآلي. يمكن القول أنه بينما يستخرج التعلم العميق الميزات تلقائيا من بنية البيانات، يجب أن يتم ذلك يدويا عن طريق التعلم الآلي ، وإذا كان يقوم بتنبؤات غير صحيحة في عملية صنع القرار ، فإن الخبير أو المبرمج يجب أن يتناول هذه المشكلة صراحة.

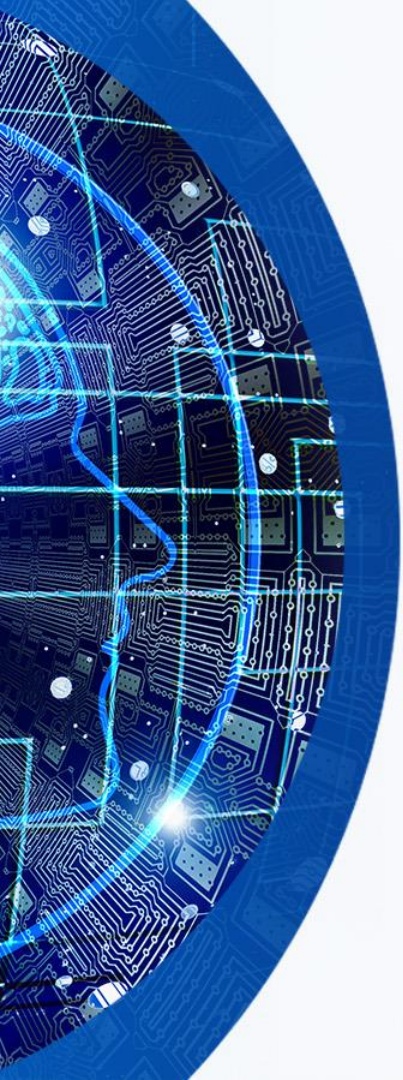
المهارات اللازمة لتعلم الآلة

• الرياضيات

- لتحسين مهارات التعلم الآلي، يجب أن يكون لدى عالم البيانات معرفة عميقة بالرياضيات.
- موضوعان مهمان في الرياضيات من حيث التطبيق في علم البيانات هما الجبر والحساب. في حين أن الجبر يدور حول دراسة المتجهات والوظائف الخطية ، فإن الحساب هو الدراسة الرياضية للتغير المستمر. يتم استخدام العديد من المفاهيم الجبرية مثل المتجهات في العديد من مجالات التعلم الآلي. وبالمثل ، فإن الحسابات المطلوبة في مجالات مختلفة من التعلم الآلي.

• الإحصاء والاحتمالات

- الإحصاء والاحتمالات من المتطلبات الأساسية في علم البيانات والمعرفة الجيدة في هذا المجال ضرورية . فنحن نعمل مع البيانات احتمالية ، هذا يعني أنه يجب أن يكون المرء على دراية بالاحتمالات والإحصاءات.

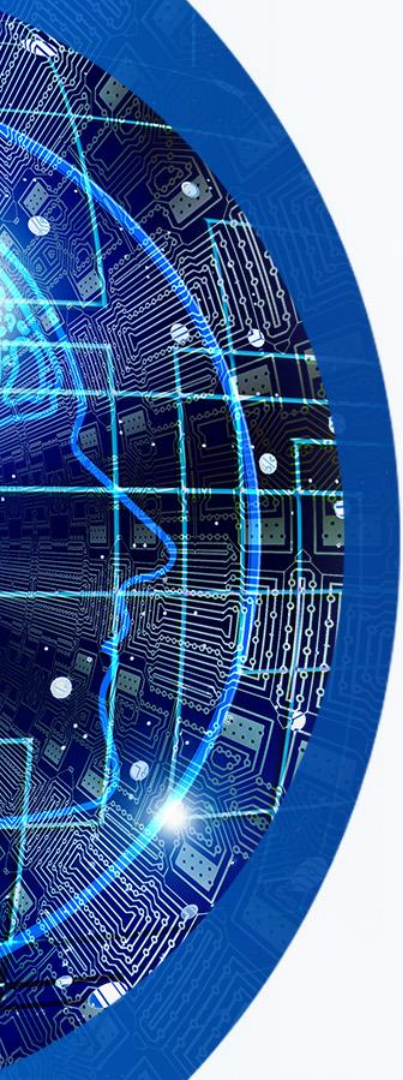


• معالجة اللغة الطبيعية

• معالجة اللغة الطبيعية هي فرع من فروع الذكاء الاصطناعي يركز على سد الفجوة بين التواصل بين الإنسان والآلة لتمكين الآلة من التفسير والفهم.

• العرض المرئي للبيانات

• يعد تمثيل البيانات أحد أهم فروع علم البيانات. ببساطة ، يتضمن الرسم التوضيحي عرض البيانات في شكل رسوم بيانية ومخططات بيانية.



• لغة البرمجة

• يجب أن يتمتع عالم البيانات ، بالإضافة إلى مهارات الكمبيوتر الأساسية مثل الكفاءة في Microsoft Excel, R, Python ، بمهارات البرمجة حتى يتمكن من استخدامها للعمل مع البيانات المعالجة والتمثيل المرئي وما إلى ذلك

• الخوارزميات

• نظرا لأن جميع أنظمة التعلم الآلي تعتمد على الخوارزميات ، فمن الضروري جدا أن يكون لدى عالم البيانات فهم أساسي للخوارزميات وكيفية تصميمها.

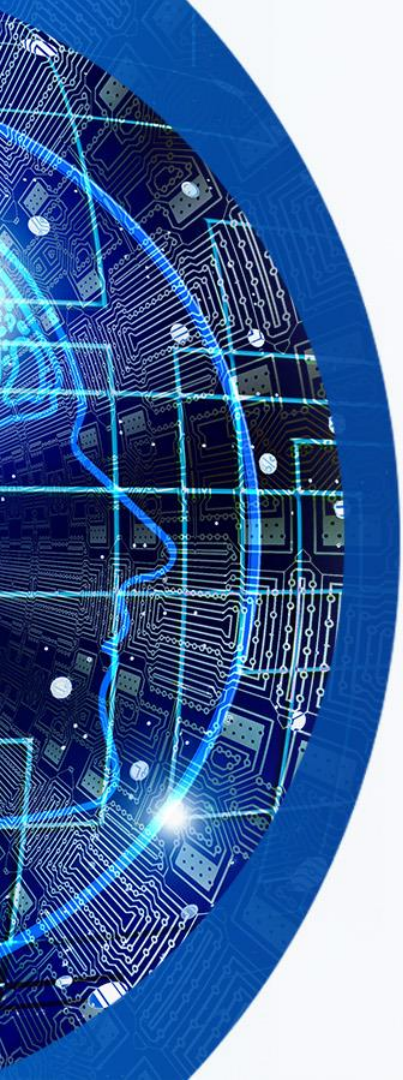
تطبيقات علم البيانات

• المواصلات

• أهم تقدم أو تطوير حققه علم البيانات في مجال النقل هو إدخال السيارات ذاتية القيادة. لقد أسس علم البيانات موطئ قدم قوي في صناعة النقل من خلال التحليل المكثف لأنماط استهلاك الوقود ، والمراقبة النشطة للمركبة وسلوك السائق ، وتوفير بيئات قيادة أكثر أمانا للسائقين، وتحسين أداء السيارة ، وإضافة الاستقلالية إلى السيارات وغير ذلك. باستخدام التعلم المعزز والاستقلالية ، يمكن لشركات صناعة السيارات بناء سيارات أكثر ذكاء وطرقا منطقية أفضل.

• كشف المخاطر والاحتيال

• تم استخدام علم البيانات لأول مرة في التمويل والمصارف. كانت العديد من المؤسسات المالية مثقلة بالديون في نهاية كل عام. لذلك ، تم اعتبار اساليب علم البيانات كحل. لتحليل احتمالية المخاطر ، تعلموا فصل البيانات بناء على مواصفات العميل والتكاليف السابقة والمتغيرات الضرورية الأخرى. وبالتالي ، يمكنهم القيام بالتسويق المستهدف بناء على إيرادات كل عميل كل عام.



• علم الوراثة والجينات الوراثية

• يساعد علم البيانات علماء الأحياء على تحليل استجابة الجينات للأدوية المختلفة. والغرض منه هو فهم ودراسة تأثير الحمض النووي على صحة الشخص، والذي يسعى إلى إيجاد روابط بيولوجية بين الأمراض والجينات والاستجابات للأدوية .

• تطوير الادوية

• يمكن استخدام خوارزميات التعلم الآلي و علوم البيانات لتبسيط عملية اكتشاف دواء جديد وتقليل الوقت اللازم للفحص الأولي لمركبات الأدوية المستخدمة في إنتاج الأدوية. يمكن أن تتنبأ الخوارزميات وعلوم البيانات أيضا بكيفية استجابة الجسم لمركبات دوائية معينة باستخدام نماذج ومحاكاة إحصائية ورياضية مختلفة. هذا أسرع بكثير من الاختبارات المخبرية التقليدية. يمكن للنماذج أيضا توقع النتائج المستقبلية بشكل أكثر دقة.



انتهى هذا المحور
شكرا لكم