

معايير التقييم

يوضح لك هذا الجدول ما يجب عليك القيام به من أجل الحصول على درجة النجاح أو التفوق أو الامتياز، والمكان الذي يمكنك فيه العثور على أنشطة لمساعدتك.

النجاح	التفوق	الامتياز
نتاج التعلم أ		
دراسة استخدامات الذكاء الاصطناعي وتطبيقاته		
A.P1	A.M1	A.D1 تقييم تأثير الذكاء الاصطناعي في المجالات المختلفة. تمرين تقييمي 21.1
وصف كيفية استخدام المفاهيم الأساسية للذكاء الاصطناعي لتلبية الاحتياجات المحددة.	تحليل فوائد الذكاء الاصطناعي ومخاطره وعيوبه وكيفية تأثيرها في المجالات المختلفة.	
تمرين تقييمي 21.1	تمرين تقييمي 21.1	
A.P2		
شرح الفوائد والمخاطر والعيوب المرتبطة بالذكاء الاصطناعي في مختلف المجالات.		
تمرين تقييمي 21.1		
نتاج التعلم ب		
تخطيط حل الذكاء الاصطناعي وإعداده لتلبية احتياجات محددة		
B.P3	B.M2	BC.D2 تقييم فعالية حل الذكاء الاصطناعي. تمرين تقييمي 21.2 تمرين تقييمي 21.3
تحديد أهداف مشروع الذكاء الاصطناعي.	مراجعة مجموعات البيانات لتحسين الحصول على أفضل جودة في حل الذكاء الاصطناعي.	
تمرين تقييمي 21.2	تمرين تقييمي 21.2	
B.P4		
جمع مجموعات البيانات المناسبة وإعدادها لحل الذكاء الاصطناعي.		
تمرين تقييمي 21.2		
نتاج التعلم ج		
تطوير حل الذكاء الاصطناعي لتلبية احتياجات محددة		
C.P5	C.M3	
تطوير حل الذكاء الاصطناعي باستخدام لغة برمجة وأدوات حوسبة مناسبة.	اختبار حل الذكاء الاصطناعي وتحسينه.	
تمرين تقييمي 21.3	تمرين تقييمي 21.3	

بدء النشاط

عندما تسمع مصطلح **الذكاء الاصطناعي**، كيف يكون رد فعلك تجاهه؟
هل تفكر في الذكاء الاصطناعي بطريقة إيجابية أم بطريقة يشوبها القلق؟
هل أحدث الذكاء الاصطناعي فرقاً في حياتك، وإذا كان الأمر كذلك، فكيف؟
لا تتردد في مشاركة تجاربك ووجهات نظرك ومناقشتها مع زملائك. هل يعتقدون الشيء نفسه؟



نتائج التعلم

ستتعلم في هذه الوحدة:

- أ { دراسة استخدامات الذكاء الاصطناعي وتطبيقاته
- ب { تخطيط حل الذكاء الاصطناعي وإعداده لتلبية احتياجات محددة
- ج { تطوير حل الذكاء الاصطناعي لتلبية احتياجات محددة



أ دراسة استخدامات الذكاء الاصطناعي وتطبيقاته

بعبارة بسيطة، الذكاء الاصطناعي هو في الأساس التكنولوجيا التي تسمح للحاسوب بالتفكير والتعلم من تلقاء نفسه، وهو ما يمنحه القدرة على حل المشكلات المحددة، واتخاذ الإجراءات بناءً على معايير محددة مسبقاً، وإجراء تنبؤات دقيقة باستخدام البيانات. وهذا يمكن أجهزة الحاسب من محاكاة قدرات البشر على حل المشكلات. تتطلب أنظمة الحاسب التقليدية الكثير من التفاعل البشري، أو البشر لإنشاء برامج لأتمتة العمليات. بمجرد الإعداد، يمكن للآلات التي تعمل بالذكاء الاصطناعي اتخاذ إجراءات دون الحاجة إلى أي تدخل بشري إضافي.

المصطلح الرئيس

الذكاء الاصطناعي (AI) – أحد مجالات علوم الحاسوب يركز على إنشاء آلات لديها القدرة على 'محاكاة تفكير الإنسان'.

أ1 استخدامات الذكاء الاصطناعي وتطبيقاته

الذكاء الاصطناعي قادر بالفعل على تنفيذ العديد من المهام التي يمكن للبشر القيام بها. على سبيل المثال، يمكنها:

- التعلم من التجربة
- فهم اللغات الطبيعية (أي المنطوقة أو المكتوبة)
- التعرف الأنماط أو الصور
- اتخاذ القرارات

المهارات

- المهارات المعرفية: العمليات والإستراتيجيات المعرفية:
- التفكير الناقد
- حل المشكلات
- المهارات المعرفية: الإبداع:
- الابتكار

هل لديك أي مساعدين افتراضيين في منزلك؟ ما الغرض من استخدامهم؟

وقفة للتفكير



قد تكون هذه أي خدمات تستند إلى السحابة يمكنك التفاعل معها للإجابة عن الأسئلة أو للتحكم في الأجهزة باستخدام الأوامر الصوتية.

تلميح

ما مدى دقة الإجابات التي تلقيتها؟ إذا كنت تستخدمها للتحكم في الأجهزة، هل كنت تتلقى دائماً الاستجابة المتوقعة؟

توسيع الأفق

الذكاء الاصطناعي القوي

- يُعرف أيضًا باسم الذكاء الاصطناعي للأغراض العامة (GPAI)
- مصمم لتطبيق الذكاء على أي مشكلة، وليس فقط مهمة محددة
- يحاكي القدرات المعرفية البشرية مثل حل المشكلات وإصدار الأحكام والتخطيط والتعلم وتجميع الخبرات في سياقات جديدة
- من الناحية النظرية، لا يمكن تمييزه عن شخص حقيقي
- ما زال هدفًا نظريًا

الذكاء الاصطناعي الضعيف

- يُعرف أيضًا باسم الذكاء الاصطناعي الضيق
- مصمم لأداء مهمة محددة
- يعمل بقدرة محدودة
- يحاكي السلوك البشري بناءً على مجموعة القواعد والبيانات المتاحة
- الذكاء الاصطناعي الضعيف يشمل المساعدين الشخصيين مثل أليكسا (أمازون®) أو سيرى (أبل®)، أنظمة توصيات البث، أنظمة تعرف الوجه، وغيرها.

الشكل 21.1 مقارنة بين الذكاء الاصطناعي الضعيف والقوي.

الاختلافات بين الذكاء الاصطناعي الضعيف والذكاء الاصطناعي القوي

عادةً ما يتم تصميم الذكاء الاصطناعي على أنه إما (ضعيف) وإما (قوي) بطبيعته. يوضح الشكل 21.1 نظرة عامة على الاختلافات الرئيسية:

الفرق الرئيس هنا هو نطاق التطبيق، فالذكاء الاصطناعي الضعيف ضيق جدًا في التركيز، في حين أن الذكاء الاصطناعي القوي (الذي ما يزال قيد التطوير) له هدف واسع وطموح، يتضمن القدرة على فهم مجموعة من الوظائف البشرية وتعلمها وأدائها – تمامًا مثلما تعمل أدمغتنا.

إن أساس الذكاء الاصطناعي القوي هو مفهوم الذكاء الاصطناعي التأسيسي، ويشير إلى النماذج والتقنيات الأساسية التي تشكل أساس تطبيقات الذكاء الاصطناعي المختلفة، وعادةً ما تكون مثل هذه النماذج واسعة النطاق وتخضع للتدريب على مجموعات بيانات متنوعة، تسمح بفهم وإنشاء نص يشبه ما يكتبه الإنسان، وتعرف الصور، ومعالجة الكلام وغير ذلك الكثير. على سبيل المثال، توجد نتيجة تم اختبارها جيدًا من تقنية الذكاء الاصطناعي التأسيسية تسمى GPT (المحول التوليدي المُدرَّب مسبقًا) في قلب أداة ChatGPT AI التي حظيت بتغطية إعلامية واسعة.

أنواع الذكاء الاصطناعي

هناك أنواع مختلفة من الذكاء الاصطناعي. يتم تصنيفها بشكل عام من حيث القدرة التي تتمتع بها التكنولوجيا.

الآلات التفاعلية

هذا نوع أساسي من الذكاء الاصطناعي لا يخزن أي ذاكرة لأفعاله أو تجاربه السابقة، لذلك لا يمكنه استخدام هذه المعلومات لاتخاذ القرارات في الوقت الحاضر. كان الحاسوب الفائق للعبة الشطرنج من IBM® المسمى ديب بلو (Deep Blue) أحد الأمثلة على الآلة التفاعلية، وكان لديه القدرة على تعرف قطع الشطرنج وكان يعرف كيف لهذه القطع أن تتحرك. يمكن أن يقوم أيضًا بعمل تنبؤات حول التحركات المحتملة. لذلك كان قادرًا على اتخاذ قرارات بناءً على تصرفات الخصم ولكنه لم يكن لديه ذاكرة للإجراءات التي تم اتخاذها في وقت سابق من اللعبة.

أجهزة الذاكرة المحدودة

هذا نوع أكثر تقدمًا من الذكاء الاصطناعي يمكنه استخدام التجارب السابقة لاتخاذ قرارات مستقبلية. الدردشة مع روبوتات خدمة العملاء هي أمثلة جيدة على الآلات ذات الذاكرة المحدودة. تقوم هذه الأنظمة بمعالجة البيانات باستخدام نماذج رياضية معقدة تسمح لها بالاستجابة للمدخلات السابقة وبالتالي محاكاة الذكاء البشري. تُستخدم روبوتات الدردشة لتقديم دعم العملاء على مدار الساعة طوال أيام الأسبوع، وغالبًا ما يتم إعدادها بنطاق محدود من الاستجابات وبالتالي فهي غير قادرة على التعامل مع المشكلات المعقدة وحلها.

المصطلحات الرئيسية

النموذج – هيكل ينشأ عن طريق عملية تسمى التعلم الآلي (غالبًا ما يُشار إليها باختصار ML) وتدعمه إحدى الخوارزميات.

الخوارزمية – هي سلسلة من الخطوات أو التعليمات المُنفَّذة لتحقيق نتيجة يمكن تعريفها.

الآلات ذات نظرية العقل

هذه هي الآلات التي تتمتع بقدرات الذكاء الاصطناعي المتقدمة التي تسمح لها بتعرف التعبيرات والعواطف والنوايا وفهمها وما إلى ذلك. يمكنهم استخدام هذا الفهم للتفاعل مع البشر. نظرًا إلى أن المشاعر البشرية يمكن أن تكون ذاتية، فيمكن أن يكون الذكاء الاصطناعي عرضة **للتحيز** بالطريقة التي يفسر بها مثل هذه المشاعر. هناك أمثلة للآلات ذات نظرية العقل في كل من الرعاية الصحية والسيارات.

الذكاء الاصطناعي ذاتي الإدراك

هذا هو مستقبل الذكاء الاصطناعي. الآلات الذكية ذاتية الإدراك ستتمتع بوعي خاص بها وستكون قادرة على فهم (والتصرف بناءً على) العواطف والمعتقدات والرغبات والاحتياجات والرغبات الخاصة بها. ما تزال الآلات التي تتوافق مع نظرية العقل والذكاء الاصطناعي ذي الإدراك الذاتي قيد البحث والتطوير. حث العديد من الأشخاص، بما في ذلك المتخصصون المشهورون عالميًا في مجال الذكاء الاصطناعي **والتعلم الآلي**، على توخي الحذر، وأثاروا المخاوف بشأن مشكلات التحكم، والسلامة، وسوء الاستخدام المحتمل. وقد أدى ذلك إلى تطوير قواعد الاتحاد الأوروبي البارزة التي تحكم استخدام الذكاء الاصطناعي. تشمل هذه القواعد مشكلات منها **المراقبة البيومترية** وكيفية تنظيم أدوات الذكاء الاصطناعي الناشئة مثل ChatGPT المتقدم من OpenAI.

المصطلحات الرئيسية

التحيز – تفضيل أو موقف له تأثير مباشر في الحيد ويرجح إصدار حكم في اتجاه معين.

التعلم الآلي – علم تصميم وتطوير الخوارزميات التي تسمح لأجهزة الحاسوب بالتعلم والاستجابة بناءً على هذا التعلم.

المراقبة البيومترية – هي أنظمة تجمع كميات كبيرة من البيانات البيومترية لتمكين تتبع الأفراد بناءً على خصائصهم الفيزيائية أو البيولوجية.

التعلم العميق – هو جزء من التعلم الآلي، وهو نفسه أحد مجالات الذكاء الاصطناعي ويهدف إلى تعليم الحواسيب كيفية التعلم من التجربة، وعادةً ما تستخدم نماذج التعلم العميق الشبكات العصبية لتحقيق هذا الهدف.

الشبكة العصبية – برنامج متطور، مستوحى من بنية الدماغ البشري ووظيفته، يهدف إلى إيضاح كيفية تعلمنا من تجاربنا.

هل سبق لك استخدام روبوت دردشة؟ متى فعلت هذا؟

وقفة للتفكير



هل طُلب منك تقديم بعض التعليقات على موقع إلكتروني استخدمته؟ أو هل دخلت في اختبار عبر الإنترنت؟ هل تعلم أنك تتحدث إلى روبوت محادثة؟

تلميح

هل شعرت حقًا وكأنك تتحدث إلى إنسان؟ هل قدم برنامج الدردشة الآلي دائمًا الردود المناسبة؟ إن كان الجواب "لا"، اذكر السبب؟

توسيع الأفق

مجموعات فرعية من الذكاء الاصطناعي

يغطي الذكاء الاصطناعي مجموعة واسعة من التقنيات المختلفة ويمكنه محاكاة مجموعة من الوظائف المعرفية البشرية المختلفة (العقلية والجسدية وكلاهما).

التعلم الآلي (Machine learning/ML)

التعلم الآلي هو مجموعة فرعية من الذكاء الاصطناعي تهتم بتطوير الخوارزميات التي يمكنها التعلم من البيانات وأداء المهام دون تعليمات محددة. هذه هي التكنولوجيا التي تدعم تعرف الوجوه وأتمتة البريد الإلكتروني وتصفية البريد العشوائي وتطبيقات تحويل الصوت إلى نص على الهاتف المحمول والنص التنبؤي. يعتبر التعلم الآلي في صميم تطوير مثل هذه الأنظمة الآلية.

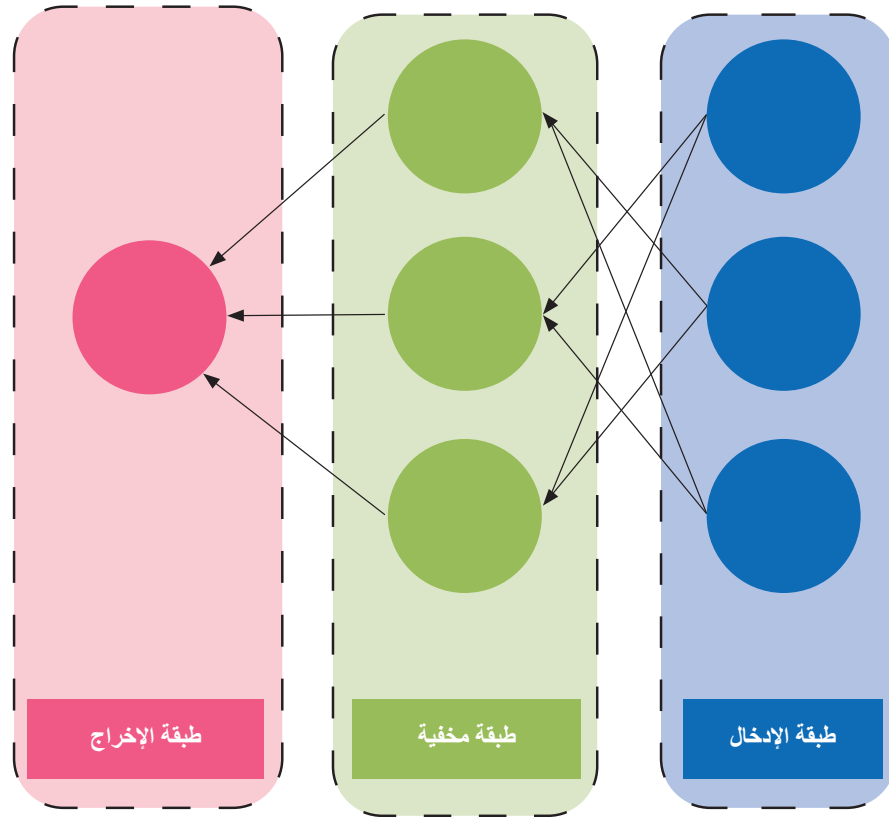
التعلم العميق (Deep learning/DL)

هذا التعلم هو مجموعة فرعية من التعلم الآلي، وعادةً ما يستخدم **التعلم العميق** الشبكات العصبية للسماح لأجهزة الحاسوب بالتعلم من التجربة.

فالشبكة العصبية هي شبكة من الوحدات المترابطة (أو 'الخلايا العصبية') التي يتم ترتيبها في طبقات، وتعالج كل خلية عصبية البيانات المدخلة وتتعلم منها ويمكنها اتخاذ قرارات أو تنبؤات بناءً على ما تعلمته.

موضوعات ذات صلة

لمزيد من المعلومات حول مزايا وعيوب الأنواع المختلفة من الذكاء الاصطناعي، راجع الجزء **استخدامات الذكاء الاصطناعي في الصناعات والقطاعات المختلفة** في هذه الوحدة.



الشكل 21.2 طبقات الشبكة العصبية التي توضح كيفية تحويل الطبقة المخفية للبيانات.

الطبقة المخفية مسؤولة عن تحويل المدخلات من طبقتها السابقة إلى شيء يمكن استخدامه لعمل تنبؤات أو تنفيذ مهام محددة. يقوم بذلك من خلال تعلم تصفية أهم جوانب بيانات الإدخال، وإضافة "وزن" أكبر أو أقل إليها، ما يؤثر في طريقة إجراء التنبؤات أو اتخاذ القرارات.

تستخدم الشبكات العصبية بشكل شائع في تعرف الصوت (أنظمة تحويل الصوت إلى نص) للهواتف المحمولة والمساعدات الشخصية وخدمات التوصية على خدمات البث الشائعة وحتى في أنظمة التشخيص الطبي.

يعد التعلم العميق وراء تطبيقات مثل أنظمة اكتشاف الاحتيال التي يتم استخدامها في الخدمات المالية لاكتشاف الاحتيال ومنعه. ويتم تطبيقه في نطاق واسع أيضًا في أنظمة إدارة علاقات العملاء (customer relationship management/CRM) لاكتشاف رؤى أعمق عن العملاء. يمكنه القيام بذلك لقدرته على تجميع البيانات التنظيمية بشكل أكثر فعالية (وبسرعة أكبر) مما يمكن لأي شخص تحقيقه.

معالجة اللغة الطبيعية (Natural language processing/NLP)

تقنية الذكاء الاصطناعي هي أصل التطبيقات الشائعة للتصحيح التلقائي والتدقيق الإملائي. كما أنه يدعم قدرة الأنظمة على الترجمة بين اللغات. تطبيق آخر للمعالجة الطبيعية للغة هو 'اكتشاف الطوارئ'. يتم استخدامه لتحديد التعبيرات أو الكلمات أو العبارات التي عادة ما تكون مدعاة للقلق أو اتخاذ إجراء فوري.

التعرف على الكلام

تقنية الذكاء الاصطناعي هذه راسخة بالفعل، وتستخدم في نطاق واسع في التطبيقات مثل القياسات الحيوية الصوتية في أنظمة الأمان، والأوامر الصوتية للأجهزة الذكية، وأنظمة تحويل الكلام إلى نص التي تمكن المستخدمين من استخدام الحاسوب دون استخدام اليدين.

بحث

هل تعلم أن الجامعة الأردنية قدمت برنامج الماجستير في الذكاء الاصطناعي والروبوتات في عام 2022؟ استخدم الإنترنت لمعرفة ما يتضمنه البرنامج ومتطلبات القبول. ناقش النتائج التي توصلت إليها مع زميل.

الرؤية الحاسوبية

هي استخدام الذكاء الاصطناعي في معالجة البيانات المرئية. يمكن للرؤية الحاسوبية تصنيف الكائنات باستخدام الخصائص المرئية. حدثت زيادة في تطبيقات الهاتف القائمة على الذكاء الاصطناعي التي تستخدم هذه التكنولوجيا. على سبيل المثال، يستخدم أحد هذه التطبيقات الرؤية الحاسوبية لتحديد النباتات وتشخيص أمراض النبات وتقديم توصيات حول العلاجات.

الأنظمة الخبيرة

تستخدم هذه الأنظمة قاعدة بيانات لمعارف الخبراء لتقديم المشورة بشأن موضوع معين. وهي الآن شائعة في مجموعة متنوعة من المجالات، بما في ذلك إدارة المعلومات ومكتب المساعدة. كما أنها تستخدم بطريقة محدودة، في الطب. تستخدم أنظمة خبراء الطب قاعدة بيانات معرفية وقواعد محددة مسبقاً لإجراء تشخيصات طبية محدودة. على سبيل المثال، يتم استخدامها في أنظمة العزل الطبي في المستشفيات. يتم استخدامها أيضاً في خطوط هاتف الطوارئ لاتخاذ قرارات بشأن إرسال سيارات الإسعاف.

وقفة للتفكير

هل رأيت أنظمة الخبراء المستخدمة في خدماتك الصحية؟

تلميح

لماذا تُستخدم؟ على سبيل المثال، هل تستخدم خدمة الإسعاف المحلية أنظمة متخصصة لفرز المرضى؟

توسيع الأفق

هل تعتقد أن مثل هذه الأنظمة دائماً فكرة جيدة؟

المهارات

مهارات التواصل الشخصي: الانفتاح الفكري:

- القدرة على التكيف
- التعلم المستمر

الروابط بين الذكاء الاصطناعي والتخصصات الأخرى

يرتبط الذكاء الاصطناعي ارتباطاً وثيقاً بالتخصصات الأخرى، بما في ذلك علوم الحاسب والرياضيات وعلم النفس واللغويات. يوضح الشكل 21.3 هذه الروابط القوية ويوضح كيف يتناسب الذكاء الاصطناعي مع التخصصات الأخرى.

- توفر علوم الحاسب مفاهيم البرامج الأساسية والنظريات الحسابية وهياكل البيانات وتصميم الأنظمة للعديد من حلول الذكاء الاصطناعي.
- وتوفر الرياضيات الخوارزميات والمنطق والإحصاءات والنمذجة التنبؤية.
- يوفر علم النفس رؤى حول العمليات المعرفية البشرية - كيف نتصرف ونتعلم ونتخذ القرارات.
- يمنحنا علم اللغة فهماً لكيفية تواصل الناس باستخدام اللغات الطبيعية، في كل من الأشكال المكتوبة واللفظية.

علوم الحاسوب

الرياضيات

الذكاء الاصطناعي

اللغويات

علم النفس

الشكل 21.3 الذكاء الاصطناعي وارتباطاته بالتخصصات الأخرى ذات الصلة. أي منها تعرف؟

التطورات في مجال الذكاء الاصطناعي

هناك عدد من التطورات الحالية في الذكاء الاصطناعي تمثل ما يمكن أن يحققه الذكاء الاصطناعي على المدى القصير. هناك أيضًا مجموعة من أهداف الذكاء الاصطناعي النظرية الطموحة التي أبرزها الباحثون - وتكون هذه الأهداف الشكل الذي سيبدو عليه الذكاء الاصطناعي في المستقبل.

الاحتمالات الحالية مع الذكاء الاصطناعي

تتخذ خطوات كبيرة إلى الأمام في العديد من المجالات، بما في ذلك ما يأتي.

التعلم الآلي والتعلم العميق

ساعدت خوارزميات التعلم الآلي، خاصة تلك التي تتبنى التعلم العميق، على تحفيز شعبية الذكاء الاصطناعي ونموه. كما هو موضح أعلاه، يستخدم التعلم العميق شبكات عصبية متعددة الطبقات لتحليل البيانات. وهي تؤدي حاليًا إلى تحسينات كبيرة في تعرف الصور والكلام ومعالجة البيانات. كما أنها تستخدم في الكشف عن الاحتيال، والخدمات المالية (المساعدة على إدارة المحافظ الاستثمارية)، والرعاية الصحية (في تطوير حلول العلاج)، وفي مجالات مختلفة (على سبيل المثال، في تعزيز سلامة الموظفين في المصانع بتحديد الوقت الذي يكون فيه الموظفون أو الأشياء على مقربة من الآلات).

الروبوتات

لقد قاد الذكاء الاصطناعي ثورة تطوير السيارات ذاتية القيادة على طرقتنا، والطائرات دون طيار، والروبوتات القادرة على أداء المهام الأكثر تعقيدًا (مثل الجراحة المجهرية). يسهم الذكاء الاصطناعي في تطوير العديد من الجوانب المختلفة للروبوتات (المتنقلة والثابتة على حد سواء). وتشمل الآتي:

- الإدراك، بما في ذلك الرؤية الحاسوبية وإدراك العمق ومداخلات أجهزة الاستشعار
- إرشادات الملاحة الأنيقة وتخطيط المسار
- التحكم في الحركة
- التكيف مع التجارب الجديدة (على سبيل المثال، الاختناقات المرورية) (التعامل مع العوائق غير المتوقعة)، وإعادة التوجيه، وتجنب المشاة، وما إلى ذلك)
- المعالجة الذاتية (حلول الذكاء الاصطناعي التي يمكنها تنفيذ المهام دون أي تفاعل بشري).

الرؤية الحاسوبية

يعد الذكاء الاصطناعي مكونًا رئيسيًا في الأنظمة التي تتعرف الوجوه (على سبيل المثال، لفتح الهواتف الذكية). يتم استخدام الرؤية الحاسوبية في أنظمة الواقع المعزز والألعاب التفاعلية وأنظمة الأمان المحسنة (على سبيل المثال، أمن الحدود في المطارات وما إلى ذلك).

يسهم الذكاء الاصطناعي في تطوير العديد من الجوانب المختلفة لرؤية الحاسوب بما في ذلك:

- تصنيف الصور (على سبيل المثال، الأحرف المكتوبة بخط اليد، والفحوصات الطبية، وما إلى ذلك)
- اكتشاف الكائن (على سبيل المثال، اكتشاف فئات معينة من الكائنات في الصور)
- تمييز الكائنات داخل الصورة (على سبيل المثال، اختيار شخص من الخلفية)
- تتبع الكائنات (على سبيل المثال، تتبع شخص عبر إطارات متعددة من تسلسل الفيديو)
- تحسين الصورة (على سبيل المثال، تقليل الضوضاء والترقية وضبط السطوع والتباين وما إلى ذلك).

هل يمكنك التفكير في أمثلة على استخدام الرؤية الحاسوبية في العالم من حولك؟

وقفة للتفكير



هل تلعب الألعاب التفاعلية؟ هل يتم استخدام الرؤية الحاسوبية في المطار المحلي؟

تلميح

هل تدرك متى تتفاعل مع الذكاء الاصطناعي؟ إذا لم يكن الأمر كذلك، قم ببعض البحث حول طرق استخدام الرؤية الحاسوبية في العالم من حولك.

توسيع الأفق

المصطلح الرئيس

اختيار الميزة – عملية اختيار مجموعة فرعية من المتغيرات المتاحة، المستمدة من مجموعة البيانات الأساسية، بحيث يمكن استخدامها على أنها مدخلات في أحد النماذج.

معالجة اللغة الطبيعية (NLP)

كما ذكرنا سابقًا، أدت التحسينات التكنولوجية إلى تحسين فهم أجهزة الحاسوب للغات الطبيعية (على سبيل المثال، الإنجليزية والفرنسية والألمانية وما إلى ذلك). وقد أتاح ذلك عددًا ضخمًا من التطبيقات العملية مثل روبوتات الدردشة (للخدمات عبر الإنترنت) والمساعدات الافتراضيين (Alexa من Amazon و Siri من Apple وما إلى ذلك) وخدمات الترجمة المختلفة (على سبيل المثال، Google Translate™). يتم استخدام الذكاء الاصطناعي في البرمجة اللغوية العصبية لتقديم خدمات محسنة مثل:

- تعرّف الكلام
 - تصنيف النص (مواضيع مختلفة ولغات مختلفة وما إلى ذلك)
 - تعرف المسميات (الأسماء الصحيحة، المواقع، التواريخ، وما إلى ذلك) في النص (على سبيل المثال، في العبارة 'جبن ذاهبة إلى مهرجان غلاستونبري الموسيقي في يونيو 2024'، سيكون النظام قادرًا على تحديد كلمة 'جبن' بوصفها اسم شخص، و'غلاستونبري' بوصفها موقع، و'يونيو 2024' بوصفها تاريخ)
 - ترجمة اللغة (على سبيل المثال، الإنجليزية إلى الفرنسية)
 - إنشاء حوار لإخراج النص (على سبيل المثال، لروبوتات المحادثة ومحركات البحث وما إلى ذلك)
 - الإجابة عن الأسئلة
 - تحليل المشاعر (على سبيل المثال، تحديد مشاعر الشخص على أنها إيجابية أو سلبية أو محايدة بطبيعتها)
 - تلخيص الموضوعات لمختلف أعمار القراءة المستهدفة.
- يعتمد استخدام الذكاء الاصطناعي في معالجة اللغة الطبيعية على القواعد (إذ يستخدم قواعد نحوية ولغوية ثابتة) وهو محدد بشكل كبير بالمهمة، بدلاً من أن يكون عامًا. ويتضمن هذا الذكاء الاصطناعي مزيجًا من التعلم تحت الإشراف والتعلم دون إشراف، فعلى سبيل المثال، في تحليل المشاعر، يخضع نظام الذكاء الاصطناعي للتدريب عادةً عن طريق البيانات المصنفة، في حين يمكن تصنيف أنواع مماثلة من المستندات عبر اختيار الميزات. (يرد شرح هذه المفاهيم لاحقًا في الوحدة.)

الأهداف النظرية للذكاء الاصطناعي

هناك قدر كبير من البحث المستمر والابتكار والتطوير في مجالات الذكاء الاصطناعي المختلفة. هذا العمل مدفوع بمجموعة من الأهداف الطموحة. يأمل الباحثون ورجال الأعمال في تطوير الذكاء الاصطناعي الذي لديه القدرة على:

- فهم وتفسير البيانات المعقدة (للتفكير واتخاذ قرارات مستنيرة)
- العمل بشكل مستقل (لكي يكون مستقلًا وقادرًا على التكيف بناءً على الظروف المتغيرة)
- تطبيق المعرفة في المواقف الجديدة (لكي يكون قادرًا على التعميم بناءً على بيانات التدريب المحدودة)
- إجراء تفاعلات طبيعية شبيهة بالإنسان، وتمييز المشاعر والنوايا المختلفة والتكيف معها
- توليد أفكار وحلول جديدة للمشكلات والأعمال الإبداعية (في الموسيقى والأدب والفن وما شابه ذلك)
- عرض الذكاء على مستوى الإنسان (للعمل بطريقة لا يمكن تمييزها عن الإنسان).

يجب أن نتذكر دائمًا أن المحرك الأساسي لتطوير الذكاء الاصطناعي يجب أن يكون بناء أنظمة قوية تعمل بأمان وبطريقة أخلاقية ومسؤولة. يجب أن يكون الهدف من أي بحث في هذا المجال هو تحسين حياة البشر والمساعدة على حل العديد من التحديات التي تواجهنا في القرن الحادي والعشرين (وما بعده).

تطبيقات الذكاء الاصطناعي

يحتوي الذكاء الاصطناعي على تطبيقات (وتطبيقات محتملة) في العديد من المجالات المختلفة، ويمكن أن يكون لكل منها تأثير مباشر في حياتنا اليومية.

هناك العديد من الأسباب لاستخدام الذكاء الاصطناعي. وتشمل الآتي:

- لزيادة الابتكار
- لسرعة وكفاءة حل المشكلات
- لتحسين الخدمات



استخدامات الذكاء الاصطناعي في المجالات والقطاعات المختلفة

أصبح الذكاء الاصطناعي يستخدم بشكل متزايد في مختلف المجالات والقطاعات، في المجالين العام والخاص. يوضح الجدول 21.1 القطاعات الرئيسية التي كان للذكاء الاصطناعي فيها بالفعل تأثير واسع النطاق؛ أما القطاعات الأخرى التي ترك فيها الذكاء الاصطناعي أيضاً انطباعاً فيرد توضيحها في الجدول 21.2.

الجدول 21.1 تطبيقات الذكاء الاصطناعي في الصناعات والخدمات الرئيسية

البيع بالتجزئة والتجارة الإلكترونية	النقل والخدمات اللوجستية	الشؤون المالية	الخدمات العامة
<ul style="list-style-type: none"> توصية مخصصة للمنتج بناءً على النشاط السابق كشف الاحتيال - فحص المعاملات لتحديد الاحتيال إدارة المخزون - استخدام نهج الإنتاج في الوقت المناسب لتقليل تكاليف التخزين روبوتات المحادثة والمساعدين الافتراضيين - يقدمون الدعم على مدار الساعة طوال أيام الأسبوع 	<ul style="list-style-type: none"> التخطيط اللوجستي، بما في ذلك التنبؤ بالطلب لتخطيط السعة تحسين مسار النقل - التخطيط المسبق لطرق التسليم لتكون فعالة قدر الإمكان التخزين الآلي مع انتقاء العناصر ومعالجتها آلياً استخدام الطائرات دون طيار لتوصيل للطرود الصغيرة إدارة الشحن لتحسين حمولة الحاويات 	<ul style="list-style-type: none"> التداول الآلي للأسهم والحصص - الشراء عندما تكون الأسعار منخفضة والبيع عندما تكون الأسعار مرتفعة التنبؤ بالسوق - استخدام بيانات العملاء والمخاطر والمعاملات للتنبؤ بالنتائج المستقبلية. يسهم هذا الأمر أيضاً في اكتشاف الاحتيال وإدارة المخاطر الخدمات المصرفية المخصصة واكتشاف الأعطال 	<ul style="list-style-type: none"> التشغيل الآلي للخدمات العامة باستخدام روبوتات المحادثة في المقام الأول، والتصعيد إلى البشر إذا لزم الأمر التنبؤ بالأزمات الصحية العامة - لقد تم القيام بذلك لسنوات عديدة في التخطيط للاستجابة للإنفلونزا الشتوية تخصيص الموارد بكفاءة تعزيز الكشف عن الاحتيال والأخطاء في أنظمة الضرائب والمزايا والمنح

الجدول 21.2 تطبيقات الذكاء الاصطناعي في القطاعات والخدمات الأخرى

التثقيف والتوعية	الرعاية الصحية	الوسائط	البنية التحتية والخدمات
<ul style="list-style-type: none"> التعلم المخصص مع خطط التعلم الفردية القائمة على المهارات الموجودة مسبقاً برنامج تصحيح تلقائي للمعلمين واللجان الامتحانية فرص التعلم عن بعد لنقل مهارات القراءة والكتابة والحساب إلى مجموعة أوسع من الطلاب دعم المتعلمين ذوي الإعاقة - على سبيل المثال، الكتب الدراسية الرقمية المدعومة بالذكاء الاصطناعي لمجموعة متنوعة من مستويات كفاءة المتعلمين 	<ul style="list-style-type: none"> خدمات التشخيص - غالباً ما تستخدم أنظمة الفرز في المستشفى الذكاء الاصطناعي لدعم الاتصال الأولي مع المرضى الأدوية والخطط العلاجية المخصصة - تقديم رؤى قابلة للتنفيذ بناءً على تحليل بيانات المرضى وبالتالي تحسين نتائج المرضى المراقبة عن بُعد للمرضى - لقد قللت الحاجة إلى الاستشفاء، ووفرت الفائدة الإضافية بأن المرضى يتحسنون بشكل أسرع لأنهم في المنزل 	<ul style="list-style-type: none"> المساعدة على تطوير الأفلام والبرامج التلفزيونية والكتب وألعاب الفيديو أنظمة التوصية لخدمات البث، واختيار التوصيات بناءً على المشاهدة السابقة حماية المحتوى للقاصرين: مراقبة المحتوى الذي يمكن أن ينشر الكراهية والقالب النمطية وأشكال التحيز الأخرى الفن والموسيقى وغيرهما من إنتاج الذكاء الاصطناعي. ابتكار عمل يحاكي عمل الفنانين. تأليف الموسيقى والتناغمات المصممة خصيصاً لأنواع أو حالات مزاجية أو سياقات محددة، استناداً إلى أعمال أخرى 	<ul style="list-style-type: none"> الحفاظ على الطاقة باستخدام الشبكات الذكية والأجهزة المتصلة لتحسين توزيع الطاقة أدوات إدارة حركة المرور، مثل تحليل أنماط حركة المرور لتحسين المسارات وتقليل استخدام الوقود وإدارة أعداد المركبات على الطرق الرئيسية تحديد الأعطال وتحديد المشكلات في شبكات الطاقة وشبكة السكك الحديدية والشبكات الرقمية صيانة البنية التحتية - تستخدم في الهندسة المدنية لمنع تجاوز التكاليف وتقليل مخاطر الحوادث

فكر في كيفية تواصلك الشخصي مع الذكاء الاصطناعي. افحص الأمثلة في الفئات الثماني المختلفة في الجدولين 21.1 و21.2.

وقفعة للتفكير



حدد أكبر عدد ممكن من مثيلات الذكاء الاصطناعي التي اتصلت بها.

تلميح

قم بإنشاء خريطة ذهنية للأمثلة وشارك خريطتك الذهنية مع طالب آخر. ما أوجه التشابه والاختلاف بين الخريطين؟

توسيع الأفق

2 الآثار المترتبة على الذكاء الاصطناعي

يمكن أن تكون المزايا العملية لاستخدام الذكاء الاصطناعي كبيرة. يمكن للذكاء الاصطناعي:

- إزالة احتمال حدوث خطأ بشري، ما ينتج عنه نتائج أكثر دقة
 - يمكن استخدامه في المناطق التي قد تكون فيها المهام خطيرة على حياة الإنسان، مثل التخلص من القنابل وصيانة خطوط الأنابيب تحت الماء
 - يمكن استخدامه لأتمتة الوظائف المتكررة
 - توفير الخدمات على مدار الساعة طوال أيام الأسبوع (على سبيل المثال، يمكن لروبوتات الدردشة المدعومة بالذكاء الاصطناعي تقديم خدمات دعم العملاء خارج ساعات العمل العادية)
 - تعمل مع كميات أكبر كثيرًا من البيانات بسرعات عالية بشكل لا يصدق.
- ولكن ما تزال هناك بعض العيوب العملية، منها:
- العديد من تطبيقات الذكاء الاصطناعي تعتبر مكلفة.
 - الذكاء الاصطناعي غير قادر على استخدام الإبداع أو العاطفة في عملية صنع القرار.
 - تقلل الأتمتة توافر الوظائف، خاصة لمن يبحثون عن وظائف تتطلب إلى القليل من المهارات.
 - الأخلاقيات هي إحدى أكبر المجالات التي تسترعي الاهتمام – يسأل الكثير من الناس السؤال الآتي: (هل يمكننا فعل ما نريد طالما أنه باستطاعتنا فعله بالذكاء الاصطناعي؟)

المهارات

- مهارات التواصل الشخصي: الانفتاح
الفكري:
- المسؤولية الشخصية والاجتماعية
- مهارات التواصل الشخصي: أخلاقيات
العمل/الضمير:
- الأخلاقيات
 - النزاهة

وقفة للتفكير

ما رأيك في فوائد تقديم خدمة للعملاء على مدار الساعة طوال أيام الأسبوع باستخدام الذكاء الاصطناعي؟

تلميح

ماذا سيفعل العملاء إذا لم تكن هذه الخدمة متاحة؟

توسيع الأفق

هل هناك أي عيوب في الوصول إلى خدمة العملاء على مدار الساعة طوال أيام الأسبوع؟ فكر في تأثير إتاحة المعلومات بسهولة في جميع الأوقات.

بحث

ماذا تعرف عن المعضلات الأخلاقية المرتبطة بالذكاء الاصطناعي؟ استخدم الإنترنت وابحث عن مقاليتين حديثتين يمكنك استخدامهما لإثراء المناقشة حول هذه المشكلة.

الاعتبارات الأخلاقية والقانونية عند استخدام الذكاء الاصطناعي

ترتبط الاعتبارات الأخلاقية والقانونية، وتتداخل إلى حد ما، في أمور قد يشوبها بعض القلق. يجب معالجتها عند النظر في كيفية تطوير أنظمة الذكاء الاصطناعي ونشرها واستخدامها في المجتمع. وهي تغطي الجوانب الآتية:

أمان البيانات والخصوصية

غالبًا ما تعالج مشاريع الذكاء الاصطناعي البيانات ذات الطبيعة الحساسة أو الشخصية، ما يجعلها أهدافًا واضحة للهجمات الإلكترونية وغيرها من انتهاكات البيانات.

لذلك فإن ضوابط البيانات هي اعتبار بالغ الأهمية، ويجب وضعها لضمان سلامة البيانات (دقتها) وسريتها (بتقييد الوصول). تشمل تقنيات حماية البيانات:

- التشفير
- التحكم المنطقي في الوصول
- الأمان المادي.

ستمنع هذه الضوابط أيضًا إفساد مشاريع الذكاء الاصطناعي من خلال التغيير الضار للبيانات المستخدمة لتدريبها. يمكن أن تقود مثل هذه التعديلات أنظمة الذكاء الاصطناعي المتأثرة إلى اتخاذ قرارات غير صحيحة قد تحقق فائدة.

نظرًا إلى أن الذكاء الاصطناعي يمكن أن يولد رؤى جديدة (وأحيانًا غير متوقعة) من البيانات الموجودة، يجب توخي الحذر لعدم دمج معلومات تحديد الهوية الشخصية (PII) عن طريق الخطأ مع مجموعات البيانات التي تم إخفاء هويتها. إذا لم يتم ذلك، فقد يُساء استخدام الاكتشافات المتعلقة بالأفراد إذا وقعت في الأيدي الخطأ.

يوجد في أوروبا تشريع شامل لأمن البيانات يسمى اللائحة العامة لحماية البيانات (GDPR). دخل هذا التشريع حيز التنفيذ في مايو 2018 لدول الاتحاد الأوروبي (EU) والمنطقة الاقتصادية الأوروبية (EEA). هدفه الأساسي هو منح الأفراد السيطرة الرئيسية على بياناتهم الشخصية. كما أنه يوفر فائدة إضافية تتمثل في توحيد لوائح البيانات للأعمال التجارية الدولية.

كان لللائحة العامة لحماية البيانات تأثير كبير في أوروبا وعلى الشركات التي تعالج البيانات الشخصية لمواطني الاتحاد الأوروبي. بموجب التشريع، تخضع مشاريع الذكاء الاصطناعي والبيانات التي تستخدمها للمتطلبات القانونية والتدقيق مثل أي مشروع تكنولوجيا معلومات آخر يستخدم بيانات الأفراد. لدى الولايات المتحدة تشريعات مماثلة لللائحة العامة لحماية البيانات الأوروبية، وكذلك اليابان. بالإضافة إلى ذلك، أدخل الأردن تشريعات مماثلة لحماية البيانات في عام 2023.

موضوعات ذات صلة

انظر الوحدة 9: إدارة مشاريع تكنولوجيا المعلومات لمعرفة المزيد عن التقييم القانوني للمشروع.

ماذا تعرف عن تشريعات حماية البيانات في بلدك؟

وقفة للتفكير

هل تعرف ما حقوقك بموجب تشريعات حماية بياناتك؟

تلميح

هل تعتقد أن التشريع فعال؟

توسيع الأفق

التحيز والتمييز

تعد قضايا التحيز والتمييز مصدر قلق أخلاقي أساسي للذكاء الاصطناعي، لا سيما عند استخدامه في مجالات حساسة مثل العدالة والتعليم. هذا بسبب خطر إمكانية تدريب أنظمة الذكاء الاصطناعي باستخدام بيانات متحيزة. يمكن أن يترجم هذا إلى قرارات تمييزية يتم اتخاذها ضد فئات معينة من الأفراد.

لذلك من الضروري التحقق من عدم وجود تحيز داخل نظام الذكاء الاصطناعي وضمان عدم التمييز ضد الأشخاص بسبب بعض الخصائص المحمية، بما في ذلك العرق والدين (أو المعتقد) والجنس والعمر والإعاقة والحالة الاجتماعية. إذا حدث هذا داخل منظمة، فقد يكون له آثار سلبية على التوظيف والإقراض والتأمين (خاصة التأمين الصحي والتأمين على الحياة) والشرعية.

اتخاذ القرار في الأنظمة التي تعد السلامة أمرًا حيويًا لها

تعتمد أنظمة الذكاء الاصطناعي، خاصة تلك التي تستخدم التعلم الآلي، بشكل كبير على بيانات التدريب الخاصة بها. إذا كانت هذه البيانات محدودة النطاق أو لا تمثل المجال أو معيبة، فقد يتخذ الذكاء الاصطناعي قرارات غير صحيحة وربما خطيرة.

- قد لا تميز السيارات ذاتية القيادة المشاة الذين يعبرون الطريق.
- قد لا تقرأ أنظمة التحكم الآلي في الفيضانات معدل ارتفاع مستوى المياه بشكل صحيح.
- قد تفقد أنظمة التشخيص الطبي علامات العدوى والأمراض الرئيسية في الدم وعينات الأنسجة والفحوصات. قد يؤدي ذلك إلى غياب أو عدم صحة تشخيص المرض.
- قد تحدث الحوادث الجوية بسبب قيام الأنظمة بتفسير البيانات بشكل غير صحيح من أجهزة الاستشعار وعدم اتخاذ إجراءات تصحيحية.

لقد حدثت بالفعل العديد من هذه الإخفاقات. وقد أسفرت عن قضايا قانونية لإثبات اللوم والمسؤولية. على سبيل المثال، تعرضت سيارة ذاتية القيادة لحادث أدى إلى مقتل امرأة في الولايات المتحدة الأمريكية. حدث هذا الأمر عند عدم تمييز أنظمة الذكاء الاصطناعي في السيارة المرأة على أنها أحد الأشخاص في أثناء عبورها الطريق. في ذلك الوقت كان هناك راكب في السيارة كان يجب أن يعمل كسائق أمان. كان ينبغي لهذا الشخص أن يتخذ الخطوات المناسبة لمنع الحادث. ومع ذلك، لم يكونوا منتبهين لأنهم كانوا يشاهدون برنامج تلفزيوني.

بعد التحقيق، تم اتهام سائق السلامة بالقتل بسبب الإهمال (التسبب في الوفاة نتيجة نقص التركيز). ولكن من كان المخطئ حقًا؟ هل يتحمل مطورو نظام الذكاء الاصطناعي أي مسؤولية؟

بحث

هناك العديد من الأمثلة الموثقة لأنظمة تعد السلامة أمرًا حيويًا لها. ولا تعمل على النحو المطلوب. ابحث عن ثلاثة أمثلة من صناعات مختلفة واكتب مقالًا قصيرًا عن استخدام الذكاء الاصطناعي في الأنظمة التي تعد السلامة أمرًا حيويًا لها.

التأثير الاجتماعي والاقتصادي

من المرجح أن يؤثر الذكاء الاصطناعي بعدة طرق في الآليات الاقتصادية التي تدفع النمو المالي والإنتاجية، وفي المجتمع بأكمله.

- **الإزاحة الوظيفية:** سيتولى الذكاء الاصطناعي المهام التي يؤديها الموظفون عادةً بشكل متزايد (على سبيل المثال، دعم العملاء). في المقابل، فإشياء أدوار جديدة متعلقة بالذكاء الاصطناعي سيحدث بلا شك، ما يؤدي إلى تحول في الفرص التعليمية والوظيفية.
- **المؤهل التعليمي:** أتاح الذكاء الاصطناعي فرص التعلم عن بعد. وقد أدى ذلك إلى فتح باب التعليم للأفراد الذين ربما لم تتاح لهم الفرصة للتعلم.
- **عدم المساواة في الثروة:** من المرجح أن يكون القسم الأول من المجتمع الذي يستفيد من أنظمة وخدمات الذكاء الاصطناعي هم أولئك الذين يتمتعون بالفعل بالأمان المالي. ونتيجة لذلك، هناك احتمال أن يتسبب الذكاء الاصطناعي في توسيع الفجوة بين الأغنياء والفقراء.
- **الصحة:** أدت مراقبة المريض عن بُعد إلى تقليل الحاجة إلى دخول المستشفى لحالات معينة. وقد حقق هذا فوائد حقيقية حيث يميل العديد من المرضى إلى التحسن بشكل أسرع إذا تم الاعتناء بهم في منازلهم. كانت هذه التكنولوجيا مفيدة على وجه الخصوص في جائحة COVID-19، لا سيما لمن يعانون حالات كبت المناعة، والذين كانوا معرضين بشكل خاص لخطر الإصابة بالمرض. وبما أنهم لم يضطروا إلى الذهاب إلى المستشفى، لم يضطروا إلى تعريض أنفسهم للعدوى.

ما رأيك في أخلاقيات فقدان الوظائف عند إسناد العمل الذي كان يقوم به الموظفون سابقاً إلى الذكاء الاصطناعي؟

وقفة للتفكير



فكر في الفرص التي قد تصاحب ذلك، بدلاً من السلبيات.

تلميح

هل يجب أن يشارك المجتمع ككل في النقاش؟

توسيع الأفق

الاستقطاب

لسوء الحظ، كما هو الحال في معظم التقنيات، يمكن للأشخاص والمنظمات عديمة الضمير استخدام الذكاء الاصطناعي لخلق الاستقطاب في المجتمع. يمكن أن يحدث هذا أيضاً عن غير قصد. يرد في ما يأتي استخدامان نموذجيان للذكاء الاصطناعي يمكن أن يخلقاً فجوة أكبر بين القطاعات المختلفة في المجتمع، بل ويقوضان العملية الديمقراطية.

- 1 الخوارزميات التي تستخدمها المنصات الإعلامية التي تروج للحقائق (الكاذبة)، والتي تعزز معتقدات الناس وتستبعد وجهات النظر المعارضة.
- 2 الدعاية المستهدفة التي تستخدم وجهات النظر والتفضيلات والميول المحددة للشخص للتأثير في آرائه السياسية.

مبادئ الذكاء الاصطناعي القابل للتفسير (explainable AI/XAI)

يلعب الذكاء الاصطناعي دوراً مهماً بشكل متزايد في حياتنا، سواء كنا على علم بذلك أم لم نكن، لذلك تستدعي الحاجة الملحة طمأنة الناس بأن الذكاء الاصطناعي آمنٌ وموثوقٌ و(الأهم من ذلك) أن بإمكانه إصدار أحكام لا تتأثر بالتحيز أو التحامل.

وهنا يأتي دور **الذكاء الاصطناعي القابل للتفسير (explainable AI/XAI)**، وهو هدف عام يستهدفه معظم الأطراف الرئيسيين في مجال الذكاء الاصطناعي، لهذا التوجه هدفان رئيسيان.

- **الشفافية** – يهدف الذكاء الاصطناعي القابل للتفسير إلى تيسير فهم عمليات صنع القرار لأنظمة الذكاء الاصطناعي بحيث يمكن فهمها. وبعبارة أخرى، فإنه يهدف إلى السماح للمستخدمين بفهم كيف (والأهم من ذلك، لماذا) تم اتخاذ قرار أو تنبؤ معين.
- **الامتثال للوائح** - من المهم أن يتوافق الذكاء الاصطناعي مع المتطلبات القانونية والتنظيمية. يساعد الذكاء الاصطناعي القابل للتفسير على إظهار ذلك.

الشفافية والامتثال مهمان بشكل خاص في تطبيقات الذكاء الاصطناعي التي تتطلب المساواة - على سبيل المثال، عندما يتم استخدام الذكاء الاصطناعي في عمليات صنع القرار الآلية المهمة المرتبطة بالتوظيف والتقييم المالي والقضايا القانونية.

المصطلح الرئيس

الذكاء الاصطناعي القابل للتفسير (XAI) – الأساليب والتقنيات التي تجعل مخرجات نماذج الذكاء الاصطناعي مفهومة للبشر.

المصطلح الرئيس

العائد على الاستثمار (RoI) – النسبة
بين صافي الدخل والاستثمار، ويشير
أيضًا إلى الفحص الذي يتم للتأكد من
أن الفوائد المالية لتنفيذ حل الذكاء
الاصطناعي تزيد على التكاليف التي تم
تكبدتها لإنشائه.

- يرد في ما يأتي اثنين من الآثار الجانبية الجديرة بالملاحظة للذكاء الاصطناعي القابل للتفسير.
 - تحسين التفاعل بين الذكاء الاصطناعي والبشر - يحدث عندما يكون من الأسهل فهم وتفسير وتصحيح (إذا لزم الأمر) الأسباب وراء قرارات الذكاء الاصطناعي.
 - يمكن لعلماء البيانات تحسين النماذج بشكل أكثر فعالية عندما يفهمونها بشكل كامل. وهذا يساعدهم على تحديد أي تحيزات أو أخطاء أو أوجه قصور في خوارزميات نماذجهم.
- كلما زادت ثقة أصحاب المصلحة في الذكاء الاصطناعي، زادت احتمالية استثمار الناس فيه، وهذا الأمر مهم لأن تطوير مثل هذه الأنظمة يمكن أن يكون مكلفًا للغاية. لذلك **فالعائد على الاستثمار (RoI)** لحلول الذكاء الاصطناعي هو أحد الاعتبارات الهامة؛ والعائد على الاستثمار هو القيمة التي تنتجها حلول الذكاء الاصطناعي مقارنةً بالتكاليف المطلوبة لإنشائه فعليًا.

مناقشة

ماذا يحدث عندما يتضمن الذكاء الاصطناعي تحيزًا غير مرغوب فيه؟
ناقش ما قد يحدث إذا بدأ أن نظام الذكاء الاصطناعي يتخذ قرارات سيئة. ماذا يحدث إذا لم يكن هذا النظام والقرارات التي يتخذها قابلة للتفسير بسهولة؟
من خلال العمل في أزواج أو مجموعات صغيرة، تحقق من أداة الذكاء الاصطناعي التي تعتقد أنها ربما تضمنت تحيزًا غير مرغوب فيه. ناقش المشكلات التي تسببها الأداة وأي نتائج نهائية لاستخدامها.

عملية تطوير الذكاء الاصطناعي/العمليات الجارية للذكاء الاصطناعي

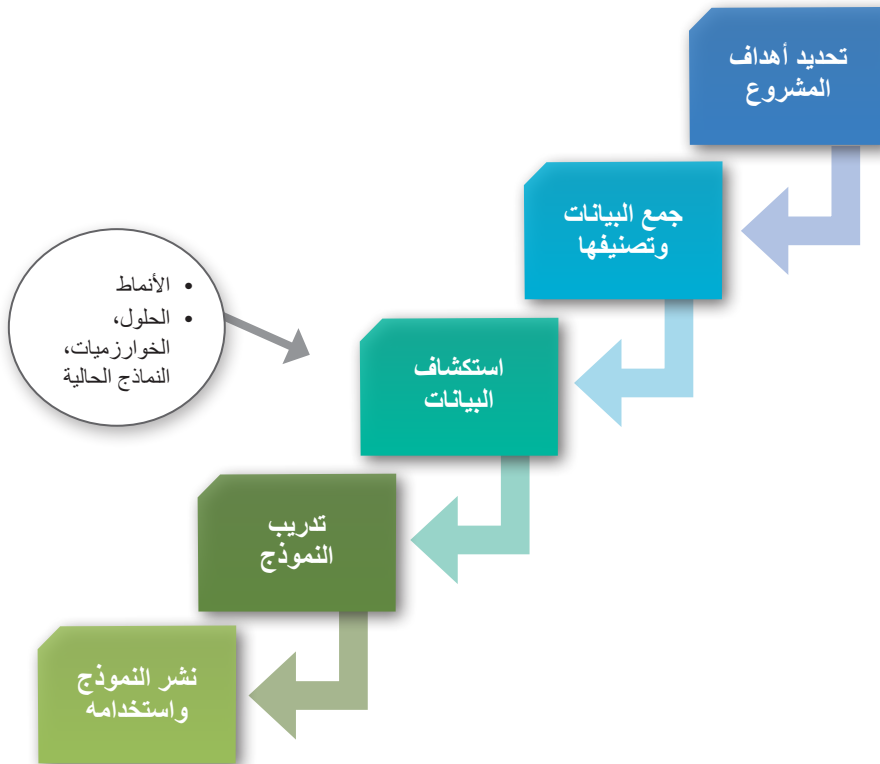
لضمان نجاح أي مشروع للذكاء الاصطناعي، فإن أفضل الممارسات المقبولة هي اتباع سير عمل أو مسار شامل وموثق جيدًا للمشروع. توضح هذه العملية الخطوات المطلوبة في تطوير ونشر وصيانة نماذج الذكاء الاصطناعي.

من الجدير بالذكر أن هناك العديد من الإصدارات المختلفة لمثل هذه المخططات، حيث غالبًا ما تكون للمراحل أسماء مختلفة وأحيانًا تظهر بعض التداخل. ومع ذلك، فإن المهام والمبادئ الرئيسية لتطبيق نهج منظم متسقة، ويوضح الشكل 21.4 نموذجًا لمخطط ذكاء اصطناعي.

المهارات

المهارات المعرفية: الإبداع:

- الإبداع
- مهارات التواصل الشخصي: الانفتاح
- الفكري:
- القدرة على التكيف



الشكل 21.4 يتبع مسار الذكاء الاصطناعي النموذجي نموذج تطوير الشلال، حيث تكتمل كل مرحلة قبل أن تبدأ المرحلة التالية - على الرغم من أنه في بعض الأحيان قد يتراجع التقدم خطوة إذا لزم الأمر.

الحاجة إلى اتباع عملية تطوير منظمة للذكاء الاصطناعي/مسار للذكاء الاصطناعي

يوجد خمس مزايا رئيسية لمتابعة عملية تطوير منظمة للذكاء الاصطناعي.

- 1 يساعد على موازنة مشاريع الذكاء الاصطناعي مع الأهداف التنظيمية.
- 2 يزيد الكفاءة إلى أقصى حد.
- 3 يضمن جودة البيانات.
- 4 يحقق دقة أكبر للنموذج.
- 5 يحافظ على المعايير الأخلاقية.

الجدول 21.3 تأثيرات وفوائد كل مرحلة من مراحل مسار الذكاء الاصطناعي النموذجي

المرحلة	التأثيرات	الفوائد
تحديد أهداف المشروع	<ul style="list-style-type: none"> - وضوح الأهداف؛ يحل المشكلات المعنية بالأساليب الصحيحة - يحدد توقعات أصحاب المصلحة - يعمل على موازنة الموارد لتحقيق النتائج المرجوة 	<ul style="list-style-type: none"> - المشروع لديه اتجاه واضح ودعم من أصحاب المصلحة - تخصيص الموارد بكفاءة - ارتفاع احتمالية النجاح ورضا أصحاب المصلحة
تدقيق البيانات	<ul style="list-style-type: none"> - يضمن كمية البيانات وجودتها (قبل كل شيء) - يضمن عدم جمع بيانات منخفضة الجودة - يمنع التحيز أو النماذج غير الدقيقة ويحافظ على الثقة والمصداقية 	<ul style="list-style-type: none"> - بيانات عالية الجودة - الحد من التحيز - تطوير نموذج قوي - نتائج محسنة - زيادة الثقة بين أصحاب المصلحة والمستخدمين النهائيين
البحث	<ul style="list-style-type: none"> - يحدد خصائص البيانات/الميزات ذات الصلة بالمشكلة - يسلط الضوء على نقاط قوة البيانات وحدودها - يعطي معلومات تفيد في تصميم النموذج 	<ul style="list-style-type: none"> - يساعد في إنشاء نماذج أكثر دقة - يزيد من فعالية النموذج - يزيد الثقة بين أصحاب المصلحة والمستخدمين النهائيين
تصميم وبناء النموذج	<ul style="list-style-type: none"> - يستهلك موارد المنظمة خلال مرحلة التدريب - يحدد الأداء النهائي لحل الذكاء الاصطناعي - يؤثر بشكل مباشر في دقة النتائج المتوقعة للنموذج ونزاهتها وشفافيتها 	<ul style="list-style-type: none"> - تعزز النماذج الجيدة إمكانات صنع القرار - يساعد على أتمتة المهام الروتينية - يعزز العائد على الاستثمار لأصحاب المصلحة - يمنح المؤسسة (والمستخدمين) ميزة تنافسية
نشر النموذج واستخدامه	<ul style="list-style-type: none"> - يدمج قدرات الذكاء الاصطناعي في العمليات التشغيلية - قد يغير سير العمل - قد يتطلب مهارات واستثمارات جديدة - يؤثر في عمليات اتخاذ القرار 	<ul style="list-style-type: none"> - يقدم تحسينات للإنتاجية - يعزز عملية صنع القرار - يوفر إمكانات لمزيد من الابتكار - يحصل أصحاب المصلحة على نتائج محسنة والمزيد من الفرص

المهارات

المهارات المعرفية: العمليات والإستراتيجيات المعرفية:

- الاستدلال المنطقي/الحجة
- اتخاذ القرار

المهارات المعرفية: الإبداع:

- الابتكار

مهارات التواصل الشخصي: أخلاقيات

العمل/الضمير:

- الإنتاجية

العوامل التنظيمية التي يجب مراعاتها عند استخدام الذكاء الاصطناعي

عندما تستخدم إحدى مؤسسات الذكاء الاصطناعي، من المهم أن نتذكر أن العديد من الناس يكرهون التغيير بطبيعتهم. وبالتالي، فمن المحتمل أن يكون من الصعب إدخال الذكاء الاصطناعي في نطاق واسع واعتماده في المؤسسة.

ومع ذلك، فإن طرح الأسئلة الصحيحة ووضع خطة منظمة بشكل معقول لتقديم حلول وخدمات الذكاء الاصطناعي هي خطوات في الاتجاه الصحيح.



عند التفكير في اعتماد الذكاء الاصطناعي، يتعين على المؤسسات ثلاثة عوامل حاسمة للنظر فيها:

- القدرة التجارية
- القدرة البشرية
- القدرة الفنية.

القدرة التجارية

هي قدرة المؤسسة على تحقيق أهدافها من خلال التخطيط الجيد، والاستخدام الأمثل للموارد المتاحة لديها، والطريقة التي تنفذ بها خططها الإستراتيجية.

يغطي رأس المال التجاري المجالات التشغيلية الآتية: المشاركة الإستراتيجية والثقافية ومشاركة أصحاب المصلحة. تتضمن الأسئلة الرئيسية التي تركز على الذكاء الاصطناعي في كل مجال ما يأتي.

الإستراتيجية

- كيف يتماشى الذكاء الاصطناعي مع الأهداف الإستراتيجية الرئيسية للمؤسسة؟
- ما المشكلات التي يمكن للذكاء الاصطناعي حلها؟
- الذكاء الاصطناعي مكلف: كيف تحدد المؤسسة عائد الاستثمار؟
- كيف يمكن للذكاء الاصطناعي أن يقود الميزة التنافسية؟
- كيف يمكن للذكاء الاصطناعي تحسين تجربة العملاء؟
- كيف يمكن للذكاء الاصطناعي توليد تدفقات جديدة من الإيرادات؟

عوامل ثقافية

- هل ستستغل المؤسسة الفرص التي يقدمها الذكاء الاصطناعي؟
- هل يمكن لإدارة التغيير أن تتكيف مع التغيير المدفوع بالذكاء الاصطناعي؟
- كيف ستتعاامل المؤسسة مع القوى العاملة وتغيرات سير العمل؟

إشراك أصحاب المصلحة

- هل ستقبل القيادة والإدارات الرئيسية والموظفون حلول الذكاء الاصطناعي وخدماته؟
- هل سيتفق العملاء في هذه الحلول القائمة على الذكاء الاصطناعي؟
- عادةً ما يتم تعزيز قدرة الأعمال من خلال تطبيق الذكاء الاصطناعي، والذي يمكنه:
- تحسين العمليات التجارية (جعلها أكثر كفاءة)
- تحسين عملية صنع القرار (جعلها أكثر سرعة وأكثر استنارة)
- دفع الابتكار وتطوير الأفكار الجديدة.
- تتضمن أمثلة استخدام الذكاء الاصطناعي لتعزيز قدرة الأعمال ما يأتي:
- استخدام روبوتات المحادثة لتحسين خدمات العملاء
- التصنيع الآلي الموجه نحو الروبوت ومراقبة الجودة
- تحسين إدارة المخاطر من خلال التنبؤ بشكل أفضل بالإخفاقات المحتملة (والمكلفة)
- إنشاء رؤية جديدة لسلوك العملاء لفهم اتجاهات السوق بشكل أفضل.

القدرة البشرية

ويشمل ذلك المعرفة والمهارات والخبرات الجماعية التي تمتلكها القوى العاملة في المؤسسة. إنه مقياس لمدى نجاحهم في أداء المهام وحل المشكلات وابتكار أفكار جديدة. يشمل رأس المال البشري المجال التشغيلي الآتي: المواهب والخبرة الداخلية. تتضمن الأسئلة الرئيسية التي تركز على الذكاء الاصطناعي ما يأتي.

المواهب والخبرة الداخلية

- هل لدينا الموظفين المهرة لتقديم مشاريع الذكاء الاصطناعي داخل الشركة؟
- ما المهارات التي يجب تطويرها؟
- ما مستوى الاستثمار المطلوب؟
- ما نوع المواهب الجديدة التي نحتاج إلى جذبها/توظيفها؟



ستتأثر القدرات البشرية للمؤسسة تأثيرًا مباشرًا بالذكاء الاصطناعي. على الأقل، سيعيد الذكاء الاصطناعي تشكيل الأدوار الحالية من خلال تعزيز المهارات الحالية، أو (للأسف) سيستبدل تلك الأدوار التي تُعتبر زائدة عن الحاجة. لحسن الحظ، سيتم إنشاء أدوار جديدة تتعلق بأدوات الذكاء الاصطناعي المستخدمة. هذا شكل من أشكال النزوح الوظيفي وهو شائع عندما يحدث أي نوع من الثورات الصناعية.

يمكن للذكاء الاصطناعي تعزيز القدرات البشرية للمؤسسة من خلال:

- تقليل المهام المتكررة وتحرير الموظفين ليكونوا أكثر إبداعًا وابتكارًا
 - تسهيل التواصل بشكل أفضل بين أعضاء الفريق
 - توفير ملخصات سريعة للمعلومات الأساسية من مصادر وتنسيقات مختلفة
 - إنشاء برامج تعليمية مخصصة لسد فجوات المهارات المحددة
 - تحسين رفاهية الموظفين من خلال اقتراح تغييرات إيجابية على سير العمل وبيئات العمل.
- يجب أن تكون النتيجة قوة عاملة أكثر ديناميكية وتكيفًا وإبداعًا.

القدرة الفنية

القدرة الفنية تصف القدرات الفنية والموارد التي تستخدمها المؤسسة لدعم عملياتها ومشروعاتها الحالية، وتطوير منتجات أو خدمات جديدة، والتكيف مع التغييرات في قطاعها. تشمل القدرات الفنية المجالات التشغيلية الآتية: البنية التحتية التكنولوجية وإدارة البيانات. تتضمن الأسئلة الرئيسية التي تركز على الذكاء الاصطناعي ما يأتي.

البنية التحتية التكنولوجية

- هل لدينا حواسيب كافية؟
- هل لدينا مستويات كافية من تخزين البيانات؟
- هل ستكون الأنظمة الحالية قادرة على الاندماج مع أنظمة الذكاء الاصطناعي الجديدة؟

إدارة البيانات

- كيف سيتم جمع بيانات التدريب/الاختبار المطلوبة؟
- كيف وأين سيتم تخزين هذه البيانات؟
- كيف سيتم إدارة هذه البيانات وجعلها آمنة؟

يمكن للذكاء الاصطناعي تعزيز القدرة التقنية للمؤسسة بعدة طرق:

- يمكن أن تكون الموارد التكنولوجية الحالية غير مستغلة كثيرًا في العديد من المنظمات. يجب أن تعزز مقدمة الذكاء الاصطناعي كيفية استخدام التكنولوجيا، ومدى جودة (وكفاءة) استخدامها.
- يمكن أن يكون للذكاء الاصطناعي تأثير كبير في تحليل البيانات عن طريق معالجة كميات هائلة من البيانات بسرعة أكبر وبدقة أعلى من الأساليب التقليدية. ويمكن أن يمنح هذا الأمر المؤسسات رؤى أعمق كثيرًا مما تنتجها بياناتها الحالية، ما يعزز قدرتها على تحديد الاتجاهات الجديدة، والتنبؤ بنتائج المشاريع واتخاذ قرارات أكثر استنارة.
- وكذلك يمكن أتمتة العديد من المهام التنظيمية اليومية المعقدة بشكل أكثر فعالية باستخدام الذكاء الاصطناعي. تتراوح هذه المهام من الجوانب العملية للتصميم والتصنيع إلى المشاركة مع العملاء.
- من منظور أجهزة وبرامج تكنولوجيا المعلومات، يمكن للذكاء الاصطناعي المساعدة على مراقبة استخدام الموارد الحالية، ما يؤدي إلى صيانة أكثر تنظيمًا واستهدافًا (تقليل حالات الفشل وحالات التعطل وفقدان الخدمة).
- يمكن أيضًا تحسين الأمن السيبراني باستخدام أنظمة تعتمد على الذكاء الاصطناعي لاكتشاف الحالات الشاذة التي ربما لم تجدها الأنظمة التي يديرها البشر ببساطة.

مناقشة

بصفتك موظفًا مستقبليًا، راعِ التأثير المحتمل للذكاء الاصطناعي في مؤسسة قد تعمل بها. كيف يمكن لمؤسستك استخدام الذكاء الاصطناعي؟ هل ستتقبل أنت وزملاؤك الموظفون الذكاء الاصطناعي؟ كيف سيؤثر ذلك في عملك؟ ما العوامل التي قد تمنع مؤسستك من استخدام الذكاء الاصطناعي؟ ناقش هذه القضايا مع زملائك.

فوائد الذكاء الاصطناعي في مختلف القطاعات

بشكل عام، يجب أن تستفيد جميع القطاعات من استخدام الذكاء الاصطناعي، حيث يمكن أن يساعد على تقديم:

- اتخاذ قرارات أفضل
 - تحسين الكفاءة من خلال التشغيل الآلي للمهام الروتينية
 - تخفيض التكاليف
 - إضفاء الطابع الشخصي على تجربة العملاء
 - المزيد من الابتكار (على سبيل المثال، من خلال تحديد المنتجات أو الخدمات الجديدة).
- كما سيجقق الذكاء الاصطناعي عددًا من المكاسب الخاصة بكل قطاع.

الشؤون المالية

لقد أفاد الذكاء الاصطناعي بالفعل القطاع المالي من خلال توفير اكتشاف محسن للاحتيال والوقاية منه. كما ساعد على حماية العملاء بشكل أكثر فعالية من خلال مراقبة المعاملات وتحديد الأنشطة غير العادية.

الرعاية الصحية

استفادت الرعاية الصحية من الذكاء الاصطناعي لأنه أدى إلى إنشاء خطط علاج شخصية يمكن أن تؤدي إلى فعالية أفضل ونتائج محسنة. على سبيل المثال، يُستخدم الذكاء الاصطناعي في نطاق واسع في إنشاء برامج العلاج الكيميائي المخصصة التي تستهدف بشكل فعال السرطان الدقيق الذي يعانيه المريض، ما يؤدي إلى تحسين فرص النجاة.

التثقيف والتوعية

أتاح الذكاء الاصطناعي إمكانية تصميم تجارب التعلم لتناسب مع احتياجات المتعلم. يمكن للذكاء الاصطناعي أن يأخذ في الاعتبار أسلوب التعلم المفضل للمتعلم. يمكن أن يضمن أيضًا تقديم التجربة بأكثر الطرق فعالية.

الوسائط

يوفر الذكاء الاصطناعي فوائد متعددة لقطاع الإعلام. على سبيل المثال، يُمكن خدمات البث من تخصيص المحتوى ليتوافق مع التفضيلات والسلوكيات وأنماط المشاركة للعملاء. عادةً ما يتم تحليل سجل خدمة البث لكل فرد لتقديم توصيات أكثر استهدافًا. (هل تعلم أنه من الممكن عادةً تنزيل سجل البث الخاص بك؟)

البنية التحتية والأصول الأخرى

يمكن أن تستفيد العديد من أنواع البنية التحتية والأصول الأخرى من تحسينات الصيانة التي يمكن للذكاء الاصطناعي تقديمها. ويمكن للذكاء الاصطناعي تتبع الحاجة إلى الصيانة والتأكد من إجراء الصيانة الاستباقية (بدلاً من الصيانة التفاعلية) على النحو المطلوب. وهذا سيقطل أي وقت تعطل ناتج عن الإصلاحات ويمكن أن يطيل أيضًا عمر الأصول.

مخاطر وعيوب الذكاء الاصطناعي في مختلف القطاعات

يمكن أن يتسبب الذكاء الاصطناعي في عدد من المخاطر الشائعة وله عدد من العيوب التي تؤثر في جميع القطاعات.

المهارات

- المهارات المعرفية: العمليات والإستراتيجيات المعرفية:
- التحليل
 - الاستدلال المنطقي/الحجة

موضوعات ذات صلة

لمزيد من المعلومات حول استخدامات الذكاء الاصطناعي في سياقات مختلفة، انظر الجزء استخدامات الذكاء الاصطناعي في الصناعات والقطاعات المختلفة في هذه الوحدة.

المهارات

- المهارات المعرفية: العمليات والإستراتيجيات المعرفية:
- التفكير الناقد
 - حل المشكلات

المصطلحات الرئيسية

الجهات الفاعلة السيئة – الاسم الجماعي القياسي للأشخاص الذين يرغبون في تعطيل و/أو استغلال الذكاء الاصطناعي لصالحهم، على الرغم من أن المصطلح يستخدم بشكل عام لوصف من يقومون بأعمال ضارة بخصوص أنظمة الحاسوب.

التسريب – هو عملية الكشف عن البيانات.

التحيز والظلم

هناك خطر أن يتسبب الذكاء الاصطناعي في نتائج غير عادلة يمكن أن تؤدي إلى التمييز ضد مجموعات معينة من الأفراد. ومن الأمثلة على هذه المشكلة نظام الذكاء الاصطناعي الذي ينتج نتائج تشخيصية أسوأ للأقليات العرقية. ربما يكون السبب في ذلك هو استخدام البيانات المتحيزة لتدريب نظام الذكاء الاصطناعي.

مخاوف الخصوصية

تمثل مخاوف الخصوصية مشكلة عند استخدام بيانات ذات طبيعة شخصية أو حساسة للغاية لتدريب نموذج الذكاء الاصطناعي. قد تشمل المشكلات الناتجة إساءة استخدام المعلومات الشخصية والأحكام الضارة ضد الأفراد أو المراقبة غير المصرح بها.

الإزاحة/الخسائر الوظيفية

يحدث الذكاء الاصطناعي ثورة في طريقة إنجاز العمل. ونتيجة لذلك، يتمثل أحد عيوبه الرئيسية في أنه سيحل محل مجموعة كاملة من الوظائف التي تتصف بالتركرار أو الروتينية (على سبيل المثال، تحديد المنتجات المعيبة على خط التجميع). ومع ذلك، فمن المحتمل أيضاً أن يتم إنشاء أدوار جديدة من خلال حلول الذكاء الاصطناعي، إلا أن هذه الوظائف الجديدة قد تتطلب إعادة بناء مهارات الأشخاص وإعادة تدريبهم.

المخاطر الأمنية

تعترف معظم (وليس كل) المؤسسات التي تستخدم التكنولوجيا بإمكانات الذكاء الاصطناعي على خلق مخاطر أمنية، وهذا يرجع إلى أن أنظمة الذكاء الاصطناعي يمكن استهدافها من **الجهات الفاعلة السيئة** التي ترغب في تعطيل و/أو استغلال الذكاء الاصطناعي لصالحها، واحتمال حدوث انتهاكات للبيانات بما تسمح بتسريب البيانات الشخصية أو الحساسة هو أيضاً أحد العيوب المعروفة.

عدم قابلية التفسير

تفتقر العديد من أنظمة الذكاء الاصطناعي إلى الشفافية، ما يجعل الناس يفقدون الثقة في أنظمة الذكاء الاصطناعي نظراً إلى أنهم يتخذون قرارات لأسباب يصعب تفسيرها. يتسبب الذكاء الاصطناعي أيضاً في عدد من الاهتمامات الأكثر وضوحاً في قطاعات صناعية وتجارية محددة.

الشؤون المالية

تستخدم العديد من أنظمة التمويل الذكاء الاصطناعي لاتخاذ القرارات، مثل تقييم مدى ملائمة الشخص للقرض أو الرهن العقاري. هناك احتمال أن تكون هذه القرارات غير صحيحة (أو تعتبر غير صحيحة). يمكن أن يسبب هذا القلق والتوتر وعلاقات العملاء السيئة (على سبيل المثال، عندما يتم رفض القروض والرهون العقارية). سيلغي عدد قليل جداً من المؤسسات المالية هذه القرارات، على الرغم من تأثيرها في المقترضين.

الرعاية الصحية

استخدم مقدمو الرعاية الصحية الذكاء الاصطناعي بنجاح لدعم مجموعة من الخدمات، بما في ذلك الابتكارات مثل مراقبة المرضى عن بُعد. ومع ذلك، فإن احتمال التشخيص الطبي غير الصحيح ما يزال قائماً، ومثل هذه الأخطاء قد تؤدي إلى استخدام علاجات غير صحيحة أو قد تؤدي إلى وفيات. ومع ذلك، هناك أدلة على أن أنظمة التشخيص القائمة على الذكاء الاصطناعي قد حققت بعض النجاحات الحقيقية، ولا سيما تشخيص السرطانات التي لم يكتشفها المتخصصون. ومع ذلك، ما تزال هذه الأنظمة غير موثوقة تماماً وما يزال فحص المتخصصين لجميع القرارات هو النظام المعمول به.

التثقيف والتوعية

من الواضح أن التعليم استفاد من الحوسبة والذكاء الاصطناعي خلال جائحة COVID-19 العالمية الأخيرة، ومن الواضح أن التكنولوجيا جعلت التعليم أكثر سهولة. ومع ذلك، ما يزال هناك احتمال أن يؤدي الذكاء الاصطناعي إلى عدم المساواة التعليمية (على سبيل المثال، لمن ليس لديهم إمكانية الوصول إلى أجهزة الحاسب). كما أن الاعتماد المفرط على التكنولوجيا والتعلم عن بعد يمكن أن يكون له تأثير في تطوير مهارات التعامل مع الآخرين التي يتم تعلمها عادةً في الصفوف الدراسية وجهاً لوجه.

الوسائط

لقد خلق استخدام الذكاء الاصطناعي في وسائل الإعلام عددًا من المشكلات. وهي مسؤولة عن انتشار المعلومات المضللة (الأخبار المزيفة) ومقاطع الفيديو متقنة التزييف والادعاءات الكاذبة بتزوير الانتخابات. وقد أثار هذا مخاوف كبيرة بشأن الخصوصية وساعد على تقويض ثقة الجمهور في الشخصيات العامة والمسؤولين.

البنية التحتية

إن استخدام الذكاء الاصطناعي في إدارة المرافق الحيوية مثل الطاقة والمياه ومراقبة حركة المرور والمستشفيات جعلها عرضة للهجمات الإلكترونية والأنشطة الإجرامية الأخرى، وكانت هناك مؤخرًا بعض الأمثلة البارزة على ذلك في جميع أنحاء العالم.

اختيار حل الذكاء الاصطناعي الصحيح لتلبية الاحتياجات التشغيلية

إن اختيار حل الذكاء الاصطناعي الصحيح لتلبية الاحتياجات التشغيلية للمؤسسة ليس بالمهمة السهلة. ومع ذلك، يوجد بعض العمليات الرئيسية التي يجب اتباعها لضمان نتائج جيدة.

تحديد الأهداف

حدد أهدافًا واضحة للمشروع من خلال تحديد ما تحاول تحقيقه بالضبط. في إطار هذه العملية، يجب أن تسأل عما إذا كان الذكاء الاصطناعي مطلوبًا حقًا أم يمكن تكييف الأنظمة الحالية لحل المشكلة.

تقييم البيانات

من أجل دعم حلول الذكاء الاصطناعي والتأكد من تلبيتها للاحتياجات التشغيلية الرئيسية، يجب تقييم البيانات المتاحة. في إطار هذه العملية، يجب عليك التفكير في المتناقضات أو المتقابلات الخمسة – الحجم والتنوع والسرعة والصدق والقيمة، وتوفر هذه المعايير أساسًا قويًا لاختيار البيانات المناسبة.

دراسة السياق

الاعتبار المهم التالي هو دراسة سياق الوسط الحقيقي الذي سيوجد فيه الذكاء الاصطناعي وسيتم استخدامه. سيتيح لك ذلك تسليط الضوء على أي اعتبارات تنظيمية وأخلاقية تحتاج إلى النظر فيها.

مراجعة مهارات الفريق

من الضروري مراجعة القدرات الفنية للفريق وأعضاء الفريق الذين سيشاركون في تطوير حلول الذكاء الاصطناعي. على الرغم من إمكانية جلب أي مهارات غائبة (إما عن طريق تدريب أعضاء الفريق وإما توظيف متعديين)، إلا أن ذلك يخلق خطرين: زيادة كبيرة محتملة في التكاليف، وكذلك احتمال عدم تمكن المتعديين من إنجاز المشروع. بالإضافة إلى ذلك، في الحالات القصوى، يمكن للمقاولين زعزعة استقرار أنشطة الفريق، خاصة عندما يتعين زيادة السرعة.

مراجعة البنية التحتية

يجب أيضًا مراجعة التوافق التقني وقدرات البنية التحتية الحالية (على سبيل المثال، الشبكات وقدرات المعالجة والتخزين وما إلى ذلك). قد يلزم شراء تكنولوجيا جديدة أو ترقية التكنولوجيا الحالية.

مقياس التفكير

سيحتاج فريق تطوير الحلول إلى النظر في كيفية توسيع حل الذكاء الاصطناعي لتلبية الطلب المحتمل. يجب عليهم بعد ذلك النظر في ما قد ينطوي عليه هذا التوسع. يمكن أن يشمل ذلك توسيع نطاق الذكاء الاصطناعي من خلال توسيع استخدامه عبر المؤسسة، أو دمجها في الوظائف الأساسية. ومن ثم ستعود التحليلات الإضافية المقدمة بالفائدة على المؤسسة وستساعد على تحفيز إستراتيجية أعمالها. كجزء من هذه الخطوة، يجب أيضًا إجراء تحليل التكلفة والعائد لتحديد عائد الاستثمار.

المصطلح الرئيس

المتقابلات الخمسة – مجموعة من المعايير المعترف بها والتي تُقِيم البيانات على أساسها لإثبات ملائمتها لغرض معين.

المهارات

المهارات المعرفية: العمليات والإستراتيجيات المعرفية:

- التفكير الناقد
- اتخاذ القرار

موضوعات ذات صلة

لمزيد من المعلومات حول المتقابلات الخمسة، راجع الجزء الخاص بالمتقابلات الخمسة للبيانات الضخمة.

المصطلح الرئيس

إثبات المفهوم (PoC) – هو عملية تطوير أحد الحلول المحدودة الذي يمكن اختباره في بيئة اختبار للتحقق من إمكانية تطبيقه.

تقييم المفهوم

في العديد من حلول الذكاء الاصطناعي، سيُقيم المفهوم للتأكد من أنه ممكن ومجدي، وإذا كانت الإجابة عن نشاط إثبات المفهوم (PoC) هي "لا"، فربما يلزم تكثيف الحل، أو قد يتأجل المشروع بأكمله.

ضمان دعم أصحاب المصلحة

سيكون عدد قليل جداً من مشاريع تكنولوجيا المعلومات قابلاً للتطبيق إذا لم يُقدّم دعم من أصحاب المصلحة. هذا لأن الحل قد يحتاج إلى استثمار وفي بعض الحالات إلى مشاركة مباشرة من أصحاب المصلحة. غالباً ما يكون من الصعب البدء في المشاريع التي لا تحظى بدعم أصحاب المصلحة، ناهيك عن اكتمالها. تتمثل إحدى الطرق الرئيسة لضمان مشاركة أصحاب المصلحة ودعمهم في التأكد من إشراك أصحاب المصلحة في العملية، لا سيما عند صنع القرار، وإطلاعهم بانتظام على التقدم المحرز.

ضمان الأمان

يجب بناء حلول الذكاء الاصطناعي بإحكام، مع بذل كل جهد لضمان مستويات مناسبة من الأمان. يستكشف هذا الموضوع تخطيط حلول الذكاء الاصطناعي. وينظر في كيفية تحديد البيانات لدعم الحلول وجمعها وإعدادها.

A.P1, A.P2, A.M1, A.D1

تمرين تقييمي 21.1

أنت تعمل لدى مدير مشروع ذكاء اصطناعي وقد طلب منك إعداد بودكاست مدته 15 دقيقة حول استخدامات الذكاء الاصطناعي وتطبيقاته. يجب إنتاجه بأسلوب يجعله مثيراً للاهتمام مع سهولة وصول أصحاب المصلحة الرئيسيين إليه.

يجب أن يغطي البودكاست الخاص بك:

- كيفية استخدام المفاهيم الأساسية للذكاء الاصطناعي في المجال لتلبية الاحتياجات المحددة
- الفوائد والمخاطر والعيوب المرتبطة باستخدام الذكاء الاصطناعي في سياقات مختلفة
- تحليل فوائد ومخاطر وعيوب الذكاء الاصطناعي وكيفية تأثيرها في المجالات المختلفة.
- تقييم تأثير الذكاء الاصطناعي في المجالات المختلفة.

التخطيط

- ضع خطة للمهمة. ضع قائمة بكل ما ستحتاج إلى القيام به وضع جدول زمني لموعد إنجاز العمل.

التنفيذ

- صمم محتوى يعالج كل جزء من البودكاست.
- تذكر توقيت كل مقطع وتأكد من عدم تجاوز الحد الزمني الإجمالي.
- سجل البودكاست الخاص بك كملف صوتي.

المراجعة

- اطلب من أحد زملائك مراجعة المحتوى من خلال الاستماع إلى البودكاست الخاص بك.
- استناداً إلى مراجعة المستمع، حدد ما إذا كنت قد أدرجت النقاط الرئيسة بشكل مناسب وبطريقة شيقة وغنية بالمعلومات.

ب } تخطيط حل الذكاء الاصطناعي وإعداده لتلبية احتياجات محددة

يستكشف هذا الموضوع تخطيط أحد حلول الذكاء الاصطناعي. ويتناول كيفية التعرف على البيانات اللازمة لدعم الحل وكيفية جمعها وإعدادها.

ب1 بيانات الذكاء الاصطناعي

يتناول هذا الجزء المفاهيم الأساسية للبيانات واستخدامها في الذكاء الاصطناعي. يدرس المفاهيم الأساسية مثل الطريقة التي يتم بها تنظيم البيانات (أو عدم تنظيمها في الواقع)، وأنواع البيانات المختلفة. كما يبحث في مصادر البيانات المختلفة وكيفية استخدام البيانات. سيساعدك هذا على تحديد مصادر البيانات المناسبة لمشاريع الذكاء الاصطناعي ويوضح لك كيف يجب إعداد البيانات للاستخدام.

المهارات

مهارات التواصل الشخصي: أخلاقيات

العمل/الضمير:

- المبادرة
- المسؤولية
- الإنجابية

أنواع مختلفة من البيانات

قد توجد العديد من أنواع البيانات في أي سيناريو تريد التحقيق فيه لمشروع الذكاء الاصطناعي. تعد معرفة الأنواع المختلفة من البيانات وفهم خصائص كل فئة أمرًا مهمًا، نظرًا إلى تأثير نوع البيانات في طريقة التعامل معها والحفاظ عليها وفحصها.

الأنواع الأربعة الرئيسية من البيانات التي تحتاج إلى مراعاتها هي:

- منظمة
- شبه منظمة
- غير منظمة
- السلاسل الزمنية

البيانات المنظمة

يتم تنظيم البيانات المنظمة بشكل جيد، حيث يتم تخزين عناصرها عادةً بتنسيق جدولي (يشبه الجدول) (على سبيل المثال، باستخدام الصفوف والأعمدة). يحتوي كل عنصر بيانات على نوع بيانات محدد (على سبيل المثال، عدد صحيح، ورقم عشري، وسلسلة، وما إلى ذلك)، ما يسهل البحث والمعالجة بسهولة بالغة. عادةً ما يتم تخزين هذا النوع من البيانات في جدول بيانات وقاعدة بيانات وملف CSV وسجل وما إلى ذلك. تصّور قاعدة بيانات يستخدمها بائع تجزئة كبير لتخزين تفاصيل حول مشتريات العميل (إما عبر الإنترنت وإما في المتجر وإما كليهما). تحتوي قاعدة البيانات هذه على بيانات منظمة. إنها كبيرة جدًا ويتم تحديثها بشكل متكرر.

نظرًا إلى أن قاعدة البيانات تنظمًا جيدًا، يمكن الاستعلام عنها (يدويًا أو برمجياً). ومع ذلك، نظرًا لوجود العديد من نقاط البيانات التي يجب مراعاتها، فمن المستحيل تقريبًا إدارتها بفعالية. علاوة على ذلك، من الصعب الحصول على أي معلومات تنبؤية أو رؤى منها ما يتعلق بتاريخ أنماط الشراء من جانب العملاء. هذا هو المكان الذي يجب فيه استخدام التعلم الآلي والذكاء الاصطناعي. لأنه يوفر الأدوات الخوارزمية والإحصائية وموارد المعالجة المرتبطة بها (القابلة للتطوير إذا كانت في السحابة) للتعامل مع الحجم الهائل لقاعدة البيانات المنظمة. ويمكن من العثور على الارتباطات الظاهرة والارتباطات المخفية داخل البيانات بسهولة وسرعة.

البيانات غير المنظمة

البيانات غير المنظمة ليست مهيكلة. من أمثلة هذا النوع من البيانات النص الخام والصور ومقاطع الفيديو والصوت والرسائل الفورية ومحتوى صفحة الويب. تعد منشورات وسائل التواصل الاجتماعي مثالًا جيدًا للبيانات غير المنظمة، حيث تحتوي عادةً على مجموعة من بيانات النص والصورة والفيديو. تشمل الأمثلة الأخرى للبيانات غير المنظمة التي تجمعها المنظمة مكالمات ملاحظات العملاء، أو فيديوهات كاميرات المراقبة (CCTV)، أو منشورات منتديات الدعم - كل هذه البيانات متغيرة في طبيعتها.

بحكم طبيعتها، يصعب للغاية معالجة البيانات غير المنظمة يدويًا. إن افتقارها إلى هيكل محدد يجعل من الصعب البحث فيها دون استخدام أدوات متخصصة. بالإضافة إلى ذلك، في كثير من الحالات، يزداد الحجم الهائل لهذا النوع من البيانات بشكل كبير بمرور الوقت ويمكن أن يكون التخزين مكلفًا للغاية.

الأدوات القائمة على الذكاء الاصطناعي قادرة على معالجة هذا النوع من البيانات بدرجة جيدة من الدقة،

المصطلح الرئيس

البيانات الوصفية – البيانات التي تقدم وصفاً للبيانات.

اعتمادًا على طريقة تدريبها. على سبيل المثال، يمكن لأدوات الذكاء الاصطناعي:

- تمييز الأشكال في الصور/الفيديو
- تحليل المشاعر على وسائل التواصل الاجتماعي ومشاركات المنتدى
- تنفيذ تقنيات معالجة اللغة الطبيعية (NLP).

عندما تعمل أنظمة الذكاء الاصطناعي بشكل جيد مع هذا النوع من البيانات، تكون النتائج مشجعة - يتم القبض على المجرمين عند نقاط مراقبة الحدود، ويتم تعرف السيارات المسرعة بنجاح من خلال لوحات التسجيل الخاصة بها ويتم تعرف الأنسجة الخبيثة بشكل صحيح من خلال عمليات المسح. ومع ذلك، يمكن أن تختلف الدقة. هذا يمكن أن يؤدي إلى مشكلات. على سبيل المثال، في المملكة المتحدة، أصدرت بعض منظمات إنفاذ القانون إشعارات البحث عن أفراد أبرياء بسبب أنظمة غير دقيقة لتمييز الوجوه بالذكاء الاصطناعي.

البيانات شبه المنظمة

البيانات شبه المنظمة هي مزيج من البيانات المنظمة وغير المنظمة، وغالبًا ما تنشأ لتحقيق التوازن بين الحاجة إلى سهولة القراءة البشرية والحاجة إلى المعالجة الآلية، إذ إن توافر بعض عناصر البيانات المنظمة يسهل تحليلها، وغالبًا ما تتضمن البيانات شبه المنظمة استخدام **البيانات الوصفية**.

يمكن اعتبار البريد الإلكتروني شبه منظم لأن التنسيق (العنوان والموضوع والنص والمرفقات وما إلى ذلك) موثوق جيدًا ومنفصل. ومع ذلك، فإن محتوى الجسم نفسه غالبًا ما يكون نصًا غير منظم. يمكن أيضًا رؤية البيانات شبه المنظمة في أمثلة JSON وXML، ما يوفر تبادلًا فعالًا للبيانات بين الخوادم والعملاء باستخدام الطلبات والاستجابات المستندة إلى واجهة برمجة التطبيقات (API).

بحث

توجد البيانات الوصفية في العديد من السياقات بما في ذلك رسائل البريد الإلكتروني وملفات الصوت والمواقع الإلكترونية. افحص بعض ملفاتك الخاصة وحدد البيانات الوصفية.

هل تحتوي البيانات الوصفية (من النوع الذي تسجله الكاميرا الرقمية) على معلومات عن الصورة؟

وقفة للتفكير



ما نوع المعلومات التي تخزنها بيانات الصورة الوصفية؟

تلميح

لماذا تكون التفاصيل حول الكاميرا المستخدمة، وإعدادات الكاميرا، والمعلومات حول مكان وتاريخ التقاط الصورة مفيدة؟

توسيع الأفق

بيانات السلاسل الزمنية

بيانات السلاسل الزمنية هي ببساطة البيانات التي يتم جمعها في نقاط زمنية محددة (على سبيل المثال، كل يوم) أو تسجيلها على فترات زمنية متتالية (على سبيل المثال، كل 30 ثانية). يرتبط بشكل أساسي بالمراقبة والتتبع - على سبيل المثال، تسجيل وتتبع درجات الحرارة، والإقبال على المتجر، والضغط، والرطوبة، والأمطار، وزوار الموقع.

من ناحية البيانات، تستفيد بيانات السلاسل الزمنية من الطابع الزمني (الذي يحتوي على بيانات التاريخ والوقت) ما يجعل كل بند من بنود المدخلات فريدًا. يمكن أن يكون معدل حقن هذا النوع من البيانات في بعض الأنظمة سريعًا جدًا (يحدث هذا الأمر عند التقاط البيانات بشكل متكرر). لهذا السبب، يجب استخدام قواعد بيانات فعالة يمكنها التعامل مع تكرار مدخلات البيانات.

هل يمكنك التفكير في ثلاثة أمثلة على الأقل لبيانات السلاسل الزمنية؟

وقفة للتفكير



فكر في الصناعات والمواقف المختلفة التي قد تولد بيانات السلاسل الزمنية.

تلميح

فكر في ما إذا كانت هذه البيانات تُجمع لفترات قصيرة أو طويلة.

توسيع الأفق

لماذا قد تستخدم المؤسسات هذه الأنواع المختلفة من البيانات؟

قد تكون البيانات متاحة بتنسيق واحد فقط وبالتالي سيتعين على المؤسسة استخدامها في هذا النموذج. على سبيل المثال، قد تتلقى شركة إعلامية بيانات غير منظمة فقط من منشورات وسائل التواصل الاجتماعي. في كثير من الأحيان، يعكس نوع بيانات المصدر احتياجات المعالجة للمؤسسة. على سبيل المثال، إذا كانت المؤسسة بحاجة إلى الوصول السريع إلى السجلات المالية أو مخزون المعدات أو سجلات التوظيف، فستطلب تسجيل ذلك وتسليمه كبيانات منظمة. كما هو موضح، يتم البحث عن هذا النوع من البيانات بسرعة ويمكن إحضاره وإصداره بسهولة إلى تطبيقات أخرى.

عندما يتم تحليلها باستخدام الذكاء الاصطناعي، يمكن أن تكون البيانات غير المنظمة أكثر فائدة للمؤسسة من البيانات المنظمة. هذا لأن حلول الذكاء الاصطناعي المتقدمة جيدة جدًا في تحليل البيانات غير المنظمة وإيجاد الاتجاهات والارتباطات والأنماط. ومن الأمثلة البسيطة على ذلك تخزين وتحليل الفحوصات الطبية للمرضى (وهي بيانات غير منظمة) لتحسين التشخيص السريري. في هذه الحالة، تكون البيانات غير المنظمة أكثر فائدة لنظام الذكاء الاصطناعي من الوصف المنظم لأخصائي الأشعة الذي تم إدخاله في قاعدة بيانات المرضى.

قامت العديد من المؤسسات بتحليل التطبيقات الكبيرة (المتجانية) (التي كانت توفر فيما سبق جميع احتياجات البيانات الخاصة بالمؤسسة) إلى عدد من الخدمات الصغيرة المختلفة التي تتواصل مع بعضها البعض باستخدام بيانات غير منظمة. ونتيجة لذلك، سيكون لدى المؤسسات كمية هائلة من البيانات التي يتم تخزينها (واستهلاكها) بهذا التنسيق.

تعد بيانات السلاسل الزمنية التي يتم الحصول عليها كمجموعة متكررة من القياسات المسجلة بمرور الوقت أمرًا بالغ الأهمية لتحليل الاتجاهات. هذا يجعلها لازمة للأشخاص والمنظمات العاملين في مجال التنبؤات. يشمل هؤلاء الأشخاص خبراء الأرصاد الجوية (التنبؤ بالطقس) أو سماسرة البورصة (التنبؤ بارتفاع أو انخفاض الأسهم) أو مهندسي البنية التحتية الذين يراقبون تدفق حركة المرور في المدينة.

المصادر الشائعة للبيانات

يتم إنشاء كميات هائلة من البيانات يوميًا بناءً على تفاعلات الأشخاص مع التكنولوجيا، ونتيجة لمعالجة مجموعة واسعة من المعاملات. يمكن تخزين هذه البيانات واستخراجها للتحليل للإجابة عن أسئلة الأعمال وبناء حلول الذكاء الاصطناعي.

الويب

يمكن استخراج البيانات من المواقع الإلكترونية المتاحة للجمهور أو صفحات وسائل التواصل الاجتماعي أو لوحات الرسائل باستخدام تقنية تسمى (مسح الويب). هذه تقنية شائعة تستخدم في مجالات مثل أبحاث السوق ومراقبة الأسعار (والمقارنة) وتوليد العملاء المحتملين (تحديد العملاء المحتملين وجذبهم) وجمع البيانات لنماذج التعلم الآلي.

المهارات

- المهارات المعرفية: العمليات والإستراتيجيات المعرفية:
- حل المشكلات
- المهارات المعرفية: الإبداع:
- الإبداع

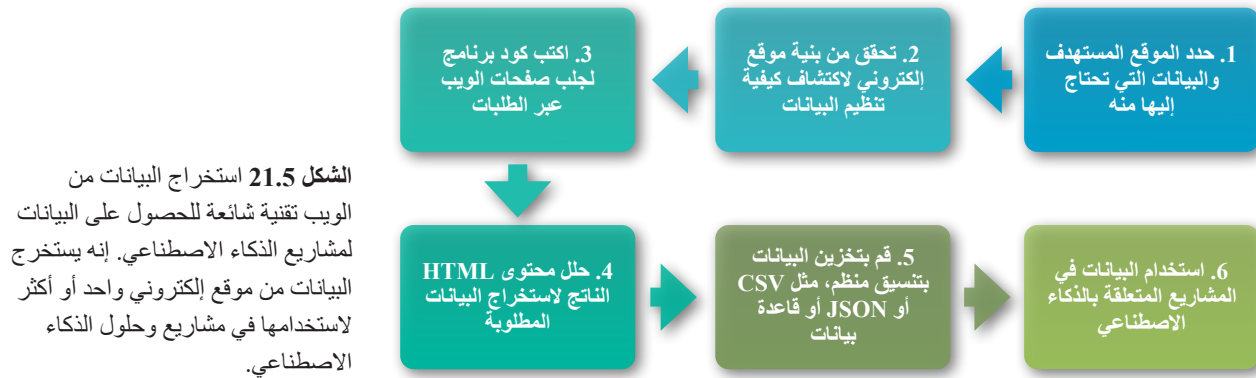
وقفة للتفكير فكر في استخدامك الخاص للإنترنت.

تلميح ما مقدار البيانات التي تعتقد أنك تولدها كل يوم؟

توسيع الأفق كم من هذه البيانات شخصية أو يمكن تعرفها بشكل شخصي؟

غالبًا ما يستخدم مصطلح (تحليل "parse") لوصف عملية استخراج البيانات المنظمة (على سبيل المثال، الأسماء والأسعار وما إلى ذلك) من البيانات غير المنظمة أو شبه المنظمة (على سبيل المثال، لغة ترميز النص التشعبي لصفحة الويب (HTML)).

الخطوات الأساسية لاستخراج البيانات من الويب موضحة في الشكل 21.5.



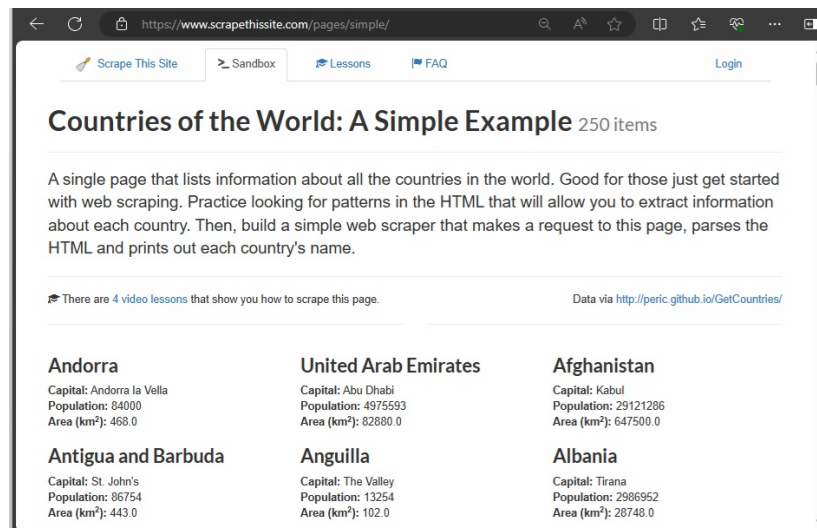
الشكل 21.5 استخراج البيانات من الويب تقنية شائعة للحصول على البيانات لمشاريع الذكاء الاصطناعي. إنه يستخرج البيانات من موقع إلكتروني واحد أو أكثر لاستخدامها في مشاريع وحلول الذكاء الاصطناعي.

عند التفكير في استخراج البيانات من الويب، من المهم أن تكون حذرًا. يمكن اعتبار هذا النشاط انتهاكًا لشروط خدمة موقع إلكتروني. يجب عليك دائمًا التحقق من أن استخراج البيانات مسموح به. تمثل المنتديات ولوحات الرسائل مصدرًا غنيًا جدًا للبيانات نظرًا إلى تحديثها باستمرار بمحتوى مستخدم جديد. عادةً ما يتضمن نوع البيانات التي يمكن استخراجها منها:

- المحتوى الذي ينشئه المستخدم (عناوين سلاسل الرسائل والمشاركات والردود وما إلى ذلك)
- ملفات تعريف المستخدمين
- البيانات الهيكلية المتعلقة بلوحات/المنتديات (على سبيل المثال، الفئات، وتنظيم الموضوعات، وما إلى ذلك)
- البيانات المستندة إلى الوقت، بما في ذلك تواريخ وأوقات الإنشاء، وتكرار المشاركات، وما إلى ذلك.
- البيانات الوصفية، بما في ذلك علامات الموضوع، والتصويت لأعلى ولأسفل، والملاحظات، وما إلى ذلك.
- الروابط المضمنة أو المراجع المستخدمة في المنشور.

مثال عملي

يوضح المثال العملي الآتي كود بايثون نموذجي يقوم بعملية استخراج بيانات بسيطة من موقع مُصمم خصيصًا للسماح للمتعلمين بممارسة هذه العملية. استخدم محرك بحث للبحث عن 'Scrape this site tool'. يعرض الشكل 21.6 صفحة الويب المحددة من خلال عميل استعراض ويب.



الشكل 21.6 مثال عملي لصفحة ويب تسرد معلومات أساسية عن دول العالم كأساس لإدراجها في نشاط استخراج البيانات من الويب.

الشكل 21.7 يظهر HTML الأساسي. هذا مجرد مقتطف لبلد واحد 'جنوب أفريقيا'.

```
<div class="col-md-4 country"> <h3 class="country-name"> <i class="flag-icon
flag-icon-za"></i> South Africa </h3> <div class="country-info">
<strong>Capital:</strong> <span class="country-capital">Pretoria</span><br/> <strong>Population:</
strong> <span class="country-
population">49000000</span><br/> <strong>Area (km<sup>2</sup>):</strong> <span class="coun-
try-area">1219912.0</span><br/> </div> </div>
```

الشكل 21.7 مقتطف HTML يوضح العلامات والبيانات المرتبطة للبلد الواحد المحدد.

يوضح الشكل 21.8 مثالاً على كود بايثون يجمع البيانات من الموقع، حيث يستخرج البيانات الرئيسة ويخزنها.

```
import requests
from bs4 import BeautifulSoup
import json

# URL of the website you want to scrape
url = 'https://www.scrapethissite.com/pages/simple/'

# Send a GET request to the website
html_content = requests.get(url)

# Parse the HTML content of the page with BeautifulSoup
soup = BeautifulSoup(html_content.text, 'html.parser')

# Create an empty dictionary to store our scraped data...
countries_data = {}

# Loop over each Country...
for country_div in soup.find_all("div", class_="col-md-4 country"):
    country_name = country_div.find("h3", class_="country-
name").get_text(strip=True)
    capital = country_div.find("span", class_="country-
capital").get_text(strip=True)
    population = country_div.find("span", class_="country-
population").get_text(strip=True)
    area = country_div.find("span", class_="country-
area").get_text(strip=True)

    # Append discovered country into the dictionary
    countries_data[country_name] = {
        "Capital": capital,
        "Population": population,
        "Area": area
    }

# Convert the dictionary to JSON and write it to a new file
with open(r'countries_data.json', 'wt') as json_file:
    json.dump(countries_data, json_file, indent=4)
```

الشكل 21.8 كود بايثون يجمع البيانات من موقع إلكتروني لاستخراج البيانات الرئيسة وتخزينها.

متابعة

يوضح الشكل 21.9 بعض البيانات المستخرجة بتنسيق JSON من الملف الذي تم إنشاؤه.

```

"Mayotte": {
  "Capital": "Mamoudzou",
  "Population": "159042",
  "Area": "374.0"
},
"South Africa": {
  "Capital": "Pretoria",
  "Population": "49000000",
  "Area": "1219912.0"
},
"Zambia": {
  "Capital": "Lusaka",
  "Population": "13460305",
  "Area": "752614.0"
},

```

الشكل 21.9 بيانات بتنسيق JSON.

كما ترى، أصبحت هذه البيانات الآن نظيفة وخالية من أي محتوى متعلق بـ HTML، ما يجعلها مناسبة للإدخال في أي عملية تعلم آلي.

المصطلح الرئيس

الطلبات البرمجية – طلب محتوى يعتمد على الويب عن طريق برنامج أو نص برمجي بدلاً من استخدام متصفح الويب يدويًا.

تجدر الإشارة هنا إلى أن بعض المواقع الإلكترونية لا تسمح بالطلبات البرمجية، وذلك لأن جزءًا من هذا الطلب يتضمن "سلسلة وكيل المستخدم". تحدد هذه القيمة ل خادم الويب الذي يقدم الخدمة اسم أداة البرنامج التي تقدم الطلب، فعلى سبيل المثال، يُعرّف المتصفح على أنه، Google Chrome®، Apple Safari®، Microsoft Edge®، وما إلى ذلك. عادةً ما تكون هذه الأمور مقبولة؛ إلا أنه عندما يقوم برنامج بإجراء طلب، فإنه عادةً لا يخفي طبيعته (مثل "لغة بايثون")، ما قد يؤدي إلى رفض خادم الويب للطلب.

من الممكن انتحال أو تزيف سلسلة وكيل المستخدم للتغلب على تردد خادم الويب، ما يجعل برنامجك يقلد عميل متصفح ويب أصلي.

قد توفر المواقع الإلكترونية الأخرى واجهات برمجة التطبيقات العامة (API) التي تسمح بالاستعلام عن مصادر البيانات الأساسية لموقع إلكتروني واستخراجها. اعتمادًا على طبيعة البيانات، قد تكون هذه الخدمة مجانية أو تتطلب استخدام مفتاح واجهة برمجة التطبيقات (API) قابل للشراء يمنح الوصول.

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/42.0.2311.135 Safari/537.36 Edge/12.246

الشكل 21.10 سلسلة وكيل المستخدم لجهاز حاسوب يعمل بنظام Windows 10 باستخدام متصفح Microsoft Edge.



الشكل 21.11 Postcodes.io.

يتمثل أحد الأمثلة ذلك في Postcodes.io (كما هو موضح في الشكل 21.11)، وهو مشروع نشأ باستخدام برنامج مفتوح المصدر بواسطة Ideal Postcodes. يقدم خدمة مجانية للمطورين للبحث عن البيانات المتعلقة بالرموز البريدية في المملكة المتحدة وتفسير الإحداثيات واستخراج البيانات. تتوفر مجموعات بيانات مماثلة لمناطق مختلفة، مثل postcode.eu لبلدان الاتحاد الأوروبي و getpostalcodes.com لمجموعات أخرى من البلدان. على سبيل المثال، يحتوي getpostalcodes.com على بيانات عن 12 منطقة بريدية في الأردن بما في ذلك عمان (العاصمة)، بالإضافة إلى مناطق مثل الزرقاء والطفيلة وإربد وجرش.

تعتبر البرامج مفتوحة المصدر، مثل تلك البرامج المستخدمة في Postcodes.io، ناجحة جدًا. مفتاح نجاحها هو أنها تسمح بالجهود التعاونية (من مجتمع المبرمجين) لتحسين المشروع بمرور الوقت، وإصلاح الأخطاء وإضافة وظائف وميزات جديدة عند الطلب.

يوفر Postcodes.io عددًا من نقاط النهاية المختلفة التي يمكن استخدامها للتحقق من بياناته. على سبيل المثال، يمكن استخدامه للتحقق مما إذا كان الرمز البريدي موجودًا بالفعل ومكان وجوده في العالم.

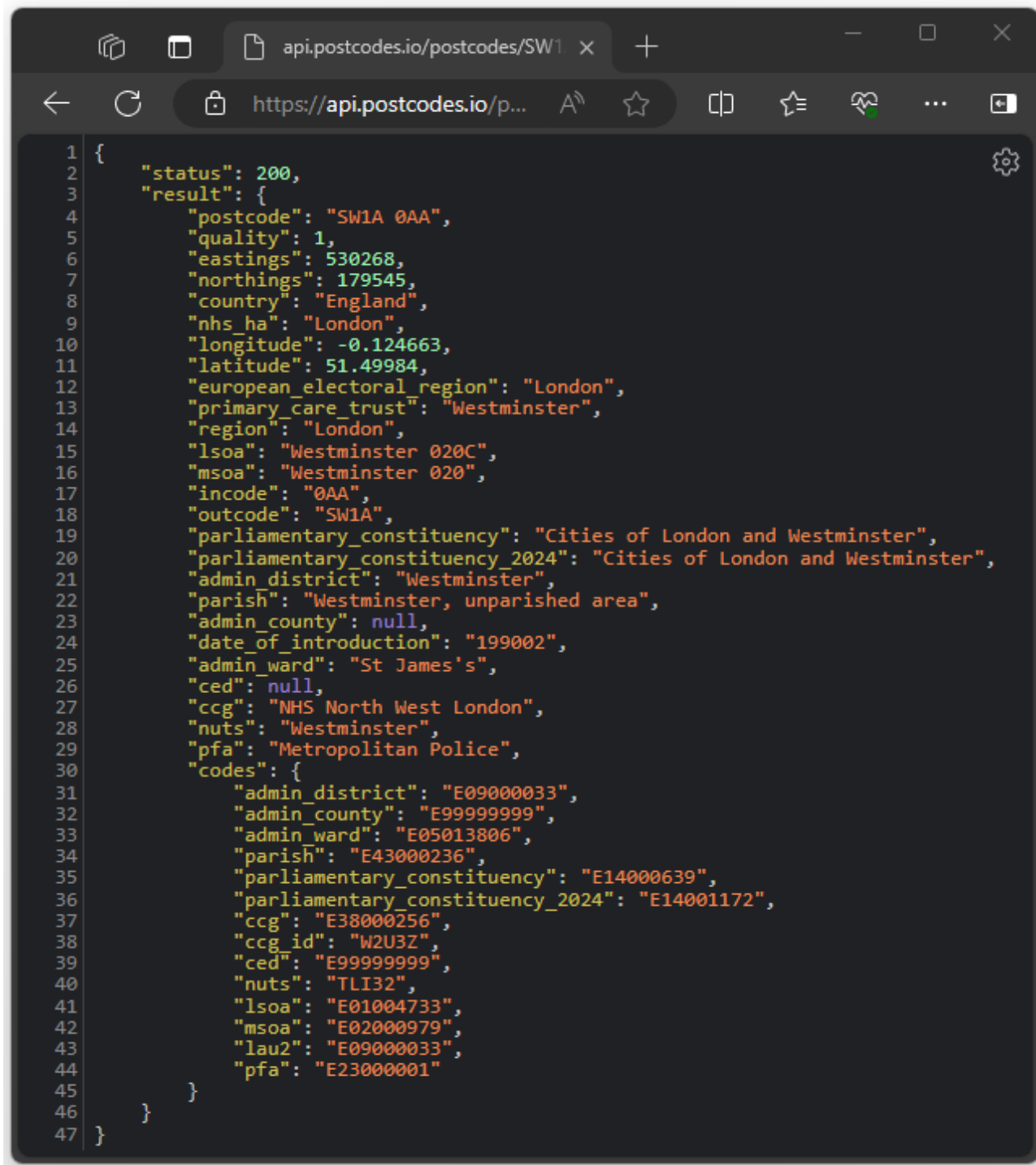


الشكل 21.12 التحقق من استخدام الرمز البريدي عبر نقطة النهاية الصحيحة لواجهة برمجة التطبيقات.

من الممكن أيضًا الوصول إلى هذه البيانات باستخدام طلب HTTP GET إما من خلال متصفح الويب (كما هو موضح في الشكل 21.13)، وإما باستخدام أداة سحب بيانات ويب مخصصة بلغة بايثون (كما هو موضح في الشكل 21.14).

المصطلح الرئيس

المصدر المفتوح – مشاريع البرمجيات التي تُتاح مجانًا للجمهور، وعادةً ما تسمح لأي شخص باستخدام الشفرة المصدرية للمشروع وتغييرها وتوزيعها.



```

1 {
2   "status": 200,
3   "result": {
4     "postcode": "SW1A 0AA",
5     "quality": 1,
6     "eastings": 530268,
7     "northings": 179545,
8     "country": "England",
9     "nhs_ha": "London",
10    "longitude": -0.124663,
11    "latitude": 51.49984,
12    "european_electoral_region": "London",
13    "primary_care_trust": "Westminster",
14    "region": "London",
15    "lsoa": "Westminster 020C",
16    "msoa": "Westminster 020",
17    "incode": "0AA",
18    "outcode": "SW1A",
19    "parliamentary_constituency": "Cities of London and Westminster",
20    "parliamentary_constituency_2024": "Cities of London and Westminster",
21    "admin_district": "Westminster",
22    "parish": "Westminster, unparished area",
23    "admin_county": null,
24    "date_of_introduction": "199002",
25    "admin_ward": "St James's",
26    "ced": null,
27    "ccg": "NHS North West London",
28    "nuts": "Westminster",
29    "pfa": "Metropolitan Police",
30    "codes": {
31      "admin_district": "E09000033",
32      "admin_county": "E99999999",
33      "admin_ward": "E05013806",
34      "parish": "E43000236",
35      "parliamentary_constituency": "E14000639",
36      "parliamentary_constituency_2024": "E14001172",
37      "ccg": "E38000256",
38      "ccg_id": "W2U3Z",
39      "ced": "E99999999",
40      "nuts": "TLI32",
41      "lsoa": "E01004733",
42      "msoa": "E02000979",
43      "lau2": "E09000033",
44      "pfa": "E23000001"
45    }
46  }
47 }

```

الشكل 21.13 استجابة JSON لطلب GET على نقطة النهاية الخاصة بواجهة برمجة التطبيقات.

```
import requests

# The URL of the API endpoint for the Postcodes.io service
url = "https://api.postcodes.io/postcodes/SW1A%20AA"

# Send a GET request to the API endpoint
response = requests.get(url)

# Specify a User-Agent string for Google Chrome
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
    AppleWebKit/537.36 (KHTML, like Gecko) Chrome/103.0.0.0 Safari/537.36'
}

# Send a GET request to the API endpoint with the custom headers
response = requests.get(url, headers=headers)

# Check if the request was successful (status 200=OK)
if response.status_code == 200:
    # Parse the response JSON and then print it
    data = response.json()
    print(data)
else:
    # Display error message
    print(f'Failed to retrieve data: HTTP Status Code {response.status_code}')
```

لاحظ استخدام سلسلة وكيل المستخدم المخادع لتوضيح كيفية تنفيذ ذلك. لحسن الحظ، لا تواجه Postcodes.io مشكلة في الرد على الاستعلامات البرمجية، لذلك هذا الأمر ليس ضروريًا جدًا في هذه الحالة. الناتج من كود بايثون موضح في الشكل 21.15.

الشكل 21.14. "كود بايثون لطلب واجهة برمجة تطبيقات بسيط لاستخراج بيانات للرمز البريدي SW1A 0AA باستخدام سلسلة وكيل مستخدم مزيفة."

```
{'status': 200, 'result': {'postcode': 'SW1A 0AA', 'quality': 1, 'eastings':
530268, 'northings': 179545, 'country': 'England', 'nhs_ha': 'London',
'longitude': -0.124663, 'latitude': 51.49984, 'european_electoral_region':
'London', 'primary_care_trust': 'Westminster', 'region': 'London', 'lsoa':
'Westminster 020C', 'msoa': 'Westminster 020', 'incode': '0AA', 'outcode': 'SW1A',
'parliamentary_constituency': 'Cities of London and Westminster',
'parliamentary_constituency_2024': 'Cities of London and Westminster',
'admin_district': 'Westminster', 'parish': 'WestminsterLondon', 'nuts': 'Westminster',
'pfa': 'Metropolitan Police', 'codes': {'admin_district', 'unparished area',
'admin_county': None, 'date_of_introduction': '199002', 'admin_ward': "St
James's", 'ced': None, 'ccg': 'NHS North West ': 'E09000033',
'admin_county': 'E99999999', 'admin_ward': 'E05013806', 'parish': 'E43000236',
'parliamentary_constituency': 'E14000639', 'parliamentary_constituency_2024':
'E14001172', 'ccg': 'E38000256', 'ccg_id': 'W2U3Z', 'ced': 'E99999999', 'nuts':
'TL132', 'lsoa': 'E01004733', 'msoa': 'E02000979', 'lau2': 'E09000033', 'pfa':
'E23000001'}}}
```

الشكل 21.15 ناتج كود بايثون.

المصطلح الرئيس

رموز حالة HTTP – رموز عالمية مكونة من ثلاثة أرقام يصدرها الخادم عندما يستجيب للطلب المقدم من العميل، ورمز الحالة الذي يبدأ بالرقم 2 يشير إلى النجاح.

يجب أن تكون البيانات المستخرجة برمجياً بهذه الطريقة مطابقة للطلب اليدوي للمتصفح، ويجب أن تدرك أنه يشيع استخدام رموز حالة HTTP للإشارة إلى الطلبات السيئة/الرموز البريدية غير الموجودة، ويوضح الشكل 21.16 هذه الإجراءات.



الشكل 21.16 رمز الحالة وتفاصيل الخطأ 404 اللذان يشاران إلى خطأ العميل.

وسائل التواصل الاجتماعي

تحتل وسائل التواصل الاجتماعي بشعبية استثنائية لدى الحكومات والشركات والأفراد (من جميع الأعمار). أدى اعتمادها في نطاق واسع إلى زيادة مذهلة في البيانات المولدة.

لا تقتصر بيانات وسائل التواصل الاجتماعي على البيانات الموجودة في الرسائل أو الصور التي تم تحميلها، بل تتضمن أيضاً جميع البيانات الوصفية المرتبطة بها (التاريخ والوقت والموقع الجغرافي وإبداء الإعجاب وعدم الإعجاب والردود وإعادة التوجيه وما إلى ذلك)

يمكن لبعض المؤسسات استخدام هذه البيانات بعدة طرق، بما يتضمن:

- في خوارزميات التوصية لترويج المحتوى للمستخدمين الآخرين
- لإنتاج إعلانات مستهدفة
- ليتم بيعها لأطراف ثالثة (اعتماداً على شروط وأحكام الاستخدام).

للحصول على فكرة عن حجم وسرعة البيانات التي تم إنشاؤها، افحص الشكل 21.17. هذا يُظهر ما يمكن لمستخدمي خدمة مثل منصة X® (المعروفة سابقاً باسم تويتر) إنتاجه خلال مدة زمنية محددة. +200 مليار تغريدة هي كمية هائلة من البيانات الجديدة.

يوضح هذا المستوى من إنشاء البيانات مدى انتشار منصة X العالمي والمشاركة المستمرة لمستخدميها من خلال تطبيقات الهواتف الذكية. نظراً لأن منصة X ليست سوى إحدى منصات التواصل الاجتماعي من بين العديد من المنصات الأخرى، يتضح أن وسائل التواصل الاجتماعي تخلق تدفقاً مذهلاً من المحتوى الجديد.



الشكل 21.17 إحصائيات توليد البيانات لموقع X (المعروف سابقاً باسم تويتر).

هل تستخدم وسائل التواصل الاجتماعي؟ هل فكرت يوماً في مدى ظهور منشوراتك؟

وقفة للتفكير



افحص إعدادات وسائل التواصل الاجتماعي الخاصة بك وقم بتعديلها إذا كنت تعتقد أنك تشارك الكثير من المعلومات مع أشخاص لا تعرفهم.

تلميح

هل ظهر أي من منشوراتك بشكل غير متوقع على الإنترنت؟

توسيع الأفق

نقطة البيع

يتم جمع كمية هائلة من البيانات من أنظمة نقاط البيع (PoS) في متاجر البيع بالتجزئة. يتضمن ذلك المعلومات الأساسية اللازمة لمعالجة مشتريات كل عميل بشكل صحيح. كما تعطي البيانات المجمعة أيضاً معلومات عن أنظمة مراقبة المخزون التي يستخدمها تجار التجزئة. هذا يسمح لأنظمة التحكم بالحفاظ على مستويات المخزون ثابتة عن طريق طلب المخزون ليحل محل البنود التي تم بيعها للعملاء.

◀ مثال على معاملة نقاط البيع مع تقديم بطاقة الدفع عبر قارئ البطاقات.



لمعرفة كيفية تصنيف البيانات من نظام نقاط البيع، انظر إلى الجدول 21.4.

الجدول 21.4 البيانات المأخوذة من المعاملة، مقسمة إلى أربع فئات رئيسية

بيانات حامل البطاقة	بيانات المعاملة
<ul style="list-style-type: none"> رقم البطاقة اسم حامل البطاقة تاريخ انتهاء الصلاحية 	<ul style="list-style-type: none"> معرف فريد تم إنشاؤه بواسطة نظام نقاط البيع التاريخ الوقت المبلغ العملة اسم بائع التجزئة رمز فئة بائع التجزئة (على سبيل المثال، نوع النشاط التجاري)
بيانات الأمان	التفويض
<ul style="list-style-type: none"> قيمة التحقق من البطاقة (CVV/CVC) بيانات شريحة EMV (يوروبي و ماستر كارد و فيزا) بيانات الشريط المغناطيسي 	<ul style="list-style-type: none"> رمز التفويض رمز الاستجابة (على سبيل المثال، موافق عليه، مرفوض)

لا يمكن الاحتفاظ بجميع هذه البيانات بعد معاملة بائع التجزئة. تخضع بعض البيانات الحساسة (رقم البطاقة الكامل، رمز التحقق من البطاقة، وما إلى ذلك) لمعايير أمان صارمة (مثل معيار أمان بيانات صناعة بطاقات الدفع (PCI DSS))، والتي تهدف إلى تقليل الاحتيال والوصول غير المصرح به. بالإضافة إلى البيانات التي يتم جمعها من خلال معاملات نقاط البيع، يقوم تجار التجزئة (مثل محلات السوبر ماركت) أيضاً بجمع البيانات من خلال استخدام بطاقات الولاء التي تُصدر للعملاء المنتظمين. تم تصميم هذه البطاقات لتزويد العميل بتجربة تسوق محسنة، بما في ذلك:

- خصومات نقدية
- الوصول إلى العروض الخاصة داخل المتجر
- عروض منتجات محسنة
- عروض تسويقية مصممة خصيصاً (على سبيل المثال، قسائم الخصم التي تتناسب مع عادات الشراء لدى العميل).

يقدم العملاء معلومات شخصية عند إنشاء حسابات بطاقات الولاء الخاصة بهم، بما في ذلك العنوان والعمر ومعلومات الأسرة. يتيح تحديد هوية العميل ببطاقة الولاء إمكانية تتبع سلوكيات وعادات التسوق الخاصة بالعميل على المدى القصير والطويل. لمعرفة كيفية تصنيف البيانات من بطاقة الولاء، انظر إلى الجدول 21.5.

منحت أنظمة نقاط البيع و بطاقات الولاء تجار التجزئة فرصة الوصول إلى بيانات أكثر كثرة مما كان متاحاً لهم في الماضي. يمكن تحليل هذه البيانات لفهم عادات شراء العملاء وتفضيلاتهم. ويتم تحليل هذا النوع من البيانات من خلال استخدام خوارزميات التعلم الآلي.

يمكن استخدام البيانات للتنبؤ بمبيعات المنتجات في أوقات مختلفة وفي أماكن مختلفة. وهذا هو السبب في أن بعض المنتجات المتوافرة في أحد فروع السوبر ماركت قد لا تكون بالضرورة متاحة في فرع آخر – لأنه قد تبين أن عددًا قليلاً جداً من العملاء يشترون السلعة في ذلك المكان.

يسمح استخدام معلومات بطاقة الولاء لبائع التجزئة بإنشاء تجارب تسوق مخصصة والتأكد من أن قسائم الخصم تتوافق بشكل وثيق مع احتياجات العملاء. بالإضافة إلى ذلك، يمكن لبائعي التجزئة استخدام هذه البيانات للترويج للمنتجات الجديدة باستخدام الإعلانات المستهدفة و قسائم الخصم.

الهدف من هذا العمل هو تحقيق الأرباح وجذب عملاء جدد والحفاظ على ولاء العملاء الحاليين.

بالطبع، يتم تنظيم جمع البيانات من خلال لوائح حماية البيانات في البلد المضيف. على سبيل المثال، في الاتحاد الأوروبي يتم تغطيتها بموجب اللائحة العامة لحماية البيانات. كما تمت مناقشته، لدى العديد من البلدان قواعد مماثلة تتطلب من تجار التجزئة جمع البيانات الشخصية وتخزينها ومعالجتها بطرق تعطي الأولوية للخصوصية والأمان.

الجدول 21.5 البيانات النموذجية وتفضيلات العملاء المخزنة في فئات مختلفة لحسابات بطاقات الولاء

بيانات المعاملة	تفاصيل العميل الشخصية
<ul style="list-style-type: none"> تاريخ الشراء تكرار التسوق حجم السلة عدد العناصر التكلفة الإجمالية طريقة الدفع بطاقة (خصم أو ائتمان) نقد استرداد القسيمة 	<ul style="list-style-type: none"> الاسم قنوات الاتصال العنوان رقم الهاتف البريد الإلكتروني رسالة نصية قصيرة تفضيلات الاتصال تاريخ الميلاد معلومات الأسرة (الحجم، الأعضاء، وما إلى ذلك)
بيانات المكافآت	البيانات السلوكية وتفضيلات المستخدم
<ul style="list-style-type: none"> نقاط الولاء التي تم الحصول عليها نقاط الولاء التي تم استردادها 	<ul style="list-style-type: none"> تاريخ استخدام القسيمة التسوق عبر الإنترنت الردود على الترفيات البريد الإلكتروني الموقع الإلكتروني رسائل البريد تفضيلات المنتج (على سبيل المثال، الولاء للعلامة التجارية) تفضيل الخدمة خدمة التوصيل إلى المنزل انقر واستلم التسوق داخل المتجر ملاحظات العملاء

تطبيق النظرية

تحتاج مدرستك إلى تحسين تشغيل كافيتيريا المدرسة. إنشاء سلسلة من شرائح العرض التقديمي التي توضح كيف يمكن لتكنولوجيا نقاط البيع في الكافيتيريا أن تحدث فرقاً. فكر في الفوائد التي يمكن أن يجلبها النظام لكل من الطلاب والمنظمة.

أنظمة التجارة الإلكترونية

يقصد بها التجارة عبر الوسائل الإلكترونية، ويشير المصطلح إلى شراء وبيع المنتجات أو الخدمات عبر الإنترنت باستخدام الإنترنت. إنه يكمل النهج التقليدي للمتاجر الفعلية أو يحل محله.

وقفة للتفكير

كم مرة تشتري منتجات أو خدمات عبر الإنترنت؟ ما مدى حرصك على خيارات الاتصال التي تختارها عند إعداد حساب جديد؟

تلميح

فكر في أنواع التواصل التي توافق عليها. فكر في فحص بعض الحسابات التي تستخدمها لتقليل رسائل البريد الإلكتروني غير المرغوب فيها التي تتلقاها.

توسيع الأفق

تقوم جميع أنظمة التجارة الإلكترونية بجمع ومعالجة بيانات المعاملات. في الواقع، يتم إنشاء كميات هائلة من البيانات في مراحل مختلفة من عملية التجارة الإلكترونية.

وُتستخدم هذه البيانات بعدة طرق بما في ذلك ما يأتي:

- التخطيط الاستراتيجي - على سبيل المثال، تُستخدم بيانات المعاملات لاتخاذ قرارات بشأن المنتجات والخدمات الجديدة التي يجب تطويرها، أو لفحص أنماط المبيعات لتحديد المنتجات أو الخدمات التي تقترب من نهاية عمرها وتحتاج إلى سحبها.
- التخطيط التشغيلي - على سبيل المثال، يتم استخدام البيانات للعثور على طرق جديدة للسوق وقنوات اتصال جديدة وما إلى ذلك. يتم استخدامه أيضًا لتوسيع نطاق الأعمال التجارية (يسمح الإنترنت ببيع العديد من المنتجات والخدمات في جميع أنحاء العالم).
- التخطيط اليومي - على سبيل المثال، يتم استخدام البيانات لإدارة الأفراد والفرق، لنقل الموظفين بين الأقسام للتغطية في حالات المرض أو الغياب، لإدارة النقل والخدمات اللوجستية الآنية استجابة لظروف الطقس، أو للعثور على أسرع الطرق عندما تكون هناك مشكلات في حركة المرور على الطرق.

تدعم أنظمة التجارة الإلكترونية أربعة أنواع رئيسية من المعاملات التجارية.

B2C (المبيعات بين الشركات والمستهلكين)

هذا هو الشكل الأكثر شيوعًا للتجارة الإلكترونية وتستخدمه متاجر البيع بالتجزئة وخدمات البث والاشتراك (مثل Amazon وNetflix وApple وما إلى ذلك). يوفر هذا النوع من المبيعات عددًا من الفوائد. على سبيل المثال، لا تفتح العديد من متاجر البيع بالتجزئة 24 ساعة في اليوم. تتيح التجارة الإلكترونية للعملاء طلب التسليم عبر الإنترنت. وهذا يمنح الشركات رؤية على مدار الساعة طوال أيام الأسبوع، دون الحاجة بالضرورة إلى فتح المتاجر عندما يكون الطلب منخفضًا.

B2B (المبيعات بين الشركات)

يُستخدم هذا النوع من التجارة الإلكترونية بين الشركات المصنعة وتجار الجملة والخدمات المالية والأمن والنقل والخدمات اللوجستية والخدمات القانونية وما إلى ذلك. على سبيل المثال، تستخدم العديد من منافذ البيع بالتجزئة التجارة الإلكترونية للتعامل مباشرة مع الشركات المصنعة وتجار الجملة. بعض تجار التجزئة يصرون على أن تستخدم المنظمات التجارية التي يتعاملون معها الأنظمة نفسها، لتسهيل نقل العروض والأسعار وتبسيط معالجة الطلبات.

C2C (المبيعات بين المستهلكين)

يُستخدم هذا النوع من التجارة الإلكترونية لدى المستهلكين لتبادل السلع والخدمات أو بيعها - على سبيل المثال، من خلال eBay®، Etsy®، وCraigslist®. يستخدم العديد من الأشخاص التجارة الإلكترونية لخفض النفقات الزائدة التي يتكفونها. أحد المجالات التي أصبح فيها هذا الأمر ممارسة شائعة هو سوق ملابس الأطفال والرضع، حيث يمكن استبدال المنتجات عالية الجودة ببساطة مقابل قيمة الشحن البريدي أو التوصيل. يرى الكثير من الناس أن هذا النوع من إعادة الاستخدام أخلاقي للغاية ومثال جيد للمسؤولية الشخصية والاجتماعية.

C2B (المبيعات بين المستهلكين والشركات)

هذا شكل أقل شيوعًا من التجارة الإلكترونية ويشمل الأفراد، مثل الفنانين والمصورين والمؤثرين، الذين يبيعون السلع أو الخدمات للشركات. منذ سنوات عديدة يكلف الفنانون والمصورون بمشاريع محددة، ولكن الأهمية المتزايدة لمؤثري الإنترنت ظاهرة حديثة.

تتطلب جميع هذه المعاملات التجارية منصات تجارة إلكترونية قوية وأمنة تقدم خدمات مثل الإعلانات والوظائف مثل مرافق البحث و(سلال التسوق) وعمليات الدفع الآمنة.

على الرغم من أنه يمكن تصميم أنظمة التجارة الإلكترونية خصيصًا للعملاء الأكبر حجمًا، إلا أن المؤسسات الأصغر عادةً ما تصمم الحلول المتاحة تجاريًا مثل Magento® وShopify®.

أجهزة إنترنت الأشياء (IoT)

إنترنت الأشياء (IoT) هو المصطلح المستخدم لوصف شبكة الأجهزة اليومية التي تجمع البيانات وترسلها عبر الإنترنت. أمثلة شائعة على أجهزة إنترنت الأشياء (الأشياء في إنترنت الأشياء) والبيانات التي تجمعها موضحة في الشكل 21.18. غالبًا ما تحتوي هذه الأجهزة على أجهزة استشعار لالتقاط البيانات. لديهم أيضًا برامج متخصصة لتوصيل وتبادل البيانات المجمعة مع أنظمة أخرى عبر اتصال بالإنترنت.

موضوعات ذات صلة

لمزيد من المعلومات حول استخدام قنوات البيانات والاتصالات في الأعمال، راجع الوحدة 1: أنظمة تكنولوجيا المعلومات.

يمكن أن تختلف أجهزة إنترنت الأشياء اختلافاً كبيراً في الحجم والتعقيد، من الآلات الصناعية إلى التكنولوجيا الشخصية القابلة للارتداء. لقد أصبحت أكثر شيوعاً وانتشاراً من أي وقت مضى وتسهم بشكل كبير في حجم البيانات التي تم إنشاؤها في القرن الحادي والعشرين.

تجمع أجزاء الذكاء الاصطناعي/التعلم الآلي في إنترنت الأشياء البيانات بطريقة تسمح بتحليلها واستخدامها لتحسين الأداء وتعزيز الكفاءة. على الرغم من قدرة إنترنت الأشياء على تحسين جودة حياة مستخدميه، إلا أنه ستكون هناك دائماً مخاوف بشأن الخصوصية واستخدام وأمان البيانات التي يتم جمعها ونقلها و(ربما) مشاركتها مع أطراف ثالثة.

ومع ذلك، فإن جميع الأجهزة المترابطة التي تسهم في إنترنت الأشياء هي مصادر محتملة للبيانات. على سبيل المثال، إذا كان لديك نظام أمان ذكي، فإن كل تفاعل مع النظام سيؤدي عادةً إلى توليد المعلومات الآتية:

- ضبط جهاز تنبيه التاريخ والوقت
- إذا كان النظام يحتوي على عدة مستخدمين برموز فردية، فسيُسجل من قام بتفعيل التنبيه
- إلغاء ضبط جهاز تنبيه التاريخ والوقت
- بيانات حول من قام بإلغاء ضبط التنبيه.

يمكن أن تحتوي بعض أنظمة الأمان على إعدادات للغرف المختلفة. على سبيل المثال، قد يتم ضبط النظام لحذف الغرف التي توجد بها حيوانات أليفة. في هذا السيناريو، عند ضبط المنبه، سيسجل النظام الغرف التي تم حذفها.



◀ التكنولوجيا القابلة للارتداء مثل هذه الساعة الذكية هي شكل شائع من إنترنت الأشياء. تستطيع هذه الأجهزة تسجيل مجموعة من المعايير المتعلقة بالصحة البدنية والنشاط.

مثال آخر هو مستخدم يرتدي ساعة ذكية تلتقط بيانات عن التمرين الذي تم إجراؤه على مدار اليوم. يمكن اختيار كل نوع مختلف من التمارين وتسجيل وقت التمرين، بالإضافة إلى السرعات الحرارية المحروقة. تسجل ساعات أبل الحديثة أيضاً معايير أخرى مثل معدل ضربات القلب وتشبع الأكسجين في الدم واكتشاف مستويات الضوضاء. حتى أنهم يمتلكون خاصية الكشف عن السقوط. يتم تسجيل جميع هذه البيانات على الجهاز ويمكن تنزيلها على الهاتف المحمول أو الحاسب الخاص بالمستخدم.

كم عدد الأجهزة المتصلة بالإنترنت التي تستخدمها كل يوم؟ قم بعمل قائمة بجميع الأجهزة التي يمكنك الوصول إليها.

وقفة للتفكير



فكر في بعض الأشياء الأقل وضوحاً مثل الثلاجات وتحديثات الحافلات التي تبث بثاً مباشراً أو أنظمة أمان المنزل.

تلميح

ماذا عن المقاييس الذكية أو المصابيح الذكية أو كاميرات المراقبة أو أنظمة مراقبة الأطفال أو الأقفال الذكية أو منظمات الحرارة أو تكييف الهواء؟

توسيع الأفق



الشكل 21.18 أمثلة على البيانات التي تولدها أجهزة إنترنت الأشياء والتي يمكن جمعها وتحليلها.

المساعدون الأنكياء

يعد المساعد المنزلي (أو الذكي) أحد الأمثلة الشائعة لجهاز إنترنت الأشياء. ومن الأمثلة المعروفة على ذلك ما يأتي:

- سيري من آبل
- أليكسا وإيكو من أمازون
- جوجل نيكست.

توجد هذه الأجهزة عادةً في المنزل أو المكتب. ويمكن أن تساعد على أتمتة المهام اليومية ويُتحكم فيها باستخدام أوامر صوتية بسيطة. تستخدم الأجهزة الذكاء الاصطناعي لتمييز الكلام المعتمد على السحابة لتحويل الأوامر الصوتية إلى نص. ثم يقوم الجهاز بتنفيذ الأمر - على سبيل المثال، عن طريق إرسال تعليمات إلى أجهزة إنترنت الأشياء الأخرى (مثل تشغيل/إطفاء الضوء)، أو عن طريق استعلام قاعدة بيانات للحصول على المعلومات، أو عن طريق إجراء عمليات بحث حية على الإنترنت. وعند الحاجة، يتم تحويل النتائج مرة أخرى من نص إلى كلام وإرسالها إلى المستخدم.

بفضل سهولة إعداد المساعد المنزلي ودمجه بسلسلة في الحياة المنزلية، أصبح المساعد المنزلي شائعًا بسرعة، خاصة لمن يواجهون تحديات جسدية أو عقلية.

على الرغم من أن معظم الأشخاص يظنون أن المساعدين الأنكياء يشاركون المعلومات مع المستخدم فقط، إلا أن هذه الأجهزة الذكية تسجل أيضًا نطاقًا من المعلومات لكل تفاعل، مثل: موقع المستخدم، ووقت التفاعل وتاريخه، والسؤال المطروح، والاستجابة المعطاة، ومقدار الوقت المستغرق للرد. يسأل معظم المساعدين أيضًا عما إذا كانت الإجابة قد ردت على السؤال بشكل صحيح أم لا. يسمح استخدام هذه المعلومات للخدمة بتحسين استجاباتها للمستخدمين الآخرين - وهذا ما يسمى حلقة التغذية الراجعة.

أنظمة التدفئة الذكية

أنظمة التدفئة الذكية الحديثة هي أجهزة إنترنت أشياء تزداد شعبيتها بشكل كبير. فهي مزودة بسلسلة من أجهزة استشعار درجة الحرارة والرطوبة التي تجمع البيانات في الوقت الفعلي. وتشمل مصادر البيانات التي قد توفرها هذه المستشعرات درجة الحرارة (درجات مئوية أو فهرنهايت) والرطوبة (كنسبة مئوية) والتواريخ والأوقات والمواقع وما إلى ذلك. تحتوي بعض هذه الأجهزة على أجهزة استشعار يمكنها اكتشاف أنماط الإشغال (عند استخدام الغرف أو عدم استخدامها). يمكن للأجهزة الأكثر تقدمًا استخدام الذكاء الاصطناعي لتحليل البيانات التي يجمعونها من أجل:

- توقع الأوقات التي سيتم فيها استخدام الغرف
 - تحديد أفضل إعدادات درجة الحرارة والرطوبة
 - تفعيل عناصر التحكم في الغرف قبل التسخين
 - إيقاف تشغيل الأنظمة قبل 60 دقيقة من احتمال مغادرة الشاغلين للغرفة.
- كل هذا يساعد على تقليل استخدام الوقود، ما يخفّض التكاليف.

أجهزة مراقبة صحية يمكن ارتداؤها

يمكن استخدام أجهزة مراقبة الصحة القابلة للارتداء بطريقتين مختلفتين تمامًا.

- يمكن استخدامها من جانب الأفراد الذين يرغبون في مراقبة صحتهم وفوائد الأنشطة البدنية في إطار برامج ممارسة التمارين وفقدان الوزن.
- يمكن للأطباء المتخصصين استخدامها في علاج الحالات الصحية ودعما – لا سيما عندما يمكن الوصول إلى البيانات عن بُعد.

يمكن لمكونات الذكاء الاصطناعي/التعلم الآلي لهذه الأجهزة استخدام البيانات التي تجمعها للتنبؤ بالأحداث الصحية المحتملة بناءً على البيانات في الوقت الفعلي، وإصدار التنبيهات بحسب الضرورة. أحد الأمثلة الراسخة على ذلك هو مراقبة وعلاج مرض السكري من النوع 1 والنوع 2. في مثل هذه الحالات، يرسل جهاز مراقبة الصحة القابل للارتداء البيانات إلى الجهاز المحمول للشخص، وينصحهم بالتصرف قبل حدوث حدث نقص أو ارتفاع السكر في الدم.

وقفة للتفكير



هل تستخدم أي تقنيات قابلة للارتداء؟ أو هل تعرف أي شخص يفعل ذلك؟ ماذا تفعل الأجهزة؟

تلميح

فكر في الأجهزة المرتبطة بالرياضة أو الموقع أو حتى أجهزة الألعاب القابلة للارتداء.

توسيع الأفق

هل يمكنك التفكير في عشرة أمثلة؟ قم ببعض البحث للعثور على المزيد.

الذكاء الاصطناعي في السيارات

استخدام الذكاء الاصطناعي/التعلم الآلي في السيارات (والمركبات الأخرى) على مدى العشرين عامًا الماضية قد أسفر عن العديد من الفوائد.

- تسمح بيانات الموقع الجغرافي للشركات برؤية تقدم العناصر التي يتم تسليمها عن طريق السكك الحديدية والطرق والبحر والجو. غالبًا ما تتم مشاركة البيانات مع العميل، الذي يمكنه رؤية الوقت المقرر للوصول على تطبيق جهاز محمول. بالإضافة إلى ذلك، تستخدم المؤسسات البيانات لتخطيط عمليات التسليم المستقبلية، لا سيما عندما يكون الوقت عاملاً مؤثرًا.
- تمتلك العديد من العائلات تطبيقات محمول على أجهزتها تشارك موقعها مع أفراد العائلة الآخرين، وهو أمر مفيد بشكل خاص عند السفر. بالإضافة إلى كونها مفيدة، يجد العديد من المستخدمين هذه المعلومات مطمئنة.
- يمكن لتطبيقات السفر إعادة توجيه رحلة السيارة تلقائيًا إذا كان هناك حادث أو عائق أمامها. ويمكنهم بعد ذلك إعادة حساب (وقت الوصول إلى الوجهة) ومشاركة المعلومات مع الآخرين في مجموعة عائلة السائق. يتم الحصول على هذه البيانات عندما يقوم المستخدمون الذين يواجهون الحادث أو العائق (بالإبلاغ) عن المشكلة من خلال تطبيق السفر. استنادًا إلى موقع المستخدم عند إعداد التقرير، يمكن للتطبيق تقدير خطورة المشكلة، ثم يشاركها مع مستخدمي الطريق الآخرين من خلال التطبيق.
- يمكن لأنظمة الذكاء الاصطناعي/التعلم الآلي التقاط معلومات حول سرعة القيادة (الحد الأدنى/الأقصى)، وأنماط التسارع والفرملة لتحليل سلوك القيادة. يمكن استخدام هذه البيانات لتقديم أسعار تأمين مخفضة. يمكن للسائقين الذين يظهرون سلوكيات غير لائقة أن يروا زيادة تكاليف التأمين الخاصة بهم.
- تستخدم الشركات بيانات استخدام وقود المركبات للتنبؤ بالتكاليف المستقبلية. تُستخدم بيانات الاستخدام أيضًا لإصدار تنبيهات وتحذيرات الصيانة التي تسمح للشركات والأفراد بحجز الخدمة وإصلاح سياراتهم قبل أن تتعطل. وهذا يسمح للشركة بالحفاظ على المستوى المأمول لخدمة العملاء.



مناقشة

فكر في مجموعة واسعة من أجهزة إنترنت الأشياء القابلة للارتداء التي أصبحت تندمج بشكل متزايد في حياتنا اليومية، مثل أجهزة تتبع اللياقة البدنية، والساعات الذكية، وأجهزة مراقبة الصحة. ناقش أنواع البيانات الشخصية التي تجمعها هذه الأجهزة عنك. فكر في هذه الأسئلة الرئيسية:

- في رأيك، ما تداعياتها على الخصوصية والأمان؟
- ما شعورك حيال التوازن بين الفوائد التي تقدمها هذه الأجهزة والبيانات الشخصية التي تجمعها؟
- كيف يمكن استخدام الذكاء الاصطناعي للحصول على رؤى من أجهزة إنترنت الأشياء المختارة حول عادات الشخص وسلوكياته؟
- شارك أفكارك وناقش أي تدابير تعتقد أنه يجب عليك اتخاذها لحماية خصوصيتك في أثناء استخدام هذه التقنيات.

مصادر بيانات الشركات

إن مجموعة مصادر البيانات التي تُنشئها وتستخدمها الشركات واسعة ومتنوعة بطبيعتها. وتُستخدم هذه البيانات لدعم عملياتهم اليومية (للشراء والإنتاج والتسويق والبيع والموارد البشرية وما إلى ذلك) ومساعدتهم على اتخاذ قرارات إستراتيجية فعالة. قد تتضمن مصادر البيانات الداخلية الشائعة:

- السجلات المالية، بما في ذلك البيانات المحاسبية
- بيانات الموارد البشرية (الموظف)
- بيانات المبيعات والتسويق
- بيانات العملاء (عادةً من خلال نظام إدارة علاقات العملاء)
- البيانات التشغيلية

تتقسم البيانات التشغيلية للشركة إلى بيانات حول:

- المخزون: المخزون/المنتجات التي تصنعها الشركة وتبيعها
- سلسلة التوريد: النظام الذي يأخذ المواد الخام ويحولها إلى منتجات ثم يسلم هذه المنتجات إلى المتاجر والعملاء
- الخدمات اللوجستية: إدارة عملية سلسلة التوريد

يمكن للشركات استخدام الذكاء الاصطناعي لتحليل البيانات بعدة طرق. على سبيل المثال، يمكن استخدام الذكاء الاصطناعي لتحليل بيانات المبيعات لتحديد الاتجاهات المستقبلية والتنبؤ بها. هذا شيء تقوم العديد من المؤسسات بالتحقيق فيه بالفعل.

عادةً ما يتم تخزين مثل هذه البيانات في قواعد البيانات الإلكترونية وجداول البيانات. قد تظل بعض المؤسسات تخزن هذه البيانات في مقراتها الخاصة ('في الموقع')، لكن العديد من الشركات الأخرى قد نقلوا بياناتهم إلى التخزين السحابي، الذي توفره شركات مثل أمازون أو جوجل أو مايكروسوفت. ويمكنهم أيضًا استخدام مراكز بيانات (العلامة البيضاء) (التي تنتجها إحدى الشركات، والتي تحمل علامة تجارية أخرى). أحيانًا قد يستخدمون حلولًا مجمعة، حيث يتم الاحتفاظ ببعض البيانات محليًا في الموقع والبعض الآخر في السحابة. الميزة الرئيسية لتخزين السحابة هي أن البيانات المخزنة تُنسخ احتياطيًا عبر عدد من الآلات الافتراضية، ما يعني تقليل فرصة فقدان البيانات.

بالنسبة لمعظم الشركات، يتضمن اختيار تخزين البيانات النظر في المهارات التقنية الداخلية، وتكاليف التخزين، والتوافر، والتشريعات المحلية ذات الصلة، والمخاوف المتعلقة بالأمان.



المهارات

المهارات المعرفية: العمليات
والإستراتيجيات المعرفية:

- حل المشكلات
- اتخاذ القرار

المهارات

المهارات المعرفية: العمليات
والإستراتيجيات المعرفية:

- التفكير الناقد
- المهارات المعرفية: الإبداع:
- الابتكار

المصطلح الرئيس

البيانات الضخمة – هي مكتبة من
البيانات سريعة النمو، وتُجمع هذه
البيانات من الإنترنت، ووسائل التواصل
الاجتماعي، وأجهزة إنترنت الأشياء
(IoT)، وأنظمة العالم الحقيقي، والهواتف
المحمولة وما إلى ذلك.

تأثير حجم مجموعة البيانات على تخزينها واستخدامها

كما تمت مناقشته أعلاه، لتخزين البيانات، ستحتاج المؤسسات إما إلى الحصول على حلول محلية (داخلية) وإما استخدام مراكز بيانات ذات علامة بيضاء وإما الاستثمار في خدمات التخزين السحابي (وإما استخدام مزيج من هذه الحلول). يجب مراعاة المشكلات الآتية:

- كلما كان حجم مجموعة البيانات أكبر، زادت سعة التخزين المطلوبة.
- يجب أن تكون أنظمة التخزين آمنة.
- يجب أن تكون أنظمة التخزين قابلة للتطوير لاستيعاب مجموعات البيانات سريعة النمو التي تجمعها الشركات.
- يجب تحقيق قابلية التوسع دون تعطل أو أي تراجع في الأداء أو زمن الوصول.

يمكن أن تكون تكاليف تخزين البيانات لمجموعات البيانات الكبيرة مكلفة للغاية، خاصة في حالة الحاجة إلى إتاحة البيانات دائماً (أي يمكن الوصول إليها عند الحاجة). يقدم العديد من مورفي الخدمات السحابية معدلات تخزين أقل تكلفة للبيانات التي يجب الوصول إليها بشكل أقل تكراراً. في مثل هذه الحلول، يتم وضع البيانات بشكل أساسي في حالة (التجميد العميق) ولا يمكن الوصول إليها عادةً على الفور.

ستكون تكلفة كل بايت لهذا النوع من تخزين البيانات أقل تكلفة بشكل ملحوظ من تكلفة تخزين البيانات المتوفرة دائماً. هذا لا يختلف عن وضع مدخلك في حساب التوفير المصرفي الذي يقدم فائدة أعلى - العائد أكبر، ولكن قد لا يتم سحب الأموال في أي لحظة.

استخدام البيانات الضخمة في حلول الذكاء الاصطناعي

تؤدي البيانات الضخمة دوراً مهماً في تطوير أنظمة الذكاء الاصطناعي وفعاليتها، فهي مكتبة من المعلومات سريعة النمو، وتُجمع هذه المعلومات من الإنترنت، ووسائل التواصل الاجتماعي، وأجهزة إنترنت الأشياء (IoT)، وأنظمة العالم الحقيقي، والهواتف المحمولة وما إلى ذلك، ما يعني ببساطة أنها الوفود الذي يشغل آلة الذكاء الاصطناعي.

يستخدم هذا الحجم الكبير من البيانات المصنفة وغير المصنفة لتدريب نماذج التعلم الآلي، ويتحقق ذلك عن طريق اختيار واستخراج الميزات ذات الصلة من كتلة البيانات الخام المتاحة.

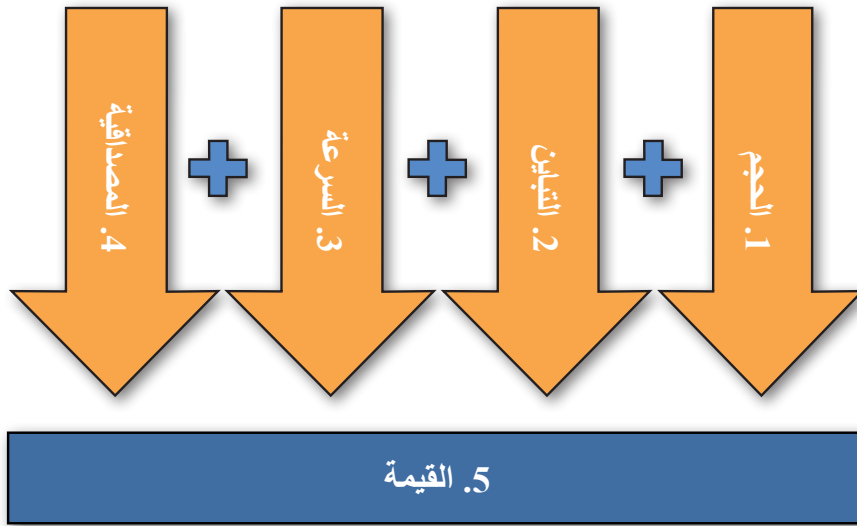
بالإضافة إلى ذلك، فإن الكميات الهائلة من البيانات المتاحة تمكن علماء البيانات من التحقق من صحة نماذجهم وتقييمها. ويضمن هذا الأمر قدرة نماذج الذكاء الاصطناعي على التعميم الجيد للبيانات التي لم تطلع عليها (حتى الآن)، بدلاً من التحيز المفرط تجاه البيانات المحدودة (أحياناً) التي تم تدريبها عليها.

أحد السمات الهامة للبيانات الضخمة هو السرعة التي يمكن بها استيعابها حديثاً في نظام الذكاء الاصطناعي. وهذا يساعد على تحديد المشكلات المحتملة (على سبيل المثال، الانحراف (الابتعاد عن الأصل)) ويضمن أن تظل التوقعات دقيقة في المواقف الواقعية المتغيرة بسرعة (على سبيل المثال، التغيرات في أسعار سوق الأسهم).

لا يمكن التحدي المتمثل في التعامل مع تعقيد البيانات الضخمة في سياق الذكاء الاصطناعي في تخزينها وتنظيمها فحسب، بل أيضاً في فهمها وتفسيرها بشكل فعال.

المتقالات الخمسة للبيانات الضخمة

المتقالات الخمسة للبيانات الضخمة هي المعايير المستخدمة لوصف الخصائص (والتحديات) الرئيسة للبيانات الضخمة، وعادةً ما تكون العناصر الأربعة الأولى (الموضحة في الشكل 21.19) أساسية لتحقيق العنصر الخامس: قيمتها الإجمالية.



الشكل 21.19 الأبعاد الأربعة (أو الخمسة) للبيانات هي المعايير المستخدمة لتقييم مدى ملائمة البيانات المكتسبة لغرض الذكاء الاصطناعي/التعلم الآلي.

الحجم

يشير إلى الكمية الهائلة من البيانات التي يتم إنشاؤها وتخزينها - يمكن أن يتراوح حجمها من تيرابايت إلى زيتابايت.

بالنسبة للذكاء الاصطناعي، تعد الكميات الكبيرة من البيانات أمراً بالغ الأهمية لأنها توفر مجموعة بيانات أكثر شمولاً لتدريب نماذج التعلم الآلي، ما يؤدي إلى دقة أفضل للتنبؤات وتحسين عملية صنع القرار.

التباين

يشير إلى الأنواع المختلفة من تنسيقات البيانات ومصادرها، بما في ذلك البيانات المنظمة وشبه المنظمة وغير المنظمة والمصادر مثل النصوص والصور ومقاطع الفيديو والسجلات.

بالنسبة للذكاء الاصطناعي، غالباً ما تعمل مجموعة متنوعة من البيانات المتاحة على تحسين عملية التعلم الآلي. ويرجع ذلك إلى أن زيادة تنوع البيانات تتيح الوصول إلى مجال أوسع من المعلومات، ما يتيح المقارنة بشكل أفضل وتحديد الاتجاهات والارتباطات، وربما لم يكن هذا الأمر ممكناً باستخدام بيانات أكثر تجانساً (أو "نفس" البيانات).

السرعة

يشير إلى السرعة التي يتم بها إنشاء البيانات وتناولها ومعالجتها.

بالنسبة للذكاء الاصطناعي، تعني السرعة العالية أن البيانات تتدفق إلى الأنظمة آنية أو شبه آنية. البيانات عالية السرعة تجعل أنظمة الذكاء الاصطناعي أكثر استجابة لتغيرات الظروف الواقعية، وتجعل توقعاتها أكثر دقة.

المصداقية

هذا هو قياس لجودة ودقة وموثوقية البيانات التي تم تناولها.

أنظمة الذكاء الاصطناعي جيدة بقدر جودة البيانات التي تُغذى بها. فجودة المخرجات هي انعكاس لجودة المدخلات. سنتنشر بيانات ذات جودة سيئة وتتضخم عن طريق نظام الذكاء الاصطناعي، ما يؤدي حتماً إلى نتائج معيبة وغير دقيقة. لهذا السبب، فإن التأكد من أن البيانات من مصدر موثوق هو خطوة مهمة في أي مشروع ذكاء اصطناعي.

القيمة

تحدد تقلبات البيانات وصحتها وحجمها ما إذا كانت البيانات ذات مغزى وجديرة بالتحليل - وبعبارة أخرى ما إذا كانت ذات قيمة.

القيمة معيار حاسم للشركات التي تستخدم أنظمة أو خدمات الذكاء الاصطناعي، لأن إنشاء أنظمة أو خدمات الذكاء الاصطناعي عادة ما يكون مكلفًا، لذلك يجب أن تكون البيانات المستخدمة عالية الجودة لإعطاء نتائج جيدة.

تحتاج الشركات إلى أن يكون لديها فكرة عن العائد على الاستثمار الذي من المحتمل أن تجنيه من استثمارها. من وجهة النظر التجارية، قد يشمل ذلك نتائج مثل:

- تحسين الكفاءة والإنتاجية
- تحسين (وزيادة سرعة) صنع القرار
- تجربة محسنة للعملاء
- منتجات وخدمات جديدة ومبتكرة
- انخفاض المخاطر والتكلفة (تحسين الربحية)
- تحسين استخدام الموارد الحالية
- تحسين مرونة وقابلية العمليات للتوسع.

البيانات النوعية مقارنة بالبيانات الكمية

هل تعرف الفرق بين البيانات النوعية والبيانات الكمية؟ اكتب خمسة أمثلة على البيانات الكمية وخمسة أمثلة على البيانات النوعية.

وقفة للتفكير



فكر في أنواع الأسئلة التي طُرحت عليك في أي استبيان.

تلميح

فكر في الأسئلة التي قدمت لها إجابات بسيطة والأسئلة التي كنت قادرًا على الرد عليها بمزيد من الكتابة.

توسيع الأفق

تأتي البيانات عادةً في شكلين: نوعي أو كمي. خصائص كل نوع من البيانات موضحة في الجدول 21.6.

الجدول 21.6 مقارنة البيانات النوعية بالبيانات الكمية مع توضيح الفروق بين نوعي البيانات

البيانات النوعية	البيانات الكمية
<ul style="list-style-type: none"> • تصف الصفات أو الخصائص • غير رقمي • تركز بقوة على 'ماذا؟'، 'لماذا؟'، 'كيف؟' • تم جمعها من خلال الملاحظات والمقابلات والأسئلة المفتوحة واستطلاعات العملاء، وما إلى ذلك. 	<ul style="list-style-type: none"> • رقمي، على سبيل المثال، الكميات، المبالغ، الدرجات، النطاقات، وما إلى ذلك. • تم جمعها عبر مجموعات البيانات المنظمة والأسئلة المغلقة • تصنيفات على نمط ليكرت، مثل '3 من 4' • سهل القياس والتحليل الإحصائي

ويمثل كلا النوعين قيمة كبيرة عند إجراء البحوث واتخاذ القرارات. غالبًا ما يكملان بعضهما بعضًا لتوفير فهم أكثر شمولًا لحالة أو موضوع. على سبيل المثال، قد يكون السؤالان التاليان (الَّذان يُطرحان على أحد العملاء) كاشفان للأمور ومفيدان بالقدر نفسه.

السؤال 1 (النوعي): (ما الذي يعجبك أكثر في المنتج؟)

السؤال 2 (الكمي): قِيم المنتج من أصل 5 نجوم: (1=ضعيف، 5=ممتاز).

سيؤثر اختيار الأسئلة النوعية أو الكمية تأثيرًا كبيرًا في مشروع الذكاء الاصطناعي بسبب البيانات الناتجة عن استخدامها. إذًا، ما نوع البيانات الأفضل لمشاريع الذكاء الاصطناعي؟ في الحقيقة، لكل نوع من البيانات نقاط القوة والضعف الخاصة به، لكن يمكن القول إن مشاريع الذكاء الاصطناعي الأكثر موثوقية تتضمن مزيجًا متزنًا من الاثنين.

البيانات الكمية

البيانات الكمية أكثر تنظيمًا من البيانات النوعية، وهي مثالية لمعظم خوارزميات التعلم الآلي (الانحدار والتصنيف وما إلى ذلك)، لأنها تستهلك المدخلات الرقمية بكل سلاسة. بالإضافة إلى ذلك، تعمل الأساليب الإحصائية التقليدية بشكل جيد، ومن السهل تتبع التغييرات في البيانات الكمية بمرور الوقت.

على الجانب السلبي، قد يتم فقدان الكثير من الفروق الدقيقة باستخدام نظام (تسجيل) بسيط. بالإضافة إلى ذلك، من الممكن أن تتضمن الأسئلة تحيزًا اعتمادًا على كيفية صياغتها، وهذا أمر قد يكون من الصعب اكتشافه إذا تم جمع البيانات الكمية فقط.

البيانات النوعية

عادةً ما توفر البيانات النوعية معلومات أكثر تفصيلًا، تحتوي على التعبير عن الإعجاب وعدم الإعجاب والمشاعر والتصورات والدوافع والأسباب الكامنة. وغالبًا ما توفر هذه المعلومات السياق الذي يشرح أي نتائج تم الحصول عليها من البيانات الكمية.

على الجانب السلبي، البيانات النوعية غير منظمة وقائمة على النصوص في طبيعتها وستتطلب معالجة أكثر تعقيدًا قبل أن تتمكن خوارزميات التعلم الآلي من استخدامها. على سبيل المثال، قد تكون هناك حاجة إلى تقنيات معالجة اللغة الطبيعية لتحويل النص إلى تنسيق بيانات للمعالجة (على سبيل المثال، من خلال التصنيف والترقيم).

هناك اعتبار آخر وهو أن البيانات النوعية يمكن أن تؤدي إلى مخرجات أكثر تعقيدًا. قد تكون هناك حاجة إلى معرفة أكثر تحديدًا بالمجال (أي فهم الموضوع المعني الذي يركز عليه مشروع الذكاء الاصطناعي) لتفسير هذه المخرجات بشكل صحيح واكتساب رؤية مناسبة.

تطبيق النظرية

تم بناء قاعة مغلقة لاستضافة الحفلات الموسيقية والمسرحيات وإقامة الفعاليات الخيرية. يتميز المكان بنهج حديث يستخدم أدوات الذكاء الاصطناعي لإدارة مناطق جلوس الزوار، وأنظمة الصوت، والردهات، ومنافذ الطعام، ومرافق الحمامات، والتنقل، وأنظمة التدفئة والتهوية، وبروتوكولات السلامة من الحرائق.

طُلب منك عمل استبيان عبر الإنترنت لجمع التعليقات من الزوار بعد حضورهم إلى المكان. بمجرد جمعها، سَتُحلل البيانات المجمعة بحلول الذكاء الاصطناعي وستُستخدم للأغراض الآتية:

- تحسين الخدمات والمرافق
- قياس رضا الزوار
- توفير المقارنة المعيارية مقابل معايير المجال والمنافسين
- تحسين علاقات الزوار ومستويات التفاعل
- السماح باتخاذ القرارات بناءً على البيانات

هذه المهمة هي مهمة تصميم لجعلك تفكر في نوع الأسئلة التي تحتاج إلى طرحها. هل سيتولد عن هذه الأسئلة بيانات كمية أم نوعية؟ أم هل يجب أن يكون هناك مزيج من النوعين؟ كيف ستؤثر اختياراتك في كيفية معالجة نظام الذكاء الاصطناعي للبيانات؟ قم بإنشاء نموذج استبيان وقم بتجريبه مع زملائك.

أنشطة معالجة البيانات واستخدامها في الذكاء الاصطناعي

يتم استخدام العديد من أنشطة معالجة البيانات الشائعة في تطوير حلول الذكاء الاصطناعي.

التحليلات الوصفية

هذه هي عملية تحليل البيانات التاريخية لفهم ما حدث سابقًا. يوفر رؤى حول الأنماط والاتجاهات التي يمكن استخدامها لتدريب نموذج التعلم الآلي بشكل صحيح.

تطبيق واقعي لهذا سيكون منظمة تجزئة تقوم بفحص مبيعاتها خلال الـ 12 شهرًا الماضية، وبعد تحديد منتجها الأكثر شعبية، تستخدم التحليل الوصفي لتحديد كيف ولماذا ومتى أصبح شائعًا بين العملاء.



سيضمن ذلك قيام الذكاء الاصطناعي بفحص بيانات المبيعات القديمة، وتحديد الاتجاهات والأنماط، وتقديم الرؤية المناسبة. في العادة، لن يستكشف هذا التحليل الوصفي الآثار المستقبلية (على سبيل المثال، في هذه الحالة كيف ستتطور مبيعات المنتج الشهير خلال الـ 12 شهرًا القادمة).

التحليلات التشخيصية

هذه عملية تحاول فهم لماذا وقعت أحداث أو سلوكيات معينة. يمكن أن يساعد تحديد هذه العوامل السببية على تحسين تنبؤات الذكاء الاصطناعي والدقة الشاملة للنتائج.

وسيكون التطبيق الواقعي لهذا هو مثال مؤسسة تجارية تدير أسطولاً من حافلات النقل العام. تريد الشركة التأكد من أن جميع مركباتها تعمل بكفاءة وأمان. كما أنها تريد تقليل الأعطال وطلبات الصيانة غير المتوقعة - وكلها تزيد التكاليف وتؤدي إلى عدم رضا العملاء.

تقوم الشركة بتزويد كل حافلة بأجهزة استشعار إنترنت الأشياء التي تراقب ضغط الإطارات ومستويات البطارية والكبح والسرعة وأداء المحرك وما إلى ذلك. وهذا يسمح لها بجمع البيانات في الوقت الفعلي واستخدام التحليل التشخيصي لتقييم حالة كل مركبة وعملياتها. هذا، بدوره، يسمح بتحديد الأنماط أو الحالات الشاذة التي قد تشير إلى تآكل غير مبرر، والتنبؤ بفشل مكون ميكانيكي أو كهربائي قبل أن يحدث فعلياً.

يمكن بعد ذلك الإبلاغ عن هذه المعلومات إلى فريق الصيانة في المؤسسة. وهذا يسمح لهم بجدولة صيانة سريعة واستباقية نسبياً. وهذا من شأنه أن يمنع حدوث الأعطال، ويزيد من مدة عمل الحافلات، ويبقي العملاء سعداء.

الاستخراج والتحويل والتحميل (ETL)

هذه معالجة هندسية للبيانات تتكون من ثلاث مراحل أساسية:

- 1 استخراج البيانات من عدد من المصادر المختلفة (على سبيل المثال، الملفات وواجهات برمجة التطبيقات وقواعد البيانات وما إلى ذلك).
- 2 تعيين البيانات من المصادر الأصلية إلى مجموعة جديدة من البيانات، بتنسيق ثابت. قد يتطلب هذا الأمر أحد أشكال التحويل (على سبيل المثال، الأحرف الكبيرة إلى الأحرف الصغيرة بالإنجليزية، والتواريخ الأمريكية إلى التواريخ الأوروبية، وما إلى ذلك).
- 3 تحميل مجموعة البيانات الجديدة هذه إلى موقع قابل للمشاركة، عادةً على السحابة.

تطور أدوات تحويل البيانات المعتمدة على السحابة وتخزين السحابة الأقل تكلفة قد شجع مهندسي البيانات على عكس مراحل 'التحويل' و'التحميل'، الأمر الذي يخلق ما يسمى بالاستخراج والتحويل والتحميل (ELT). أصبح هذا بسرعة الخيار المفضل.

التحليلات التنبؤية

استخدام التقنيات الإحصائية ونماذج التعلم الآلي للتنبؤ بالأحداث المستقبلية بناءً على البيانات التاريخية. والمثال الجيد على استخدام هذه التقنية هو شركة تأمين تستخدم التحليلات التنبؤية لتحديد ملف المخاطر لعمل محتمل. وهو ما من شأنه أن يسمح للشركة بتحديد علاوة واقعية للعميل، وهي علاوة توازن بين المطالبات المحتملة والتكاليف التشغيلية وربما رباً مقبولاً للخدمات التي تقدمها.

التحليلات الوصفية

هذه هي عملية اقتراح الإجراءات التي تجعل نتائج معينة أكثر احتمالاً أو تقلل فرصة حدوث بعض المخاطر على الإطلاق.

التطبيق الواقعي لهذه العملية سيتم في عالم البيع بالتجزئة. يمتلك العديد من تجار التجزئة الكبار مراكز توزيع متعددة ومتاجر بيع بالتجزئة تنتشر في جميع أنحاء البلد. وهم بذلك يواجهون التحدي المتمثل في موازنة مستويات الأسهم عبر شبكات البيع. إنهم بحاجة إلى التأكد من أن متاجرهم تحتوي على الكميات المناسبة من السلع لتلبية طلبات العملاء المحليين. كما أنهم بحاجة إلى تجنب مخاطر المخزون الزائد غير الضرورية.

يمكن للأنظمة التي تعتمد على الذكاء الاصطناعي معالجة مستويات البيانات التي تم جمعها من المواقع المختلفة. ويمكن لهذه الأنظمة بعد ذلك تحليل مستويات المخزون وأوقات الشحن وتوقعات طلب العملاء وعوامل أخرى. ويمكنها استخدام هذه المعلومات لمعرفة أنماط السلوك والتنبؤ بالمطلوبات المستقبلية للعناصر المختلفة في مواقع مختلفة. يمكن لشركة بيع بالتجزئة استخدام هذه الأفكار لبيان حجم التخزين الصحيح، وحركة وتخزين العناصر عبر شبكتها - ما يقلل الهدر، ويقلل فرص المبيعات المفقودة، ويزيد الأرباح.

تمثيل البيانات

هو عملية تقديم البيانات في شكل رسومي (مثل الرسوم البيانية والرسوم التوضيحية، وما إلى ذلك) التي تساعد على توصيل المعلومات الرقمية أو الفنية الأكثر تعقيدًا بطريقة أكثر قابلية للفهم للفئة المستهدفة.

ب2 جمع البيانات وتجهيزها

يركز هذا الموضوع على كيفية جمع البيانات وإعدادها. سيساعدك ذلك على فهم العملية المستخدمة لتحديد البيانات اللازمة من مصادر البيانات المناسبة. أحد أهم العوامل هو فهم أن البيانات قد لا تكون بتنسيق مناسب وقد لا تكون منظمة كما هو مطلوب. هذا يعني أن هناك حاجة إلى عمل إضافي لإعداد البيانات للاستخدام. بالإضافة إلى ذلك، يغطي الموضوع الاعتبارات الأخلاقية والتشريعية والأمنية التي يجب مراعاتها.

مبادئ وعمليات جمع البيانات وإعدادها لضمان بيانات عالية الجودة تناسب حلول الذكاء الاصطناعي

يوضح هذا الجزء أفضل الممارسات التي يجب اتباعها عند جمع البيانات وإعدادها. يجب أن نتذكر دائمًا أن أحد الأسباب الأساسية لفشل مشاريع الذكاء الاصطناعي هو سوء اختيار البيانات وجمعها. يوجد خمس نقاط رئيسية يجب تذكرها عند جمع البيانات وإعدادها لحل الذكاء الاصطناعي.

1 فهم المتطلبات بشكل كامل

تأكد من أن البيانات التي تم جمعها ذات صلة وتفي بمتطلبات مشروع الذكاء الاصطناعي الذي يتم العمل عليه.

حدد أنواع البيانات المطلوبة وتنسيقاتها ومعدل تكرار جمعها (كل ساعة، يوميًا، شهريًا، وما إلى ذلك).

2 التأكد من دقة البيانات ومن اتباع ضوابط النزاهة

تحقق من سلامة البيانات - تأكد من أن مصدر بياناتك الفعلي قانوني وآمن. يعد عقد البيانات طريقة جيدة لضمان هذه المعلومات وتسجيلها.

تحقق من جودة البيانات - تحقق من عدم الاتساق والبيانات المكررة والبيانات المفقودة والأخطاء.

3 الحد من التحيز غير المقصود من خلال ضمان تنوع البيانات

تأكد من أن البيانات تحتوي على القدر والفئة المناسبة من التنوع (على سبيل المثال، الديموغرافي والجغرافي، وما إلى ذلك). التأكد من أن ارتباط البيانات بسياق/طبيعة المشكلة التي تتم معالجتها.

4 التأكد من تطهير البيانات بشكل صحيح

تطهير القيم المفقودة، والقيم الشاذة، والبيانات العشوائية غير الضرورية في مجموعة البيانات (البيانات العشوائية هي في الأساس بيانات غير ذات صلة يمكن أن تؤدي إلى مواقف غير متوقعة).

تحويل البيانات باستخدام تقنيات التخطيط بحيث تكون جاهزة للاستخدام في نموذج تعلم آلي، وما إلى ذلك.

المهارات

المهارات المعرفية: العمليات والإستراتيجيات المعرفية:

- حل المشكلات
- التفسير

مهارات التواصل الشخصي: أخلاقيات العمل/الضمير

- الإنتاجية

موضوعات ذات صلة

راجع أجزاء مفهوم التحيز والحد من تأثير التحيز في مجموعات البيانات. اسأل نفسك عن سبب أهمية معالجة قضايا التحيز لجودة النتائج.

موضوعات ذات صلة

راجع الجزء الخاص بجمع البيانات وتجهيزها. فكر في ما يجب القيام به من أجل إعداد البيانات لاستخدامها في مشروع. لماذا يجب إعداد البيانات؟

5 اتباع الإرشادات الأخلاقية والاعتبارات القانونية في جميع الأوقات

قم بإخفاء هوية معلومات التعريف الشخصية (PII) حيثما أمكن ذلك.
واحترم قوانين حماية البيانات والخصوصية المحلية (على سبيل المثال، اللائحة العامة لحماية البيانات في أوروبا).

بحث

تتخصص مؤسستك في تقديم حلول تكنولوجيا المعلومات المصممة خصيصًا للقطاع المالي. نظرًا للظروف الصعبة في هذا المجال، تواجه الشركة صعوبة في الاحتفاظ بالموظفين ذوي المستوى الفني المناسب.

ويبدو أن إحدى المسائل الرئيسية هي أن عدد الطلبات الواردة لأي وظيفة شاغرة مُعلن عنها مرتفع للغاية بحيث يتعذر على فريق الموارد البشرية معالجتها بسرعة.
تم تقديم اقتراح لبناء حل تنبؤي بالذكاء الاصطناعي. سيتم تدريب هذا الحل على محتويات نماذج الطلبات التي سبق تقديمها ونتائج المقابلات. سيتم استخدامه بعد ذلك لتوفير نظام فرز أولي لتصديق نطاق الطلبات، قبل أن يقوم فريق الموارد البشرية بمعالجتها.
ابحث عن المشكلات التي يجب أن تضعها في اعتبارك عند محاولة ضمان جمع بيانات عالية الجودة والتحضير لهذا النوع من حلول الذكاء الاصطناعي.

اختيار البيانات ومصادر البيانات المناسبة

عادة ما يكون نظام الذكاء الاصطناعي جزءًا من سلسلة التوريد الكاملة، يستخدم البيانات ويوفر النتائج التي تساعد المستخدمين النهائيين على اتخاذ قرارات مستنيرة.
ولكي يعمل نظام الذكاء الاصطناعي بفعالية، من الضروري اختيار نوعية البيانات ومصادر البيانات المناسبة، وهو ما ينطوي على عدد من الخطوات الرئيسية.

تحديد أهداف المشروع أو المهمة

يساعد فهم أهداف مشروع الذكاء الاصطناعي على تحديد نوع البيانات المطلوبة للإجابة عن السؤال (الأسئلة المحدد).

على سبيل المثال، فإن النموذج المطلوب لإجراء التنبؤات الجوية سيحتاج إلى بيانات الطقس السابقة لبيت تدريبه بنجاح. ومع ذلك، إذا كنت تريد من الذكاء الاصطناعي التنبؤ باحتمالية حدوث السرقات في منطقة معينة، فستحتاج إلى الوصول إلى سجلات حوادث الشرطة وأوامر المطالبات التأمين على المنازل. ببساطة، تتطلب الأهداف المختلفة بيانات مختلفة.

مراجعة البيانات المتاحة

من الممكن أن يكون لديك بالفعل حق الوصول إلى البيانات التي تطلبها. إذا كان هذا هو الحال، فلا داعي للحصول عليها بشكل منفصل.

على سبيل المثال، قد تقوم المؤسسة بشكل روتيني بجمع البيانات المتعلقة بالأعمال التي تحتاج إليها في إطار عملياتها اليومية، لذلك يجب أن تحظى هذه البيانات بتخزين آمن. قد لا تكون بالتنسيق الصحيح، لكنها تمثل نقطة انطلاق جيدة، ما يوفر للشركة الوقت والمال الثمينين.

تحديد مدى سلامة وملاءمة البيانات الحالية المتاحة

إذا لم تكن البيانات متاحة بسهولة داخل الشركة، فيجب الحصول عليها. لذلك فإن السؤال التالي هو (من أين؟) من المحتمل ألا يكون أي مصدر بيانات موجود يفي بالمعايير الخاصة بك بسهولة. في الواقع، هذا ليس نادرًا. إذا كان هذا هو الحال، فقد يكون من الضروري إنشاء أنظمة للمراقبة أو التغذية الراجعة لجمع البيانات التي تحتاج إليها. على سبيل المثال، قد تقوم شركة بستنة ترغب في الحصول على بيانات لزيادة إنتاجها بتركيب مستشعرات إنترنت الأشياء في دفيئة لمراقبة ونقل درجة الحرارة والرطوبة كل ساعة لإنشاء سلسلة زمنية موثوقة من البيانات.

المهارات

- المهارات المعرفية: العمليات والإستراتيجيات المعرفية:
- اتخاذ القرار

موضوعات ذات صلة

لمزيد من التفاصيل حول مصادر البيانات، راجع الجزء الخاص بتحديد مصادر البيانات المناسبة والموثوقة والوصول إليها في هذه الوحدة.

المصطلح الرئيس

عدم الملاءمة – عندما يكون نموذج البيانات بسيطاً لأن البيانات غير كافية لتحديد العلاقات بين المتغيرات بدقة.

موضوعات ذات صلة

لمزيد من التفاصيل حول مصادر البيانات، راجع الجزء الخاص بالمتقابلات الخمسة للبيانات الضخمة في هذه الوحدة.

مناقشة

في عام 2017، قالت الإيكونوميست، 'إن المورد الأكثر قيمة في العالم لم يعد النفط، بل البيانات.' ناقش مع زملائك لماذا قد يكون هذا هو الحال.

بمجرد تحديد المصدر، يجب أن يكون الاعتبار الحاسم هو: هل تحتوي هذه البيانات على الميزات الضرورية لمعالجة هدف مشروع الذكاء الاصطناعي؟ بمعنى آخر، هل تجيب عن السؤال الأساسي للمشروع؟ من المهم ملاحظة أن هذا السؤال ينطبق بالتساوي على المصادر الداخلية والخارجية.

يرد في ما يأتي قائمة مرجعية للأسئلة التي يجب طرحها.

- هل من السهل الوصول إلى البيانات؟
- هل مصدر البيانات موثوق وصادق؟
- هل البيانات دقيقة بشكل عام؟
- هل تحتوي البيانات على تنوع مناسب (تنوع النطاق لتجنب التحيز)؟
- هل تحتوي البيانات على الحجم الكافي لكل من التدريب والاختبار (لتجنب عدم ملاءمة النموذج)؟
- هل البيانات حديثة (أي، هل تم تجميعها في فترة قريبة تجعلها صالحة؟)
- هل البيانات ذات نوعية جيدة؟
- هل هناك أي قيود قانونية أو فنية على استخدام البيانات؟
- هل هناك أي مشكلات أخلاقية في استخدام هذه البيانات؟ (على سبيل المثال، هل هناك أي مشكلات في معلومات تحديد الهوية الشخصية أو تأثيرات مجتمعية محتملة؟)
- إذا كان المشروع يحتاج إلى تدفق مستمر من البيانات (على سبيل المثال، لإعادة التدريب)، فهل سيكون هذا متاحاً باستمرار؟

بشكل حاسم، قد تلاحظ أن العديد من هذه الأسئلة تشير مباشرة إلى المتقابلات الخمسة التي تمت مناقشتها بالفعل. هذا لأن الهدف من طرح هذه الأسئلة هو تحديد البيانات ذات القيمة الجيدة.

من الجيد إجراء تحليل أولي للبيانات للإجابة عن هذه الأسئلة. وإذا كانت الإجابات عن هذه الأسئلة تشير أي شكوك حول مصدر البيانات الذي تفكر فيه، فقد ترغب في تجربة مصدر بيانات آخر، إن أمكن.

تحديد أوجه القصور في البيانات الحالية في ما يتعلق بالأهداف

عادةً ما يكون إجراء اختبار تجريبي صغير لتقييم مدى جودة تدريب البيانات للخوارزمية التي اخترتها فكرة جيدة. سيسمح لك باختبار دقة النموذج. هذا مفيد دائماً قبل الالتزام بشراء الترخيص لمصادر البيانات الخارجية، والتي قد تكون مكلفة للغاية.

ستتمكن من إثبات/نفي ما إذا كانت البيانات المتاحة تساعد على الإجابة عن السؤال المحدد. يعد الحصول على تعليقات من أصحاب المصلحة الآخرين (علماء البيانات الآخرين وخبراء المعرفة في المجال وما إلى ذلك) مفيداً أيضاً في هذه المرحلة.

جمع البيانات وتجهيزها للمشروع

يجب على مستخدمي البيانات الحرص على الحصول على البيانات التي ستكون مناسبة للمشروع، سواء من حيث الهيكل أو الشكل. قد يؤدي عدم القيام بذلك إلى عدد من الصعوبات المحتملة في مخرجات المشروع. في هذا الجزء ستتعرف بعض الطرق التي يمكن أن يحدث بها ذلك والتأثير الذي سترتب عليها.

وقفة للتفكير

هل سبق لك أن جمعت بيانات لمشروع ما؟ هل قدم شخص ما بيانات؟ هل كانت البيانات متاحة بسهولة على الإنترنت؟ أم كان عليك إنشاء استبيان؟

تلميح

فكر في السياق الذي كنت بحاجة فيه إلى البيانات والأسباب التي دفعتك لاختيار مصدر بيانات معين.

توسيع الأفق

لماذا احتجت إلى استخدام هذا المصدر بالذات؟ ما الذي كان بإمكانك فعله أيضاً؟

المصطلح الرئيس

تمام الملاءمة – الحالة التي تتناسب فيها البيانات مع نموذج إحصائي بدقة بحيث لا يمكن استخدامها في أمثلة أكثر عمومية، أي أن تركيزها ضيق أو محدد للغاية.

مفهوم التحيز

إحدى أكثر المشكلات شيوعاً التي تحدث عند إعداد مجموعات البيانات لاستخدامها في مشروع هي وجود تحيز في البيانات. يحدث التحيز عندما يتم تمثيل ميزات معينة من مجموعة البيانات بشكل متكرر و/أو يتم ترجيح استخدامها بشكل أكبر. يوجد في الواقع العديد من أشكال التحيز (على سبيل المثال، التحيز في الاختيار، التحيز في التمثيل، التحيز الثقافي، التحيز التاريخي، وما إلى ذلك) وبعضها يمكن أن يصعب تحديده للغاية. هذا الأمر يصبح أكثر صعوبة بسبب حقيقة أننا، بصفتنا بشرًا، لدينا أصلاً مجموعة من التحيزات اللاواعية التي يمكن أن تؤثر في قدرتنا على اتخاذ القرارات.

يمكن أن تؤدي التحيزات غير المحددة (أو غير المكتشفة) إلى:

أخطاء في التحليل الناتج: يمكن أن يكون جمع البيانات متحيزاً عن غير قصد تجاه مجموعة معينة، ما قد يتسبب في حدوث أخطاء في التحليل الناتج. تعمل خوارزميات التعلم الآلي بشكل أفضل عندما تكون عامة، أي قادرة على التكيف مع البيانات الجديدة بسهولة.

مستويات دقة أقل: غالباً ما تتأثر مستويات الدقة عندما لا تمثل البيانات المحددة جميع الجوانب ذات الصلة، وتشير مستويات الدقة المنخفضة بشدة إلى أن النموذج لا يعمل على النحو الأمثل (ما يعني تمام ملاءمة النموذج للبيانات المتحيزة).

نتائج منحرفة ومشوهة: عند اختراق مستويات الدقة، يُحتمل جداً إنتاج نتائج منحرفة عند معالجة البيانات المعيبة وتحليلها. هذا يمكن أن يسبب مشكلات بسهولة في العالم الحقيقي. يرد في ما يأتي بعض المخاوف المحتملة:

- توصيات متحيزة في توظيف موظفين جدد
- توصيات متحيزة عند الموافقة على القروض والرهون العقارية وما إلى ذلك.
- توصيات متحيزة عند مراجعة عروض التأمين الصحي
- توصيات متحيزة عند تحديد المجرمين المحتملين في لقطات كاميرات المراقبة.

في حين أنه من غير الصحيح القول بأن (كل التحيز سيء)، فمن المحتمل أن يكون من العدل القول إن (التحيز غير المحدد - خاصة في بيانات التدريب - من المحتمل أن يمثل مشكلة كبيرة).

كما ذكرنا أعلاه، هناك العديد من أنواع التحيز المختلفة.

تحيز غير مقصود

يمكن أن يحدث التحيز غير المقصود لعدد من الأسباب المختلفة، بما في ذلك طرق جمع البيانات الفعلية المستخدمة (على سبيل المثال، بسبب مشكلات في أخذ العينات أو القياس).

كما يوحي الاسم، فإن هذا النوع من التحيز يرجع في الغالب إلى الجهل أو الإهمال، وغالباً ما يحدث بسبب خطأ بشري. تحدث مشكلة متكررة وشائعة عندما يُطلب منه إجراء تنبؤات بناءً على بيانات غير واقعية استخدمت في تدريب الذكاء الاصطناعي.

تضمنت إحدى الحالات الشهيرة للتحيز غير المقصود قيام منظمة رائدة بتدريب نظام الذكاء الاصطناعي الخاص بها لتصنيف المتقدمين للوظائف، حيث كان هناك شعور بأن هذا سيجعل العملية أكثر كفاءة. ومع ذلك، فإن البيانات التدريبية المستخدمة كانت تتكون من عينات جاءت أصلاً من متقدمين ناجحين من الرجال. نتيجة لذلك، قام نظام الذكاء الاصطناعي بتقييم طلبات النساء بشكل سيء للغاية. وبعد أن بدا الأمر وكأن الذكاء الاصطناعي يميز ضد النساء، سحبت المنظمة الخدمة.

التحيز غير المقصود تسبب أيضاً في مشكلات عندما استُخدمت أداة فحص الذكاء الاصطناعي لتمييز أعراض سرطانات الجلد. أوضحت البيانات المتحيزة أن الأعراض لدى الأشخاص ذوي البشرة الداكنة لم يتم تحديدها بدقة. كان هذا بسبب تغذية بيانات التدريب بمزيد من صور الوجوه البيضاء.

كلا الحالتين نشأتا من مشكلات في تنوع البيانات التي أدت إلى التحيز ونتائج أسوأ لبعض فئات السكان. ومن نتائج هذا النوع من التمييز هو أن الثقة في أنظمة الذكاء الاصطناعي تتضرر.

المصطلحات الرئيسية

التحيز المؤسسي – يُقصد به التحيز المترتب على سياسات وإجراءات المنظمة المصممة للعمل بطريقة معينة تصب في مصلحة بعض الفئات الاجتماعية أو الأفراد أو تلحق الضرر بها.

التحيز الهيكلي – هو مفهوم أوسع نطاقاً يشير إلى الحفاظ النظامي على عدم المساواة بين مجموعات اجتماعية مختلفة أو أفراد.

التحيز الخوارزمي

يحدث التحيز الخوارزمي بسبب الجمع بين طبيعة بيانات التدريب ونموذج الذكاء الاصطناعي المستخدم. من المهم أن نتذكر أن بيانات التدريب والخوارزميات المستخدمة لإنشاء نماذج الذكاء الاصطناعي عادةً ما يختارها علماء البيانات (البشريين) ويجرون عليها التعديلات. لذلك من المألوف نقل تحيزاتهم اللاواعية، وقد يصعب جداً تحديد هذا النوع من التحيز وفك تشابكه.

التحيز النظامي

التحيز النظامي هو مصطلح واسع المعنى يصف الحالة التي تُصمم فيها عملية لدعم مجموعة معينة من النتائج، ونظراً لأن هذا الأمر يحدث غالباً داخل المؤسسات الكبيرة، ففي معظم الأحيان يُستخدم المصطلح إلى جانب التحيز المؤسسي أو الهيكلي.

وغالباً ما يحدث هذا النوع من التحيز بسبب التحيز المعرفي – الطريقة التي يطور بها الناس نظرة ذاتية (غير موضوعية) للعالم ويتخذون قرارات لا تبدو عقلانية (على سبيل المثال، الحكم على قدرة الناس بناءً على مظهرهم، بدلاً من الأمور التي يمكنهم القيام بها). ومن الممكن أن ينتقل هذا التحيز إلى نظام الذكاء الاصطناعي.

هل يمكن أن يكون التحيز جيداً في الذكاء الاصطناعي؟

هناك عدد من الحالات التي يتم فيها دمج التحيز في نظام الذكاء الاصطناعي.

- بعض أنظمة الذكاء الاصطناعي على نمط التوصية ستكون متعمدة التحيز لإعطاء وزن أكبر لعلاقات بيانات معينة. يتم ذلك للتأكد من أن العميل يوصى بالموسيقى أو مقاطع الفيديو أو المواد الغذائية التي قد يرغب فيها. التحيز المستخدم هو أن (الأشخاص الذين أحبوا X، أحبوا أيضاً Y).
- في السيارات ذاتية القيادة، سيتم تدريب الذكاء الاصطناعي (أي المتحيز) على توكي الحذر الشديد عند القيام بمناورات يحتمل أن تكون خطيرة (على سبيل المثال، تجاوز سيارة أخرى).
- يرى معظم الناس هذين المثالين بشكل إيجابي. كما هو الحال مع العديد من الأشياء في الحياة، فإن التحيز يعتمد على السياق.

ماذا يمكننا أن نقول باختصار عن التحيز؟

- في حين أن جميع أشكال التحيز ليست سيئة بطبيعتها، فمن المحتمل أن تكون لها عواقب سلبية على مشروع الذكاء الاصطناعي إذا لم يتم تحديدها بوضوح.
- إذا لم يتم معالجة التحيز، فإن هذا سيضر (تماماً) بثقة الجمهور في التكنولوجيا.
- من خلال الاعتراف بوجود التحيز وفهم كيفية ظهوره، يمكن لعلماء البيانات اتخاذ إجراءات لضمان عدم تأثير الذكاء الاصطناعي به.

الحد من تأثير التحيز في مجموعات البيانات

يوجد عدد من الطرق للحد عملياً من تأثير التحيز في مجموعات البيانات. يرد في ما يأتي خمس نصائح أساسية عليك مراعاتها.

- 1 احذر من التحيز في أخذ العينات.
- 2 تأكد من أن البيانات لها تمثيل متنوع، باستخدام مجموعة من المصادر المختلفة لمنع التمثيل الناقص لعناصر معينة من السكان.
- 3 نظف البيانات بعناية قبل استخدامها لاكتشاف أي تحيزات محتملة أو بيانات مفقودة أو انحرافات غير متوقعة.
- 4 استخدم الخوارزميات التي يمكنها اكتشاف التحيز والإبلاغ عنه.
- 5 تحقق من صحة القرارات وقيمتها باستخدام مقاييس العدالة - ابحث بنشاط عن أي تحيزات غير متوقعة تحدث (وقم بمواجهتها).

نصيحة مهمة أخرى هي التفاعل مع مجموعة متنوعة من أصحاب المصلحة خلال جميع مراحل مشروع الذكاء الاصطناعي - من البداية إلى النهاية. وجود وجهات نظر متعددة حول نتائج الذكاء الاصطناعي يمكن أن يكون مفيداً في تسليط الضوء على التحيز وتحديد القرارات السيئة المحتملة.

تحديد مصادر البيانات المناسبة والموثوقة والوصول إليها

عادةً ما يكون إنشاء مصدر بيانات مناسب لحلول الذكاء الاصطناعي مهمة مهندس البيانات، إذ سيقوم بالجمع الأولي للبيانات، وينظفها، ويخططها، ويوحدها – ربما من عدد من مصادر البيانات المتباينة (على سبيل المثال، استجابات واجهة برمجة التطبيقات، ملفات CSV، قواعد البيانات، وما إلى ذلك). يوجد عدد من الطرق المختلفة التي يمكن لمهندس البيانات من خلالها الحصول على مصادر موثوقة للبيانات. وتشمل الآتي:

- مقابلات أصحاب المصلحة – عادةً ما تكشف هذه المقابلات عن مصادر البيانات الرئيسية المستخدمة في عمليات سير العمل وممارسات الأعمال وما إلى ذلك.
- واجهات برمجة التطبيقات الموثوقة – غالبًا ما تكون واجهات برمجة التطبيقات المستندة إلى الويب والخدمات الأخرى مصحوبة بوثائق احترافية توضح مصادر البيانات المختلفة ونقاط النهاية للوصول إليها وتنسيق الطلبات والاستجابات اللازمة.
- مجموعات البيانات العامة/المفتوحة – تنشر العديد من المنظمات (بما في ذلك الوكالات الحكومية والهيئات الصناعية والمنظمات البحثية) مجموعات بيانات قابلة للتنزيل مجانًا.
- حوكمة البيانات – تقوم المؤسسات عادةً بإنشاء إجراءات حوكمة للامتثال لمتطلبات شهادة معالجة البيانات الدولية وأفضل الممارسات. غالبًا ما تحتوي هذه المعلومات على معلومات حول مصادر البيانات المفيدة.
- مستودعات البيانات وأنظمة الملفات – تخزن العديد من المنظمات بياناتها في مقراتها أو تستخدم مراكز بيانات ذات علامة بيبضاء أو خدمات تخزين سحابية. قد تكون بعض البيانات في شكل من أشكال التحكم في الإصدار. تصف الوثائق وسجلات الوصول عادةً مواقعها.
- قوائم جرد البيانات – الإجراء المثالي هو توثيق جميع مصادر البيانات، مع استخدام (عقود) البيانات الرسمية لتحديد المصادر والملكية وإمكانية الوصول والاستخدام المسموح به ومعلومات الاتصال وما إلى ذلك.

بالإضافة إلى ذلك، من الممكن استخدام أدوات (اكتشاف) البيانات الآلية لتحديد مصادر البيانات (الملفات وقواعد البيانات وواجهات برمجة التطبيقات وما إلى ذلك) التي تستخدمها التطبيقات الحالية. الوجهة المعتادة لأي مصدر بيانات جديد هي السحابة، ويرجع ذلك أساسًا إلى حجمها وتوافرها وقابليتها للتطوير.

بحث

الموقع الرسمي للحكومة الإلكترونية الأردنية متاح عبر الإنترنت ويستضيف منصة البيانات الحكومية المفتوحة الخاصة بهم.

تحتوي هذا البوابة المصنفة بشكل جيد والقابلة للبحث على مجموعات بيانات مفتوحة عالية الجودة متاحة بتنسيقات قابلة للقراءة آليًا (على سبيل المثال، XLSX، CSV، وما إلى ذلك) يمكن تنزيلها ومراجعتها.

يمكن استخدام هذا المورد العام لتحديد مصادر البيانات المناسبة للمعلومات الآتية:

- استهلاك الطاقة
- عدد حالات الحروق في مستشفيات وزارة الصحة لعام 2021
- عدد القضايا المستلمة والمسجلة بحسب الموقع ونوع المحكمة لعام 2023
- المواليد الأردنيون لعام 2022 بحسب الولادة والجنس.

ابحث وراجع مصدر بيانات مفتوح آخر ذي صلة يمكنك العثور عليه وتحديد مدى ملاءمته وجودته وتنسيقات البيانات المتاحة. اسأل نفسك: هل البيانات موثوقة جيدًا؟

إعداد البيانات الجاهزة للاستخدام

من المهم إعداد البيانات الجاهزة للاستخدام. قد يقوم مهندس البيانات في البداية بتنفيذ المهام القياسية الآتية على البيانات الأولية.

تنظيف البيانات

- هناك العديد من التحديات المختلفة التي ينطوي عليها تنظيف البيانات كجزء من عملية التحضير.
- تتمثل المشكلة الأولى في إزالة أي إدخال غير قابلة للتطبيق/غير متوفرة قد تكون موجودة في مجموعة البيانات. يجب أيضًا إزالة الفراغات لأنها لا تضيف أي قيمة بيانات وقد تمنع المعالجة.
 - في بعض الأحيان قد تحتاج حقول البيانات إلى التقسيم أو الدمج. على سبيل المثال، قد يلزم تقسيم حقل واحد إلى حقلين (على سبيل المثال، الاسم إلى الاسم الأول واسم العائلة)، أو قد يلزم ضم حقلين في حقل واحد (على سبيل المثال، الاسم الأول واسم العائلة في الاسم). سيعتمد هذا على طبيعة المعالجة المطلوبة.
 - من الشائع أيضًا العثور على تناقضات في البيانات. خطأ شائع يحدث عند استخدام تنسيقات حالة مختلفة لنفس عنصر البيانات (على سبيل المثال، مزيج من الأحرف الصغيرة، الأحرف الكبيرة في اللغة الإنجليزية، حالة العنوان، وما إلى ذلك). يجب أن يكون هذا موحدًا.
 - يمكن أن يكون تنسيق البيانات الموجهة (على سبيل المثال، تاريخ يحتوي على جوانب اليوم والشهر والسنة) مشكلة أيضًا إذا كان تنسيقها يختلف (على سبيل المثال، بين الأوروبي (يوم/شهر/سنة) والأمريكي (شهر/يوم/سنة)).
 - قد تمثل القيم المفقودة مشكلة أيضًا (على سبيل المثال، في سلسلة زمنية حيث قد تكون نقاط البيانات غائبة أحيانًا). عادةً ما يتم تعيينها أو إزالتها أو وضع علامة عليها.
 - إخفاء هوية البيانات أمر بالغ الأهمية أيضًا، لا سيما عند مراعاة تشريعات حماية البيانات المحلية ومخاوف حوكمة البيانات بشأن المعالجة الأخلاقية للبيانات الشخصية. قد يعني هذا إزالة معلومات التعريف الشخصية غير الضرورية من مجموعة البيانات.

إزالة التكرار

- في بعض الأحيان يمكن أن تحتوي مجموعات البيانات على بيانات مكررة. في أفضل السيناريوهات، قد يعني هذا فقط إزالة بعض الحقول التي ليس لها دور في عملية التعلم الآلي لحل الذكاء الاصطناعي، وهذه الحقول ببساطة ليست مطلوبة.
- في بعض الأحيان تكون المشكلة أكثر تعقيدًا، وهذه هي الحالة عندما لا تكون الحقول زائدة عن الحاجة فحسب، بل تتسبب في تحديات لعملية التعلم الآلي.
- يمكن أيضًا إزالة الأعمدة المحسوبة (أي تلك التي يتم فيها توليد القيم من أخرى) بأمان حيث يمكن إعادة توليدها بسهولة عند الحاجة.

دمج/تجميع مجموعات البيانات

- غالبًا ما تقتصر مجموعات البيانات إلى التنسيق الصحيح وقد تحتاج إلى الدمج. على سبيل المثال، قد يكون من الضروري دمج جداول قاعدة بيانات متعددة معًا. قد يكون من الضروري أيضًا التحويل من تنسيقات CSV/JSON/XML المختلفة إلى تنسيق واحد متفق عليه يسهل معالجته.

التحقق من صلاحية البيانات (معقوليتها) وليس دقتها

- غالبًا ما تحتوي البيانات على قيم لا معنى لها. على سبيل المثال، عمر إنسان بالسنوات يزيد عن 120 عامًا، أو طلب العميل كمية بقيمة سالبة من العناصر المطلوبة.
- يجب تحديد هذه الأنواع من القيم والتحقق منها وإزالتها إذا تبين أنها غير صالحة. بالطبع، قد يكون عمر الشخص بالشهور وليس بالسنوات، ما يعني أن البيانات صالحة، ولكن وصف الحقل غير صحيح.
- سيقوم عالم البيانات بمزيد من العمل على مجموعة البيانات عندما يبدأ تفكير في اختيار الميزة. يلزم توافر مستوى من المعرفة بالمجال (البيانات وطبيعة المشكلة) للقيام بذلك. يمكن أن يتضمن العمل التحقق من التحيز، وما إلى ذلك.

وقفه للتفكير



عندما تستخدم البيانات، هل تفترض فقط أن البيانات ستكون صحيحة ومناسبة للغرض وجاهزة للاستخدام، أم تقضي وقتًا في محاولة العثور على بيانات مناسبة؟ كيف تحققت من مشكلات مثل التحيز؟

تلميح

هل قمت بفحص البيانات النموذجية المتاحة على الإنترنت؟ إذا لم يكن الأمر كذلك، فابحث عن مصادر البيانات المتاحة في بلدك.

توسيع الأفق

فكر في كيفية تشويه البيانات المتحيزة لتتأجك.

مثال عملي

يوضح المثال الآتي إنشاء مجموعة من البيانات التركيبية المخترقة عمدًا وإعدادها لاحقًا للاستخدام في مشروع ذكاء اصطناعي بسيط. ميا تدوير مقهى. كونها شركة جديدة نسبيًا وغير مختبرة، قررت تثبيت أجهزة إنترنت الأشياء لقياس وتسجيل حركة العملاء في متجرها. سيسمح لها ذلك بنمذجة النزوات والانخفاضات في الطلب طوال كل يوم عمل. تحليل هذه البيانات باستخدام الذكاء الاصطناعي سيمكنها رؤية تأمل أن تحقق هدفين: تحسين خدمة العملاء بشكل عام وزيادة المبيعات.

يستخدم كود بايثون في الشكل 21.20 لتوليد مجموعة البيانات الاصطناعية - يتم نمذجتها بناءً على القواعد الآتية.

- متجر ميا مفتوح من الاثنين إلى السبت (سنة أيام).
 - ساعات العمل في متجر ميا هي من 9 صباحًا إلى 5 مساءً.
 - يتم تلخيص عدد الزوار لكل مدة زمنية بالساعة.
 - يتم تسجيل البيانات لمدة عشرة أسابيع.
 - الفترات من الساعة 9 صباحًا و 1 مساءً عادةً ما تكون الأكثر ازدحامًا.
- أحيانًا تفشل الأجهزة ولا تسجل البيانات بدقة. في هذه الحالات، تظهر القيمة 'n/a' (غير قابلة للتطبيق). يحدث هذا بنسبة قدرها 10 في المائة. يجب أن تكون جميع البيانات أعدادًا صحيحة.

```
# Import required libraries
import pandas as pd
import numpy as np

# Create a time series of customer footfall for 6 working days a week (Monday
to Saturday)
# Hourly slots from 9AM to 5PM (inclusive)
# For 10 weeks
dates = pd.date_range(start='2023-05-01', periods=10*6, freq='B')
hours = pd.timedelta_range(start='9 hours', periods=9, freq='H')
index = pd.MultiIndex.from_product([dates, hours], names=['date', 'time'])
data = pd.DataFrame(index=index, columns=['footfall'])

# Generate random customer footfall data with given conditions
np.random.seed(42)
for date in data.index.get_level_values(0).unique():
    day_hourly_data = []
    # For each hour interval every day...
    for hour in range(9):
        # Normal footfall, with peaks at opening time and lunchtime
        if hour == 0 or hour == 4: # Peaks at 9AM and 1PM
```

الشكل 21.20 (متابعة)

متابعة

```

        footfall = np.random.randint(25, 45)
    else:
        footfall = np.random.randint(10, 30)

    # Make some timeslots n/a randomly (10% chance of IoT failure)
    if np.random.rand() < 0.1:
        footfall = 'n/a'

    day_hourly_data.append(footfall)

data.loc[(date,), 'footfall'] = day_hourly_data

# Simplify and display the created data frame
data.reset_index(inplace=True)
data['time'] = data['time'].apply(lambda t: (t +
pd.Timestamp("today").normalize()).time())

# Print the first 18 rows to see a sample of the first 2 days' data
print(data.head(18))

```

	date	time	footfall
0	2023-05-01	09:00:00	31
1	2023-05-01	10:00:00	24
2	2023-05-01	11:00:00	16
3	2023-05-01	12:00:00	20
4	2023-05-01	13:00:00	28
5	2023-05-01	14:00:00	n/a
6	2023-05-01	15:00:00	11
7	2023-05-01	16:00:00	n/a
8	2023-05-01	17:00:00	10
9	2023-05-02	09:00:00	n/a
10	2023-05-02	10:00:00	n/a
11	2023-05-02	11:00:00	25
12	2023-05-02	12:00:00	24
13	2023-05-02	13:00:00	36
14	2023-05-02	14:00:00	12
15	2023-05-02	15:00:00	16
16	2023-05-02	16:00:00	n/a
17	2023-05-02	17:00:00	23

الشكل 21.20 (متابعة) كود بايثون والمخرجات التي تظهر النقاط خطوات الأقدام باستخدام بيانات اصطناعية.

نظرًا لطبيعتها، تُعرف هذه النقاط البياناتية مجتمعةً باسم "سلسلة زمنية". لقد حُزّنت في كائن Pandas DataFrame (هيكل يشبه الجدول مع صفوف وأعمدة معنونة) للمعالجة.

متابعة

لسوء الحظ، فإن وجود قيم 'n/a' في مجموعة البيانات يكسر استمرارية وموثوقية هذه المجموعة الزمنية من البيانات. لذلك، يجب تنظيف البيانات، وإلا فقد تواجه خوارزميات التعلم الآلي صعوبة في معالجتها. هناك بعض الخيارات الأساسية.

- 1 **استبعادها** – إزالة نقاط البيانات التي تحتوي على قيم 'n/a'. لا بأس بذلك إذا كانت الكمية صغيرة نسبيًا ولم يؤثر محوها تأثيرًا كبيرًا في السلامة العامة لمجموعة البيانات.
 - 2 **الاستبدال** – استبدال القيم المفقودة 'n/a' بقيم بديلة.
 - يمكنك استخدام حسابات المتوسط أو الوسيط أو الوضع.
 - يمكنك استخدام آخر قيمة متاحة بخلاف 'n/a' - تُعرف هذه القيمة باسم LOCF (تم ترحيل الملاحظة الأخيرة).
 - أو يمكنك استخدام الملاحظة التالية المتاحة (أي NOCB (تم نقل الملاحظة التالية للخلف)).
 - 3 **استخدم تقنيات مثل ARIMA** (الموضحة أدناه) للتنبؤ بالقيم المفقودة من خلال مراقبة الاتجاهات، الموسمية، وما إلى ذلك.
 - 4 **ضع علامة عليها** حتى تتمكن خوارزمية التعلم الآلي من التعويض عند إنشاء النموذج.
- تم تحديد الخيار 2 وتم إنشاء برنامج نصي قصير لتنظيف البيانات. سيؤدي ذلك إلى أخذ إطار البيانات التركيبي الحالي وإسناد القيم المفقودة باستخدام استبدال LOCF. الكود الإضافي موجود في الشكل 21.21.

```
# Convert 'n/a' to NaN (Not a Number) type to allow use of pandas functions
data['footfall'] = pd.to_numeric(data['footfall'], errors='coerce')
```

```
# Fill NaN missing values using Last Observation Carried Forward (LOCF)
data['footfall'].fillna(method='ffill', inplace=True)
```

```
# Display the cleaned and processed DataFrame
print(data.head(18))
```

```
0 2023-05-01 09:00:00    31.0
1 2023-05-01 10:00:00    24.0
2 2023-05-01 11:00:00    16.0
3 2023-05-01 12:00:00    20.0
4 2023-05-01 13:00:00    28.0
5 2023-05-01 14:00:00    28.0
6 2023-05-01 15:00:00    11.0
7 2023-05-01 16:00:00    11.0
8 2023-05-01 17:00:00    10.0
9 2023-05-02 09:00:00    10.0
10 2023-05-02 10:00:00    10.0
11 2023-05-02 11:00:00    25.0
12 2023-05-02 12:00:00    24.0
13 2023-05-02 13:00:00    36.0
14 2023-05-02 14:00:00    12.0
15 2023-05-02 15:00:00    16.0
16 2023-05-02 16:00:00    16.0
17 2023-05-02 17:00:00    23.0
```

الشكل 21.21 كود بايثون الذي يحل أي قيم غير متوفرة باستخدام LOCF.

متابعة

هل تم اتخاذ القرار الصحيح؟ تم تكرار ذروة وقت الغداء (في الساعة 1 مساءً) البالغة 28 باستخدام تقنية LOCF على نقطة البيانات في الساعة 2 مساءً. من المحتمل أن يؤدي ذلك إلى تحريف البيانات إلى حد ما - لاحظ أن الزيادة نفسها لا تحدث في اليوم التالي، على سبيل المثال. لذلك، يجب إعطاء المزيد من الاعتبار قبل الاستقرار على تقنية تنظيف البيانات المناسبة للاستخدام. على سبيل المثال، ماذا يحدث إذا استُخدم المتوسط الحسابي لليوم بدلاً من ذلك؟

تجربة بعض التقنيات المختلفة وملاحظة التأثيرات المحتملة هو النهج الموصى به عادةً. هذا ليس علمًا دقيقًا وسيخبرك معظم علماء البيانات أن هذه الأنواع من القرارات هي المكان الذي يصبح فيه (فن) هذا التخصص مهمًا.

اختيار واستخدام هياكل وتنسيقات البيانات المناسبة

- يركز هذا الموضوع على التحدي المتمثل في العمل مع البيانات المخزنة في هياكل وتنسيقات مختلفة، ما يعود بتأثير كبير في كيفية التقاطها وتخزينها ومعالجتها.
- يمكن أن تأتي البيانات المستخدمة في المشاريع المتعلقة بالذكاء الاصطناعي في عدد من التنسيقات المختلفة ومن مصادر مختلفة. تشمل مصادر البيانات الشائعة:
- واجهات الأجهزة على أجهزة إنترنت الأشياء مثل أجهزة الاستشعار والكاميرات والميكروفونات وما إلى ذلك.
 - ملفات البيانات
 - ملفات نصية غير منظمة
 - ملفات نصية منظمة، مثل القيم المفصولة بفواصل (CSV) ولغة الترميز القابلة للتوسعة (XML) وتكوين كائن (JSON) (JavaScript)
 - ملفات ثنائية مثل الصور والصوت والفيديو
 - قواعد البيانات
 - قواعد بيانات SQL (لغة الاستعلام المنظمة)
 - قواعد بيانات NoSQL (وليس SQL فقط)
 - واجهات برمجة التطبيقات المستندة إلى الويب، والتي تعرض (عادةً) ملفات JSON أو XML.

تنسيقات البيانات

لفهم تنسيقات البيانات، راعي بعض نماذج بيانات المتعلم:

- معرف المتعلم: 1023، الاسم الأول أراف، الاسم الأخير: باتيل، العمر: 20
- معرف المتعلم: 2045، الاسم الأول مي، الاسم الأخير: وونغ، العمر: 22
- معرف المتعلم: 3567، الاسم الأول كارلوس، الاسم الأخير: هيرنانديز، العمر: 19
- معرف المتعلم: 4789، الاسم الأول فاطمة، الاسم الأخير: الفارسي، العمر: 21

تختلف تنسيقات ملفات CSV في الجودة. وغالبًا ما تكون البيانات مفقودة وقد لا يكون صف العنوان موجودًا. تحتوي بايثون على وحدة CSV محددة يمكنها تحميل هذه البيانات مباشرة في قاموس لسهولة المعالجة.	learner_id,first_name,last_name,age 1023,"Aarav","Patel",20 2045,"Mei","Wong",22 3567,"Carlos","Hernandez",19 4789,"Fatima","Al-Farsi",21	CSV
ملفات JSON تشبه في الأساس قوائم القواميس في بايثون. هذا الأمر يجعل من السهل جدًا استيرادها واستخدامها مع وحدة JSON في بايثون.	[{ "learner_id": 1023, "first_name": "Aarav", "last_name": "Patel", "age": 20 },]	JSON

الشكل 21.22 (متابعة)

(متابعة)	<pre>{ "learner_id": 2045, "first_name": "Mei", "last_name": "Wong", "age": 22 }, { "learner_id": 3567, "first_name": "Carlos", "last_name": "Hernandez", "age": 19 }, { "learner_id": 4789, "first_name": "Fatima", "last_name": "Al-Farsi", "age": 21 }]</pre>	
<p>غالبًا ما تمثل ملفات XML تحديدًا أكثر صعوبة. بينما يمكن تحويلها إلى قوائم من القواميس، إلا أنها ليست دائمًا النوع الأسهل من المعالجة لتحقيقه، حتى في بايثون. من الجدير بالذكر أيضًا أن XML لديها ثغرة أمنية معروفة: الكيانات الخارجية.</p>	<pre><learners> <learner> <learner_id>1023</learner_id> <first_name>Aarav</first_name> <last_name>Patel</last_name> <age>20</age> </learner> <learner> <learner_id>2045</learner_id> <first_name>Mei</first_name> <last_name>Wong</last_name> <age>22</age> </learner> <learner> <learner_id>3567</learner_id> <first_name>Carlos</first_name> <last_name>Hernandez</last_name> <age>19</age> </learner> <learner> <learner_id>4789</learner_id> <first_name>Fatima</first_name> <last_name>Al-Farsi</last_name> <age>21</age> </learner> </learners></pre>	XML

الشكل 21.22 (متابعة) مقارنة مرئية بين أكثر ثلاثة تنسيقات نصية منظمة شيوعًا.

ثم ألق نظرة على السجلات كما هو موضح في الشكل 21.22. يتم تمثيلها في تنسيقات النص المنظم الثلاثة الأكثر شيوعًا.

في هذا المثال، أُجريت العديد من الخيارات، على النحو الآتي.

يتم تخزين حقل learner_id كعدد صحيح (رقم صحيح). قد لا يكون هذا مناسبًا لأن العدد الصحيح سيفقد أي أصفار بادئة. من غير المحتمل أيضًا أن ترغب في إجراء أي عملية حسابية عليها. وبالتالي، قد يكون تخزينها كسلسلة أكثر ملاءمة.

بالنسبة لملف CSV، تم اقتباس حقلَي first_name و last_name. تعتبر هذه ممارسة جيدة لأنها تمكن أي جزء من اسم المتعلم من احتواء مسافة، إذا لزم الأمر. سيساعد ذلك على المعالجة المستقبلية.

يتم تخزين حقل العمر أيضًا كعدد صحيح. هذا مفيد وربما يكون الاختيار الصحيح لأنه من المحتمل أن يكون هناك حاجة إلى نوع من المقارنة الحسابية أو الرقمية كجزء من أي عملية تعلم آلي.

المزايا والعيوب والاستخدامات الشائعة لتنسيقات البيانات

كل صيغة بيانات لها مزاياها وعيوبها واستخداماتها الشائعة، ويتم توضيحها في الجدول 21.7.

الجدول 21.7 مقارنة ميزات CSV وJSON وXML

XML	JSON	CSV
<ul style="list-style-type: none"> شكل من أشكال لغة الترميز الممتدة على العلامات، مثل HTML (لغة ترميز النص التشعبي) يُستخدم للبيانات المنظمة مصمم ليكون من السهل على البشر والآلات قراءته أكثر تفصيلاً من ملف CSV يدعم أدوات التحقق من صحة البيانات التي تفضل سلامة البيانات يمكنه استخدام العلامات والسمات المخصصة العديد من الاستخدامات بما في ذلك مع أجهزة برمجة التطبيقات المستندة إلى الويب، وتخزين البيانات وملفات التكوين 	<ul style="list-style-type: none"> على الرغم من الاسم، إلا أنه تنسيق بيانات مستقل عن لغة البرمجة من السهل على البشر قراءته وكتابته من السهل على أجهزة الحاسب إنشاؤه وتحليله يدعم الهياكل البيانية المعقدة، مثل القوائم والقواميس المترابطة المفتاح-القيمة أكثر تفصيلاً من ملف CSV غالبًا ما يُستخدم للتفاعل مع واجهات برمجة التطبيقات المعتمدة على الويب، على سبيل المثال، للطلبات والاستجابات، وتخزين البيانات في قواعد بيانات NoSQL، وما إلى ذلك. 	<ul style="list-style-type: none"> الشكل الأقدم والأكثر شيوعًا لتبادل البيانات كل سطر يمثل سجلًا واحدًا يتم فصل الحقول بفواصل يحتوي صف العنوان الاختياري على أسماء الحقول بسيط وخفيف الوزن دعم ضعيف للهياكل البيانية الأكثر تعقيدًا، مثل القوائم، وما إلى ذلك. ممتاز لاستيراد وتصدير البيانات الجدولية (مثل، إلى ومن جداول البيانات، جداول قواعد البيانات).

ستجد أن العديد من أدوات الذكاء الاصطناعي متوافقة مع واحد أو أكثر من تنسيقات البيانات الثلاثة هذه. وبالتالي من المفيد للغاية فهم الفروق بينهما، ومعرفة كيفية إنشائها باستخدام التطبيقات المناسبة (مثل برامج تحرير النصوص)، والقدرة على قراءتها وتفسيرها بحسب الضرورة.

بحث

على الرغم من أن CSV وJSON وXML هي تنسيقات شائعة لتخزين وتبادل البيانات، إلا أنها تأتي أيضًا مع مجموعة من المشكلات الشائعة التي يمكن أن تُعقّد التعامل مع البيانات ومعالجتها وتحليلها. ابحث في كل تنسيق وأنشئ قائمة بها ثلاث مشكلات شائعة لكل منها والكيفية التي يمكن بها أن تؤثر هذه المشكلات في جمع البيانات لمشروع الذكاء الاصطناعي. تحديد الاستخدامات المحتملة لكل من هذه الصيغ البيانية الشائعة (بخلاف تلك المذكورة أعلاه).

المهارات

مهارات التواصل الشخصي: أخلاقيات العمل/الضمير:

- المسؤولية
- الأخلاقيات
- النزاهة

الاعتبارات التشريعية والأخلاقية والأمنية

كما رأينا، تعد البيانات مطلبًا أساسيًا لأي مشروع للذكاء الاصطناعي. ومع ذلك، عند جمع البيانات وتجهيزها واستخدامها لمشروع الذكاء الاصطناعي، يوجد العديد من المخاوف القانونية والأخلاقية والأمنية التي يجب وضعها في الاعتبار والاستجابة لها.

بالإضافة إلى التفكير في التحديات التقنية لمشروع الذكاء الاصطناعي، من المهم بالقدر نفسه النظر في هذه القضايا. هذا الأمر مهم، لأن من الممكن أن يكون له تأثير كبير في نجاح مشروعك وسهولة استخدامه.

القضايا التشريعية المحلية الحالية وذات الصلة

يجب عليك الالتزام بقوانين حماية البيانات المحلية والدولية - على سبيل المثال، اللائحة العامة لحماية البيانات (GDPR) في الاتحاد الأوروبي (EU). لدى معظم البلدان خارج الاتحاد الأوروبي تشريعات مماثلة تعكس المبادئ الأساسية لللائحة العامة لحماية البيانات في ما يتعلق بجمع البيانات والاحتفاظ بها ومعالجتها. المبادئ هي:

- المشروعية القانونية، والعدل، والشفافية
- تحديد الغرض
- تصغير البيانات
- دقة البيانات
- النزاهة والسرية.

مشكلات أخلاقية

تشمل القضايا الأخلاقية الرئيسية حقوق ومسؤوليات الأفراد والمنظمات، ومن المهم جدًا ضمان إخفاء الهوية الفردية مع الحفاظ على جودة البيانات وتعزيز الإنصاف والمساواة.

يعد استخدام معلومات التعريف الشخصية (PII) في مشاريع الذكاء الاصطناعي مصدر قلق خاص بسبب مخاطر الخصوصية والأمان المحتملة التي يسببها. كما يوحي اسمها، فإن معلومات تحديد الهوية الشخصية هي أي بيانات يمكن أن تحدد هوية الفرد. وقد تتضمن الأسماء وأرقام الضمان الاجتماعي والعناوين وعناوين البريد الإلكتروني. هناك العديد من الخطوات التي يمكن اتخاذها لاستخدام معلومات تحديد الهوية الشخصية بشكل أخلاقي. وتشمل الآتي:

- طلب موافقة مستنيرة من الأفراد قبل جمع معلومات تحديد الهوية الشخصية الخاصة بهم أو معالجتها أو استخدامها
- جمع معلومات تحديد الهوية الشخصية الضرورية للغاية فقط للغرض المحدد لمشروع الذكاء الاصطناعي المقصود
- إخفاء الهوية أو استخدام أسماء مستعارة (إلغاء تعريف) لمعلومات تحديد الهوية الشخصية حيثما أمكن ذلك.

هناك عدد من الخطوات التي يمكن اتخاذها لتعزيز العدل والإنصاف. وتشمل الآتي:

- تقليل التحيز في اختيار البيانات
- تقليل خطر التمييز (على سبيل المثال، في مجالات التعليم والعدالة)
- تعزيز الشفافية في مشاريع الذكاء الاصطناعي من خلال ضمان وضوح العمليات وقابليتها للتفسير.

ضمان أمن وخصوصية البيانات الشخصية والحساسية وحماية سلامة البيانات من التلف أو الفساد

كما رأينا، تعد أنظمة الذكاء الاصطناعي هدفًا واضحًا للهجمات الإلكترونية أو الجهات الفاعلة الخبيثة. دوافع هؤلاء الأشخاص تختلف، ولكن أهدافهم المشتركة هي:

- الوصول إلى البيانات الشخصية أو الحساسة
- تقليل الدقة أو تعديل نتائج خدمة الذكاء الاصطناعي.

يتطلب كلاهما تنفيذ العمليات واتخاذ الاحتياطات الآتية:

- النسخ الاحتياطي المنتظم للبيانات لمنع فقدان البيانات أو تلفها بسبب الأضرار العرضية أو الخبيثة
- تدابير الأمن السيبراني القوية لمنع انتهاكات البيانات والتسلل (إزالة البيانات بشكل غير قانوني دون إذن) والفساد (على سبيل المثال، جدران الحماية وضوابط الوصول المناسبة وتشفير البيانات في حالة السكون (عدم النقل) وفي أثناء النقل وما إلى ذلك)
- إدارة جودة البيانات لمراقبة صحة البيانات المستخدمة في أنظمة الذكاء الاصطناعي والحفاظ عليها
- التدقيق الأمني المنتظم
- تطوير بروتوكولات قوية للتعامل مع انتهاكات البيانات والإبلاغ عنها إلى السلطات المختصة (والأفراد المتضررين).

بالإضافة إلى ذلك، يتمتع علماء البيانات بإمكانية الوصول إلى مجموعات أدوات توفر الدعم لتعلم الآلة العدائي (AML). يمكن أن تساعد مجموعات الأدوات هذه على تحديد محاولات تسميم بيانات التدريب الخاصة بنموذج التعلم الآلي، وما إلى ذلك.

اختيار وإعداد مجموعات بيانات التدريب والتحقق والاختبار لمشروع الذكاء الاصطناعي

يعتمد نجاح مشروع الذكاء الاصطناعي بشكل أساسي على جودة البيانات المستخدمة لتدريب نموذج (نماذج) التعلم الآلي الخاص به. لا يمكن المبالغة في تقدير مدى أهمية اختيار وإعداد مجموعات البيانات المستخدمة في التدريب والتحقق من اختبار حل الذكاء الاصطناعي في الواقع. البيانات التي لم يتم اختيارها وإعدادها بشكل مناسب لن تكون قابلة للاستخدام بشكل أساسي، وإذا تم إدخال بيانات غير قابلة للاستخدام أو ذات جودة رديئة في النموذج، فلن يتم الحصول إلا على نتائج غير قابلة للاستخدام أو ذات جودة رديئة.

إذا كانت جودة بيانات التدريب جيدة (راجع المتقابلات الخمسة) فستعكس مدخلات العالم الحقيقي التي سيراهها النظام عندما يكون قيد الاستخدام. ونتيجة لذلك، ستكون عملية التعلم فعالة وستساعد النظام على أن يكون أكثر دقة. على العكس من ذلك، يمكن للبيانات ذات الجودة الرديئة أن يكون لها العديد من الآثار السلبية على نجاح مشروع الذكاء الاصطناعي. لذلك من المهم جداً التركيز على عمليات البيانات الأولية الآتية.

اختيار البيانات

يجب عليك اختيار البيانات بحيث تكون ممثلة وتقل التحيز. يجب أيضاً احترام خصوصية الأفراد ومعالجة جميع الاعتبارات الأخلاقية ذات الصلة (لا سيما في ما يتعلق بالمعلومات الحساسة أو الشخصية).

إعداد البيانات

يجب عليك تحديد وتصحيح أي أخطاء في البيانات (على سبيل المثال، القيم المفقودة أو غير المتسقة أو التكرارات أو القيم المتطرفة الواضحة). بالإضافة إلى ذلك، يجب عليك تحويل البيانات الأساسية إلى ميزات تمثل المشكلة الأساسية بشكل أفضل. قد يتضمن ذلك تطبيع البيانات أو تحجيمها أو دمجها.

التحقق من صحة البيانات

يجب عليك التأكد من صحة أنواع البيانات، وأن البيانات تقع ضمن نطاق معقول وأنها موزعة بشكل صحيح. على سبيل المثال، إذا كنت تجمع آراء هيئة طلابية، فتأكد من أن العينات مأخوذة من جميع الفئات العمرية وليس من الطلاب الأكبر سناً فحسب.

المهارات

المهارات المعرفية: العمليات والإستراتيجيات المعرفية:

- اتخاذ القرار

مهارات التواصل الشخصي: أخلاقيات العمل/الضمير:

- الإنتاجية

موضوعات ذات صلة

للحصول على مزيد من المعلومات حول خطوات إنشاء البيانات وتنظيفها، راجع الجزء الخاص بجمع البيانات وتجهيزها للمشروع.

إنشاء مجموعات بيانات اختبار

يجب عليك التأكد من أن بيانات التدريب والاختبار معزولة لمنع تلوثها. يجب عليك استخدام تقنيات التقسيم للحفاظ على النسبة نفسها من أمثلة الفئات المختلفة للبيانات في كلا المجموعتين. على سبيل المثال، استمرراً لمثال "الطالب" السابق، ستضمن أن بيانات التدريب والاختبار تستخدم النسب نفسها لكل فئة عمرية. قبل كل شيء، أثبت أن البيانات واقعية وتمثل نوع البيانات التي سيواجهها النظام في (العالم الحقيقي).

تطبيق النظرية

معظم الإخفاقات غير المقصودة في مشاريع الذكاء الاصطناعي ناتجة عن الاختيار السيئ لبيانات التدريب والاختبار. يوجد العديد من الأسباب لمثل هذه المشكلات، لكن يمكن تجنبها في الغالب من خلال إجراء فحوصات منطقية في المجالات الأربعة الرئيسية الموضحة.

دراسة حالة



تأخر رحلات الطيران

تم توفير مجموعة بيانات أولية تحتوي على معلومات مفصلة عن تأخيرات الرحلات الجوية الأمريكية بين عامي 2018 و2022. للعثور على مجموعة البيانات هذه، يجب عليك استخدام محرك بحث والبحث باستخدام مجموعة المصطلحات الآتية:

تنزيل ملف kaggle للرحلات الجوية المجمعة 2022 بصيغة csv من robikscube

انقر على نتيجة البحث 'توقع حالة الرحلة'.

استفسرت سلطات إدارة المطار عما إذا كان من الممكن تحليل البيانات المقدمة وتدريب نموذج الذكاء الاصطناعي بنجاح بحيث يمكن التنبؤ بحالات التأخير المحتملة في المستقبل.

ملحوظة: البيانات نفسها ليست نظيفة تمامًا وهناك بعض المشكلات الواضحة التي يجب تحديدها وإصلاحها. بمجرد أن تجمع وتفحص البيانات، انتقل إلى التمرين التقييمي 21.2 وقم بإعداد (تنظيف وتكرير) مجموعة البيانات، ثم قم بتقييم مدى ملاءمتها للإجابة على السؤال المطروح.

لا تحاول تطوير حل الذكاء الاصطناعي الكامل حتى الآن.



B.P3, B.P4, B.M2, B.D2

تمرين تقييمي 21.2

أنت تعمل لصالح مدير مشروع طلب منك التخطيط والإعداد لإنشاء حل ذكاء اصطناعي لتلبية احتياجات العميل المحددة. خطوتك الأولى هي تحديد أهداف مشروع الذكاء الاصطناعي بوضوح وجمع البيانات المناسبة لتدريب النموذج.

هذه فرصتك لإظهار مجموعة من المهارات الشخصية التي طورتها خلال هذه الوحدة مثل التحليل وحل المشكلات واتخاذ القرار والاستدلال/الجدال والمسؤولية والإبداع. يجب أن تفكر مليًا في كيفية التعامل مع كل مهمة.

في نهاية المهمة، يمكنك أيضًا مراجعة تطبيقك لهذه المهارات، بالإضافة إلى تقييم فعالية حل الذكاء الاصطناعي.

يرجى قراءة دراسة الحالة المرتبطة، 'تأخيرات المطار'، لفهم المهمة المحددة.

ستحتاج إلى:

- تحديد أهداف مشروع الذكاء الاصطناعي.
- جمع مجموعات البيانات المناسبة وإعدادها لحل الذكاء الاصطناعي
- مراجعة وتحسين مجموعات البيانات لتحسين جودة حل الذكاء الاصطناعي
- تقييم فعالية حل الذكاء الاصطناعي

التخطيط

- ضع خطة للمهمة. ضع قائمة بكل ما ستحتاج إلى القيام به وضع جدول زمني لموعد إنجاز العمل.

التنفيذ

- حدد مصادر البيانات الموثوقة (والقانونية) التي يمكن استخدامها لحل الذكاء الاصطناعي الخاص بك.
- قم بإعداد البيانات من خلال البحث عن القيم المفقودة أو المكررة أو ذات التنسيق غير المتسق وأي قيم شاذة واضحة.
- راجع مجموعات البيانات المعدة وحسبها بحسب الضرورة – الاستعانة بزميل أو معلم لمراجعة البيانات سيعطيك حكمًا مستقلًا بشأن مدى ملاءمتها للمهمة.

المراجعة

- قم بإنشاء دليل قصير يشرح كيف قمت بجمع وإعداد بياناتك لمشروع الذكاء الاصطناعي الخاص بك.

لا تحاول تطوير حل الذكاء الاصطناعي الكامل حتى الآن.



تطوير حل الذكاء الاصطناعي لتلبية احتياجات محددة

المصطلحات الرئيسية

خوارزمية التجميع – تستخدم لتحديد ما إذا كان هناك أي مجموعات في مجموعة البيانات.

خوارزمية تقليل الأبعاد – تُستخدم لتقليل عدد الميزات في مجموعة البيانات دون المساس بخصائص البيانات الأصلية.

معاملات الضبط – قيم إضافية ليست جزءًا من مجموعة البيانات نفسها لكن يمكن تكوينها وتغيير طريقة عمل عملية التعلم الآلي.

بعد التخطيط والإعداد لإنشاء حل للذكاء الاصطناعي، يقدم لك هذا الموضوع الإجراءات اللازمة للمضي قدمًا بالخطوة وتطوير الحل رسميًا. في هذا الجزء، ستستكشف مجموعة من الخوارزميات المكتوبة بلغة بايثون.

ج1 نماذج الذكاء الاصطناعي

نقطة البداية في النقاش حول تطوير حلول الذكاء الاصطناعي هي مفهوم نموذج الذكاء الاصطناعي. يتم إنشاء نموذج الذكاء الاصطناعي من خلال عملية تسمى التعلم الآلي (ML). هناك طريقتان مختلفتان: التعلم الآلي تحت الإشراف والتعلم الآلي دون إشراف.

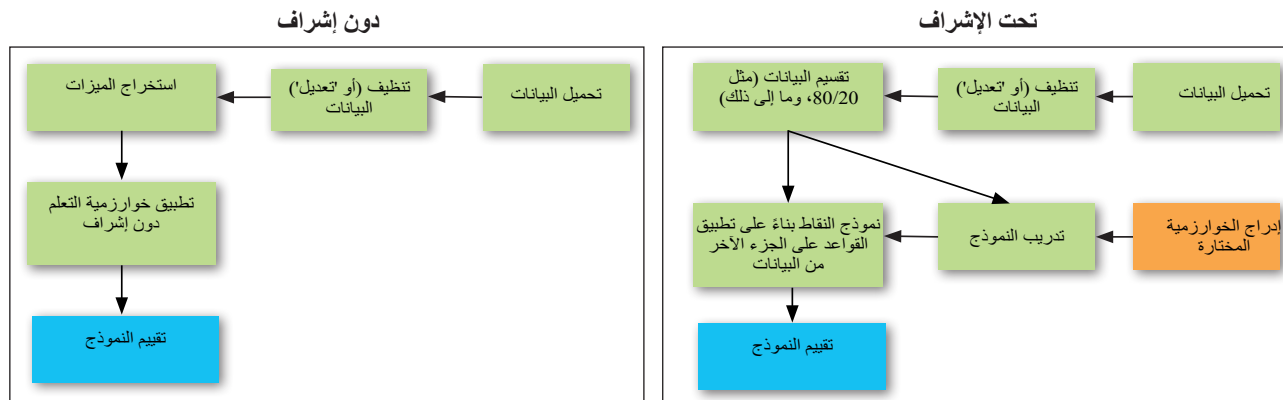
التعلم تحت الإشراف

في التعلم الآلي تحت الإشراف، يتم اختيار خوارزمية مناسبة لحل المشكلة المطروحة ويُستخدم جزء من البيانات المتاحة (عادة 80 في المائة، ولكن يمكن أن يختلف) لتدريب النموذج بشكل مناسب، يجب اختباره قبل استخدامه. وعادة ما ينطوي ذلك على استخدام نسبة 20 في المائة المتبقية من البيانات وتقييم النتيجة (على سبيل المثال، تقييم الدقة الإجمالية للتنبؤاتها).

التعلم دون إشراف

في المقابل، يعتمد التعلم الآلي دون إشراف على تطبيق الخوارزميات (مثل التجميع أو تقليل الأبعاد) للكشف عن الأنماط أو الهياكل أو العلاقات المخفية في البيانات.

يقارن الشكل 21.23 العمليات المعممة لكل من التعلم تحت الإشراف ودون إشراف، وكلاهما يبدأ بتحميل البيانات، وكلاهما ينتهي بتقييم النموذج الذي تم إنشاؤه، لكن تختلف طريقة تقديمهم قليلًا.



الشكل 21.23 عمليات إنشاء نموذج باستخدام التعلم الآلي - تحت الإشراف مقارنةً بالتعلم الآلي دون إشراف.

قد لا تؤدي المحاولات الأولية لإنشاء نموذج ذكاء اصطناعي من خلال عملية التعلم الآلي إلى تنبؤات شديدة الدقة. إذا كانت هذه هي الحالة، سيحتاج عالم البيانات إلى تعديل عملية التعلم عن طريق تعديل اختيار ميزات النموذج أو الأوزان، أو عن طريق تغيير معاملات الضبط التي تتحكم في الخوارزمية نفسها. بمجرد القيام بذلك، يمكنه تشغيل عملية التدريب مرة أخرى، على أمل أن تكون النتائج أفضل.

المصطلحات الرئيسية

الميزات الرئيسية – الميزات المحددة الشائعة في العديد من المثيلات من نوع الكائن نفسه.

الدقة العامة – مقياس يصف فعالية النموذج بشكل عام عند التنبؤ بالبيانات أو تصنيفها على نحو صحيح في أثناء مرحلة الاختبار، فإذا خضع أحد النماذج للتدريب على اكتشاف البريد الإلكتروني العشوائي وتمكن من تحديد 45 رسالة من أصل 50 رسالة بريد إلكتروني بوصفها محتوى غير مرغوب فيه بشكل صحيح، فستكون دقته العامة 90 في المائة (وهي في الواقع منخفضة بعض الشيء).

لمساعدتك على الفهم، عليك مراعاة السيناريو الآتي؛ تخيل أنك عالم بيانات، وقد جمعت الكثير من صور القطط وحددت الميزات الرئيسية التي ستعتبرها مؤشرًا بصريًا للقطط:

- شكل عيونها
- وجود شوارب
- شكل أذنانها
- وجود فراء
- نسيج الفراء الخاص بها.

يمكن تعريف هذه الميزات بوضوح من الصور التي حددتها. تمثل هذه الصور (والميزات الرئيسية التي تحتوي عليها) بيانات التدريب. هدفك هو استخدامها لإنشاء نظام ذكاء اصطناعي يمكنه تصنيف الصورة بشكل صحيح كقطعة (أو لا).

أنت تختار الخوارزمية الخاصة بك ونهج تعلم الآلة الخاص بك. قد تتضمن المعلمات الفائقة لنظام تعلم الآلة الخاص بك:

- مدى سرعة تعلم الخوارزمية (معدل التعلم)
- عدد الصور التي تغذي الخوارزمية دفعة واحدة (حجم الدفعة).

ستشغل عملية التعلم الآلي الخاصة بك وستحصل على نتيجة أولية. ستبدأ بعد ذلك في تعديل الميزات المحددة والمعايير الفائقة المستخدمة، ما يترتب عليه إمكانية تحسُّن **الدقة العامة** لتوقعات نموذج الذكاء الاصطناعي الخاص بك (ومن الوارد أن تتدهور)، وستواصل إجراء المزيد من التجارب حتى تصل إلى أفضل تركيبة.

بمجرد أن يتم تدريب النموذج الخاص بك بنجاح، يمكنك عرض صور جديدة عليه وسؤاله عما إذا كانت الصورة تحتوي على قطعة (أم لا). من الناحية النظرية، إذا كان معدل الدقة العامة للنموذج مرتفعًا، فيجب أن يكون قادرًا على الإجابة عن السؤال بنجاح.

اختيار وتطبيق نماذج الذكاء الاصطناعي لتلبية متطلبات المشروع

لا يمكن للذكاء الاصطناعي الإجابة عن كل سؤال. ومع ذلك، يمكن تدريب الذكاء الاصطناعي (وعملية التعلم الآلي الأساسية) بنجاح للإجابة بسرعة عن العديد من الأسئلة التي ستستغرق من الإنسان وقتًا وجهدًا كبيرين.

نوع السؤال سيحدد أنواع نماذج الذكاء الاصطناعي التي تريد إنشاؤها. مفتاح ذلك هو فهم الأنواع المختلفة من الخوارزميات التي يمكن استخدامها لإنشاء أنواع مختلفة من النماذج. يستكشف هذا الجزء الأنواع الأكثر شيوعًا.

يرد في ما يأتي بعض الأمثلة على الأسئلة التي يمكن تدريب الذكاء الاصطناعي على الإجابة عنها.

- هل يحتوي هذا الملف على فيروس يمكن تعرّفه؟
- ما لوحة تسجيل هذه السيارة المسرعة؟
- متى يجب أن أستثمر في الأسهم والحصص في النشاط التجاري X؟
- إذا استثمرت 1000 دولار الآن، فما العائد المحتمل على استثماري في غضون خمس سنوات؟
- للإجابة عن هذه الأسئلة باستخدام الذكاء الاصطناعي، يجب عليك تحديد:
- نوع المشكلة، وسيترتب عليه تحديد الخوارزمية التي سيتم استخدامها
- نوع تقنية (تقنيات) التعلم التي يجب استخدامها.

تصنيف النموذج المطلوب بحسب الإدخال

يمكنك تصنيف النموذج المطلوب بناءً على مدخلاته، حيث يحدد ذلك طريقة تعلمه.

يمكن عادةً تحقيق ذلك باستخدام التعلم تحت الإشراف أو التعلم دون إشراف (كما هو موضح أعلاه). يسلط الجدول 21.8 الضوء على الخصائص الرئيسية لهذين النهجين.

المهارات

المهارات المعرفية: العمليات والإستراتيجيات المعرفية:

- التفكير الناقد
- حل المشكلات

المهارات المعرفية: الإبداع:

- الابتكار

مهارات التواصل الشخصي: أخلاقيات العمل/الضمير:

- المسؤولية

المصطلحات الرئيسية

خوارزمية الانحدار – تعثر هذه الخوارزمية على أفضل خط أو منحني لملاءمة نقاط البيانات باستخدام ثلاثة مقاييس رئيسية (التباين، والخطأ، والتحيز).

خوارزمية التصنيف – هي خوارزمية تقيم وزن خصائص البيانات المختلفة بناءً على أهميتها للنتائج المطلوبة.

الجدول 21.8 التعلم تحت الإشراف مقارنةً بالتعلم دون إشراف، يُقارن النهجين والخصائص الرئيسية في كل حالة

التعلم تحت الإشراف	التعلم دون إشراف
<ul style="list-style-type: none"> مدخلات بيانات التدريب معنونة يدوياً مقترن بمخرجات متطابقة يمكن أن يكون مستهلكاً للوقت ومكلفاً الهدف هو التنبؤ بمخرجات البيانات غير الموجودة في مجموعة التدريب الخاصة بها من خلال علاقات التعلم بين المدخلات والمخرجات. تشمل الخوارزميات الكلاسيكية التصنيف والانحدار التطبيقات الشائعة: تصفية البريد الإلكتروني أو تعرف الصور أو الصوت وما إلى ذلك. 	<ul style="list-style-type: none"> مدخلات البيانات غير معنونة يتم تحديد الأنماط داخل مدخلات البيانات الهدف هو تحديد الأنماط أو الهياكل المخفية في البيانات؛ واكتشاف أوجه التشابه والاتجاهات وما إلى ذلك. أقل تكلفة مع انخفاض التفاعل البشري تشمل الخوارزميات الكلاسيكية التجميع وتقليل الأبعاد التطبيقات الشائعة: تحليل السوق واكتشاف الحالات الشاذة وما إلى ذلك.

يؤثر نوع البيانات وطبيعة النتيجة التي تحاول تحقيقها مع مشروعك للذكاء الاصطناعي في نوع خوارزمية التعلم الآلي المستخدمة في النموذج. في بعض الأحيان يكون هذا الأمر اختياراً سهلاً، فعلى سبيل المثال، إذا كنت تُجري تنبؤاً بسيطاً (لديك فيه متغير مستقل وتريد التنبؤ بمتغير تابع)، فقد يُحل هذا الأمر باستخدام **خوارزمية الانحدار الخطي** (على سبيل المثال، إذا أدخلت القيمة 'x'، فستتنبأ بالقيمة 'y').

في مواقف أخرى، ليس من السهل دوماً فهم خوارزمية تعلم الآلة التي ستريد استخدامها بناءً على البيانات المتاحة التي ترغب في إدخالها، فعلى سبيل المثال، قد تحين أوقات يكون فيها استخدام المصنف (**خوارزمية التصنيف**) أكثر ملاءمة من استخدام تقنيات التجميع (خوارزمية التجميع).

يعتمد النهج الأفضل إلى حد كبير على ما إذا كنت تستخدم بيانات منظمة أو غير منظمة، أو تفكر في تقنيات التعلم تحت الإشراف أو دون إشراف. لمعرفة أمثلة محددة عن كيفية تأثير المدخلات المختلفة في اختيار خوارزمية التعلم الآلي لاستخدامها في النموذج، راجع أمثلة الأكواد في هذا الجزء. والتي توضح ميزات مجموعات البيانات التي تم استخدامها كمدخلات وخوارزميات التعلم الآلي المرتبطة التي قامت بمعالجتها لإنشاء نموذج عملي.

يجب أن تلاحظ أن الأمور نادراً ما تنجح في المرة الأولى عند بناء مشاريع الذكاء الاصطناعي وقد تحتاج إلى إعادة المحاولة لعدة مرات قبل تحقيق مستويات دقة مرضية.

تصنيف النموذج بحسب المخرجات المطلوبة

من الممكن أيضاً تصنيف النموذج بحسب نوع المخرجات التي يُطلب من الذكاء الاصطناعي توليدها. تشمل أنواع المخرجات:

- الرقم - يتم تحقيق ذلك باستخدام خوارزمية الانحدار
- الفئة - يتم تحقيق ذلك باستخدام خوارزمية التصنيف
- باقية من المجموعات - يتم تحقيق ذلك باستخدام خوارزمية التجميع
- الميزات الرئيسية في البيانات المعقدة - يتحقق ذلك باستخدام تقليل الأبعاد (أي باستخدام خوارزمية تقليل الأبعاد).

ستدرس في الجزء الآتي كلاً منها بمزيد من التفصيل.

تطبيق واستخدام النماذج لمشكلات الانحدار

تُستخدم نماذج الانحدار للتنبؤ بمتغير النتيجة المستمر (التابع) استناداً إلى عدد من ميزات الإدخال (أو الميزات المستقلة). نماذج الانحدار تتنبأ بالأرقام.

المهارات

- المهارات المعرفية: العمليات
والإستراتيجيات المعرفية:
- حل المشكلات

هناك ثلاثة أنواع أساسية يجب مراعاتها:

الانحدار الخطي

قد تكون على دراية بالانحدار الخطي والمعادلة:

$$y = mx + c$$

وذلك عندما:

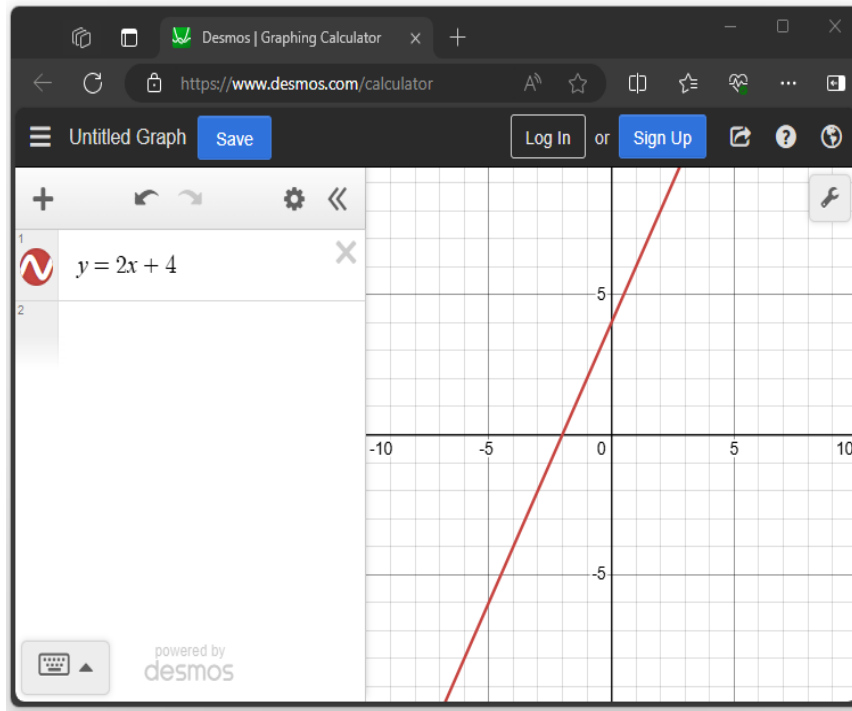
x متغير مستقل يمثل المدخلات

y يمثل المخرجات المتوقعة، وهو متغير تابع

m يمثل ميل خط الانحدار

c يمثل نقطة تقاطع المحور y لخط الانحدار (ببساطة، قيمة y عندما تكون x تساوي 0).

على سبيل المثال، رسم $y = 2x + 4$ يعطي الرسم البياني في الشكل 21.24.



الشكل 21.24 آلة حاسبة رسومية ديسموس $y = 2x + 4$.

باستخدام هذه التقنية، يجب أن يكون من الممكن أخذ أي قيمة للمحور x والعثور على قيمتها التابعة (المتوقعة) للمحور y . يمكن تحقيق ذلك بصريًا عن طريق رسم خط عمودي من القيمة المختارة على المحور x والتوقف عند مواجهة التدرج ثم رسم خط أفقي من تلك النقطة إلى المحور y وقراءة القيمة الناتجة.

في الذكاء الاصطناعي، يمكنك استخدام التعلم الآلي لرسم الخط الأنسب من البيانات المحددة، بدلاً من استخدام معادلة بسيطة.

يتم ذلك عن طريق تقليل دالة الخسارة، عادةً متوسط الخطأ التربيعي (MSE). بشكل عام، يشير انخفاض متوسط الخطأ التربيعي إلى أداء أفضل للنموذج، على الرغم من أن الرقم الدقيق يختلف بناءً على عوامل مختلفة (على سبيل المثال، طبيعة البيانات، معرفتك الأساسية بالبيانات في هذا المجال، وما إلى ذلك).

في الشكل 21.25، يمكنك رؤية هذا مطبقًا في بايثون باستخدام مكتبة `scikit-learn` ومجموعة بيانات تركيبية.

المصطلح الرئيس

متوسط الخطأ التربيعي (MSE) – يقيس
متوسط الفرق التربيعي بين القيم المتوقعة والقيم الفعلية.

```

import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Generate synthetic data
np.random.seed(42)
X = 2 * np.random.rand(100, 1) # Generate 100 random values between 0 and 2
y = 4 + 3 * X + np.random.randn(100, 1) # y = 4 + 3x + random noise

# Train linear regression model
lr_model = LinearRegression()
lr_model.fit(X, y)

# Plot the data points
plt.scatter(X, y, color='blue', label='Data Points')

# Plot the best-fit line
plt.plot(X, lr_model.predict(X), color='red', label='Best-Fit Line')

# Plot settings
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Linear Regression')
plt.legend()
plt.grid(True)
plt.show()

# Calculate Mean Squared Error (MSE)
y_pred = lr_model.predict(X)
mse = mean_squared_error(y, y_pred)
print(f'Mean Squared Error (MSE): {mse}')

# Test of x to find y
test_x = 1.5 # Test x value
test_y = lr_model.predict([[test_x]]) # Predict corresponding y value
print(f'For x = {test_x}, predicted y = {test_y[0][0]:.2f}')

```

الشكل 21.25 كود بايثون يوضح استخدام مكتبة scikit-learn ومجموعة بيانات اصطناعية. كما ترى، من أجل كتابة هذا البرنامج النصي، كان من الضروري تضمين وحدات معينة في الجزء العلوي من نص بايثون (على سبيل المثال، numpy و matplotlib و sklearn). في بعض الأحيان تكون الوحدات التي تقوم بتضمينها جزءاً من مكتبة بايثون القياسية وسيتم تثبيتها بالفعل على جهاز الحاسب الخاص بك. في حالات أخرى، تكون هذه مكتبات تابعة لجهات خارجية (أي لم يتم إنشاؤها بواسطة مصممي بايثون) ويجب تثبيتها بشكل منفصل. عادةً ما يتم تحقيق ذلك في بايثون باستخدام أداة 'pip'، التي تقوم بتنزيل الحزم من فهرس حزم بايثون المعتمد (pypi.org).

إذا حاولت تشغيل برنامج بايثون النصي الذي لا يتضمن الوحدات الصحيحة وما إلى ذلك، فمن المحتمل أن يحدث خطأ، على سبيل المثال:

المتوسط ([1, 2, 4])

Traceback (most recent call last):

File "<pyshell#0>", line 1, in <module>

mean([1, 2, 4])

NameError: name 'mean' is not defined

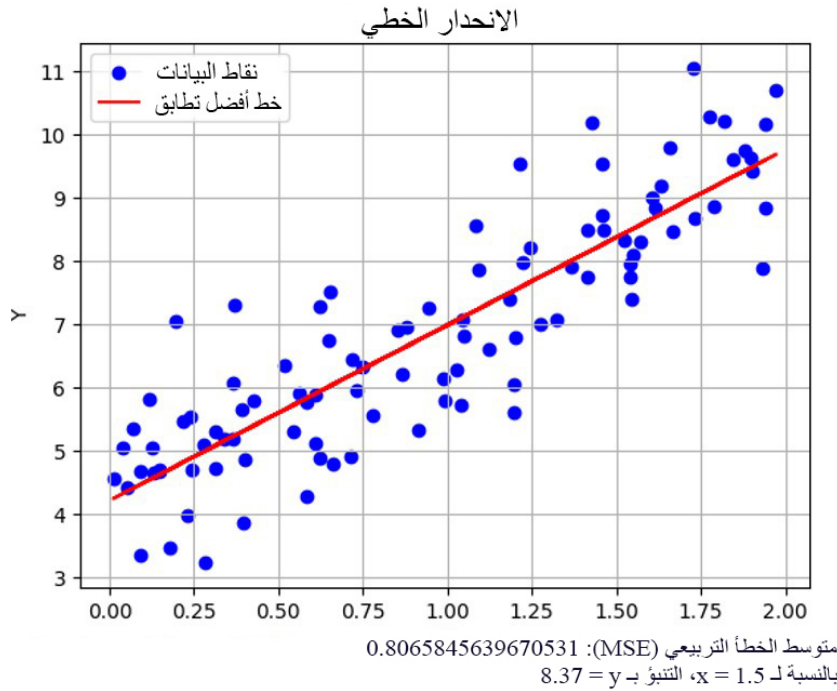
ستحتاج إلى تحديد الوحدة المفقودة، (ربما تثبيتها، ثم تضمين سطر (في الجزء العلوي من البرنامج النصي الخاص بك) لاستخدامها. كما ترى، سيؤدي ذلك إلى حل المشكلة.

من متوسط الإحصاءات الواردة

المتوسط ([1, 2, 4])

2.3333333333333335

يظهر المخرج الناتج في الشكل 21.26.



الشكل 21.26 خط الانحدار الأفضل المحسوب من خلال نقاط البيانات الملتقطة.

كما ترى، تم تدريب النموذج ورسم الخط الأفضل ملائمة واختباره باستخدام قيمة عينة للمحور X ، ما أدى إلى توليد قيمة تنبؤية للمحور y . يجب أن تكون قادرًا على التحقق من ذلك تقريبًا من خلال العمل من المحور x (للمحور X المستقل) إلى الخط الأكثر ملائمة ثم عبر المحور y (للمحور y التابع).

تُستخدم نماذج الانحدار الخطي عادةً للإجابة عن الأسئلة المستندة إلى الأرقام، كما يأتي.

- إذا استثمرت X دولار، فماذا سيكون عائدي المالي في غضون 10 سنوات؟
- إذا اشتريت منزلًا مقابل X دولار، فما قيمته المحتملة في غضون 20 عامًا؟
- إذا خصصت 5 في المائة من راتبي الإجمالي للحصول على معاش شخصي، فما القيمة المحتملة عندما أ تقاعد في غضون X سنوات؟

الانحدار اللوجستي

في بعض الأحيان، لا يكون تحديد علاقة خطية بين X و Y هو أفضل نهج دائماً. على سبيل المثال، تخيل أنك تفترض أن هناك علاقة مباشرة بين عدد الساعات التي تخطط للدراسة والعلامة المتوقعة التي ستحصل عليها للاختبار.

يمكنك بالتأكيد تدريب تلك البيانات، ورسم نقاط البيانات المتناثرة، ورسم خط أفضل ملائمة ومحاولة استخدامه بوصفه تنبؤاً. لكنها تفترض أن النتيجة ستزداد مع كل ساعة إضافية من الدراسة - وهو أمر قد لا يكون صحيحاً في الواقع.

قد يتخذ الانحدار اللوجستي نهجاً مختلفاً قليلاً، فسيسألك عما إذا كنت أكثر عرضة للنجاح أو الفشل، بناءً على ساعات الدراسة المختلفة. سيؤدي هذا بشكل فعال إلى رسم خط منحنى عبر مخطط تشتت لساعات الدراسة ونتائج النجاح/الفشل الناتجة. هذا الخط المنحني سيساعدك ببساطة على فهم احتمال النجاح أو الفشل. تقع هذه الاحتمالية في مكان ما بين 0 (أقل احتمالاً) و 1 (أكثر احتمالاً).

هذا يعني أن الانحدار اللوجستي هو انحدار فئوي (أي، يحدد الفئات 'نجاح' أو 'فشل') بدلاً من تقديم قيمة مستمرة (مثل نسبة التنبؤ) كما يفعل الانحدار الخطي.

يضرب الشكل 21.27 مثالاً باستخدام بايثون لإظهار هذا النموذج في العمل. يظهر مخطط الدقة والتصنيف الناتج في الشكل 21.28.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

# Generate synthetic data
np.random.seed(42)
hours = np.random.randint(1, 16, 100) # Random study hours between 1 and 15
pass_prob = 1 / (1 + np.exp(-(hours - 8))) # Probability of passing increases
with hours, plateauing around 8 hours
outcome = np.random.binomial(1, pass_prob) # Random pass/fail outcome based
on pass probability

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(hours.reshape(-1, 1),
outcome, test_size=0.2, random_state=42)

# Fit logistic regression model
logistic_model = LogisticRegression()
logistic_model.fit(X_train, y_train)

# Predictions
y_pred = logistic_model.predict(X_test)
```

الشكل 21.27 (متابعة)

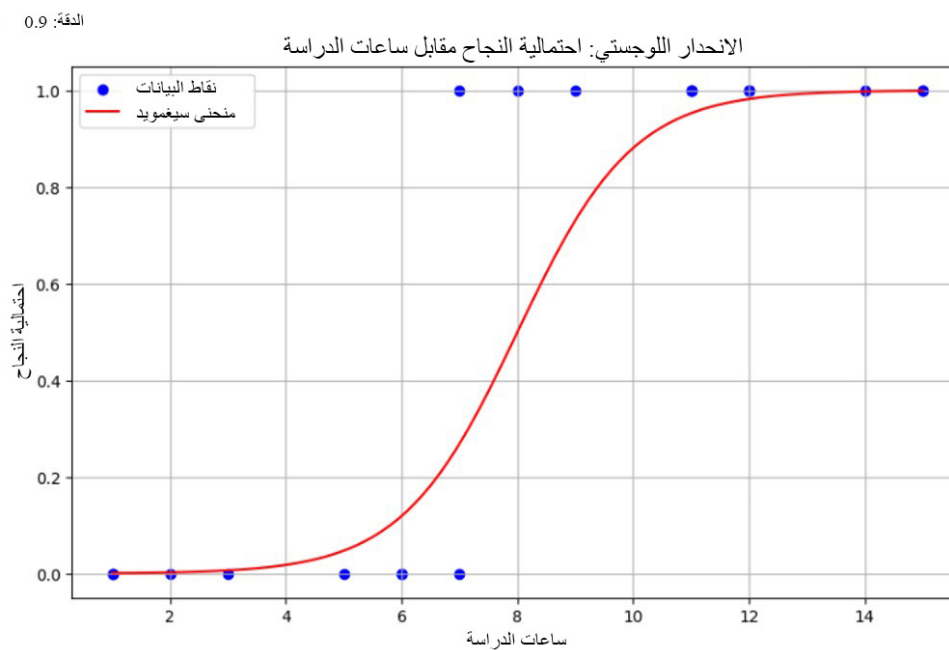
```
# Calculate Accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy}')

# Plot data and decision boundary
plt.figure(figsize=(10, 6))
plt.scatter(X_test, y_test, color='blue', label='Data Points')

# Plot the sigmoid curve (decision boundary)
X_values = np.linspace(1, 15, 100)
Y_proba = 1 / (1 + np.exp(-(X_values - 8))) # Probability of passing for each
X value
plt.plot(X_values, Y_proba, color='red', label='Sigmoid Curve')

plt.xlabel('Study Hours')
plt.ylabel('Probability of Passing')
plt.title('Logistic Regression: Probability of Passing vs Study Hours')
plt.legend()
plt.grid(True)
plt.show()
```

الشكل 21.27 (متابعة) كود الانحدار في بايثون الذي يحسب احتمال اجتياز الامتحان مقابل عدد ساعات الدراسة.



الشكل 21.28 منحنى سيغمويد.

كما ترى، يصبح احتمال النجاح شبه مؤكد (1.0) عندما تصل إلى 12 ساعة من الدراسة للامتحان، ثم ينخفض. يتم رسم التحسين من خلال استخدام منحني على شكل حرف S لتمثيل العلاقات في العالم الحقيقي (وقت الدراسة مقابل احتمال النجاح) الموضح هنا. ومع ذلك، من الممكن مرة أخرى البحث عن أي قيمة X (ساعات الدراسة) والتحقق من احتمالية النجاح، على الرغم من أنه يُرجى ملاحظة أن دقة النموذج تبلغ 90 في المائة فقط.

ARIMA

هذه تقنية قوية تستخدم في تحليل بيانات السلاسل الزمنية. يتم استخدامها للتنبؤ بالقيم المستقبلية بناءً على الملاحظات السابقة أو القيم (المتأخرة).

ومع ذلك، يمثل اسمها الميزات الثلاث التي تجعلها تقنية شائعة - الانحدار الذاتي والمتكامل والمتوسط المتحرك (كما هو موضح في الجدول 21.9).

الجدول 21.9 مقارنة بين تقنيات AR (الانحدار الذاتي) و I (التكامل) و MA (المتوسط المتحرك)، مع توضيح ميزات كل منها.

AR	I	MA
<ul style="list-style-type: none"> • AR = الانحدار الذاتي • نماذج العلاقة بين الملاحظة والقيمة السابقة • يفترض أن القيمة في وقت معين تعتمد على قيمها السابقة 	<ul style="list-style-type: none"> • I = متكامل • يستخدم تقنية تسمى التمييز يتضمن ذلك محاولة إزالة الاتجاهات والفترات من البيانات (والتي تسبب تقلبات) • يهدف إلى جعل السلسلة أكثر ثباتًا وقابلية للتنبؤ • يُستخدم مصطلح 'الثبات' لوصف هذا المعنى 	<ul style="list-style-type: none"> • MA = المتوسط المتحرك • تقنية تُستخدم "التسوية" للتقلبات أو الضوضاء في مجموعات البيانات من خلال تحديد الاتجاهات الأساسية • تستخدم (نافذة) متحركة لنقاط البيانات • بيانات جديدة تحل محل البيانات القديمة لمنع تأثير المتوسط بالقيم العليا والدنيا غير العادية، على سبيل المثال، تأثيرات موسمية مثل زيادة مبيعات الأيس كريم في الصيف

مثال عملي

يحاكي كود بايثون في الشكل 21.29 نموذج من نوع ARIMA باستخدام بعض البيانات الزمنية الاصطناعية. يظهر الإخراج الناتج في الشكل 21.30. يمكنك أن ترى تأثير التسوية الذي قلل التقلبات والضوضاء الموجودة في مجموعة البيانات.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA

# Generate synthetic time series data
np.random.seed(42)
data = np.random.randn(100)
```

الشكل 21.29 (متابعة)

متابعة

```

# إنشاء مجموعات التدريب والاختبار
train_size = int(len(data) * 0.8)
train_data, test_data = data[:train_size], data[train_size:]

# Fit ARIMA model
order = (1, 1, 1) # Example ARIMA(1,1,1) order
arima_model = ARIMA(train_data, order=order)
arima_result = arima_model.fit()

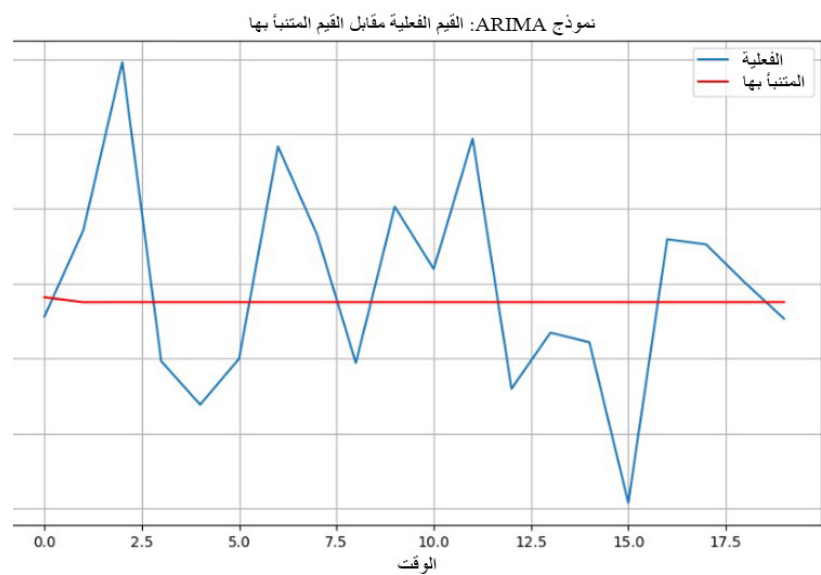
# Make predictions
predictions = arima_result.predict(start=len(train_data), end=len(train_data)
+ len(test_data) - 1)

# Plot actual vs. predicted values
plt.figure(figsize=(10, 6))
plt.plot(test_data, label='Actual')
plt.plot(predictions, color='red', label='Predicted')
plt.title('ARIMA Model: Actual vs. Predicted Values')
plt.xlabel('Time')
plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.show()

# Make a test prediction
test_prediction = arima_result.forecast(steps=1)[0]
print(f"Test Prediction for Next Time Step: {test_prediction}")

```

الشكل 21.29 (متابعة) كود بايثون لمحاكاة نموذج من نوع ARIMA باستخدام بيانات سلسلة زمنية اصطناعية.



الشكل 21.30 التسوية باستخدام نافذة متحركة للتوسيط.

المهارات

- مهارات التواصل الشخصي: أخلاقيات العمل/الضمير:
- المسؤولية
 - الإنتاجية

تطبيق واستخدام النماذج لمشكلات التصنيف

هناك ثلاثة أنواع شائعة من نماذج التصنيف التي تحتاج إلى أخذها في الاعتبار.

SVM (آلات المتجه الداعم)

SVM هو نموذج تعلم آلي قوي ومتعدد الاستخدامات للغاية. يمكنه إجراء تصنيفات خطية أو غير خطية لمجموعات البيانات المعقدة الصغيرة والمتوسطة الحجم.

مثال عملي

المصطلح الرئيس

مشكلة التصنيف – هي مشكلة تقيس متوسط الفرق التربيعي بين القيم المتوقعة والفعالية.

الكود البرمجي في الشكل 21.31 يحاكي آلة الدعم النقطية (SVM) باستخدام تصنيف المتجهات الداعمة (SVC) – نوع من آلات الدعم النقطية مصمم خصيصًا لمهام التصنيف في مكتبة scikit-learn.

يستخدم الكود مجموعة بيانات Iris، وهي مورد شائع ومعروف جدًا في مجال علم البيانات وتعلم الآلة، وغالبًا ما يستخدمه الطلاب عند ممارسة مهاراتهم في **مشكلات التصنيف** وتصور البيانات. تحتوي مجموعة البيانات على قياسات محددة بالسنتيمتر لأربعة ميزات رئيسة لنباتات سيتوزا، وفيرسيكلور وفيرجنكا من فصيلة نبات السوسن:

- طول الكأس
- عرض الكأس
- طول البتلة
- عرض البتلة.

يمكن استخدام مزيج هذه الميزات (بشكل موثوق إلى حد ما) للتنبؤ بخصائص زهرة القزحية. يُرسم التصنيف الناتج في الشكل 21.32.

```
# Importing the required libraries etc.
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC

# Load the Iris dataset
iris = datasets.load_iris()
# Using only the first two features for visualization
X_2d = iris.data[:, :2]
y = iris.target

# Standardizing the features in the data
scaler = StandardScaler()
X_2d_std = scaler.fit_transform(X_2d)

# Training the SVM model with a linear kernel
svm_2d = SVC(kernel='linear', random_state=42)
svm_2d.fit(X_2d_std, y)
```

الشكل 21.31 (متابعة)

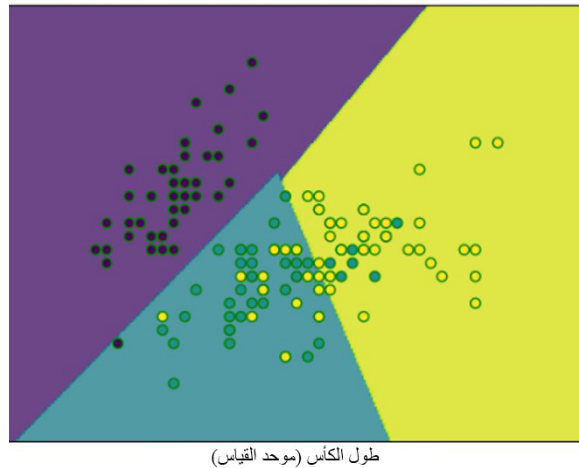
```
# Creating a mesh to plot
x_min, x_max = X_2d_std[:, 0].min() - 1, X_2d_std[:, 0].max() + 1
y_min, y_max = X_2d_std[:, 1].min() - 1, X_2d_std[:, 1].max() + 1
xx, yy = np.meshgrid(np.arange(x_min, x_max, 0.02),
                     np.arange(y_min, y_max, 0.02))

# Predicting each point on the mesh
Z = svm_2d.predict(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)

# Plotting the decision boundaries (as two-dimensional)
plt.contourf(xx, yy, Z, alpha=0.8)
plt.scatter(X_2d_std[:, 0], X_2d_std[:, 1], c=y, edgecolors='g')
plt.xlabel('Sepal length (standardized)')
plt.ylabel('Sepal width (standardized)')
plt.xlim(xx.min(), xx.max())
plt.ylim(yy.min(), yy.max())
plt.xticks(())
plt.yticks(())
plt.title('SVM Linear Kernel on Iris Dataset (First 2 Features)')
plt.show()
```

الشكل 21.31 (متابعة): كود بايثون لمحاكاة SVM باستخدام تصنيف الدعم المتجه (SVC).

استخدام نموذج SVM بنواة خطية على مجموعة بيانات Iris (أول ميزتين)



الشكل 21.32 مخطط SVM يوضح ثلاث حدود في الفضاء ثنائي الأبعاد.

كما يمكنك أن ترى، تم فصل الفئات الثلاثة المختلفة من زهرة السوسن بشكل عام باستخدام الميزتين المختارتين (عرض الكأس وطول الكأس للزهرة). توفر الحدود الملونة تمثيلاً مرئياً واضحاً لمكان وجود الفئات الثلاث المختلفة (بشكل عام) في مجموعة البيانات بناءً على توزيع هذه الميزات.

المصطلحات الرئيسية

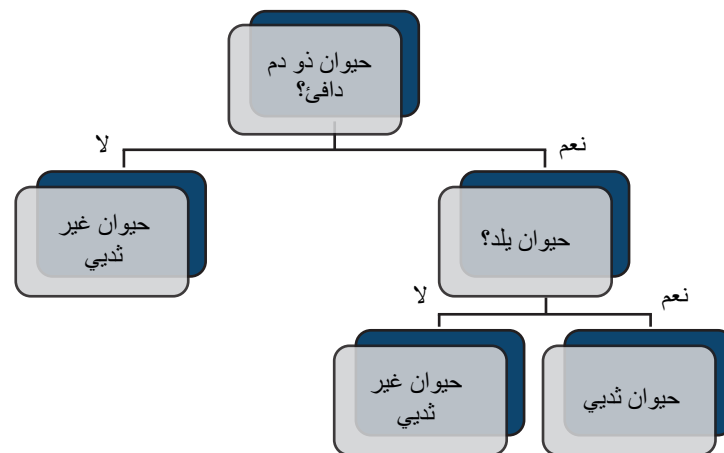
شجرة القرار – هي هيكل هرمي حيث تمثل فيه كل 'عقدة' منفصلة اختباراً من نوع ما (على سبيل المثال: هل تمطر؟).
الغابة العشوائية – مجموعة من أشجار القرارات التي تنشأ لتحسين دقة تصنيفاتها.

الأشجار والغابات

في سياق التعلم الآلي للتصنيف، يشير المصطلحان "الأشجار" و "الغابات" إلى أشجار القرار والغابات العشوائية.

شجرة القرار هي هيكل هرمي تمثل فيه كل "عقدة" منفصلة اختباراً من نوع ما (على سبيل المثال: هل تمطر؟). يمثل كل فرع من فروع الشجرة نتيجة مختلفة لهذا الاختبار (على سبيل المثال، (نعم) أو (لا)). تمثل العقد الورقية (تلك التي لا تحتوي على عقد فرعية) تسمية الفئة - وهو قرار يتم اتخاذه. يوفر لك اتباع المسارات من العقدة الجذرية (العلوية) إلى الورقة قواعد يمكنك استخدامها للتصنيف.

توفر شجرة القرارات في الشكل 21.33 نظرة ثاقبة حول كيفية إجراء سلسلة من الاختبارات لتحديد ما إذا كان المخلوق من الثدييات أو غير الثدييات.



الشكل 21.33 شجرة قرارات توضح نتيجة سلسلة من أسئلة القرار يُجاب عليها بنعم/لا وتُستخدم للوصول إلى التصنيف.

يشار إلى الغابة العشوائية عادةً باسم أسلوب التعلم (الجماعي). هذا لأنه يبني أشجار قرارات متعددة (حرفياً غابة) لتحسين دقة تصنيفاتها. ويقوم بذلك عن طريق تدريب كل شجرة على مجموعة فرعية عشوائية من البيانات المتاحة. ثم يقوم بحساب متوسط هذه التوقعات للحصول على نتيجة أكثر دقة.

تحظى هذه النوعية من تقنيات التعلم بشعبية بسبب فعاليتها العامة عبر مجموعة واسعة من أنواع البيانات المختلفة ومشكلات العالم الحقيقي، مثل التصنيف والانحدار واكتشاف الحالات الشاذة. والأهم من ذلك أنها سهلة التفسير نسبياً.



مثال عملي

يوضح كود بايثون في الشكل 21.34 استخدام مصنف شجرة القرار. الدقة وشجرة القرار التي يولدها الكود موضحة في الشكل 21.35.

```
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree

# Load the Iris dataset
iris = load_iris()
X = iris.data
y = iris.target

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)

# Initialize the Decision Tree Classifier Algorithm
dtree = DecisionTreeClassifier(random_state=42)

# Train our model
dtree.fit(X_train, y_train)

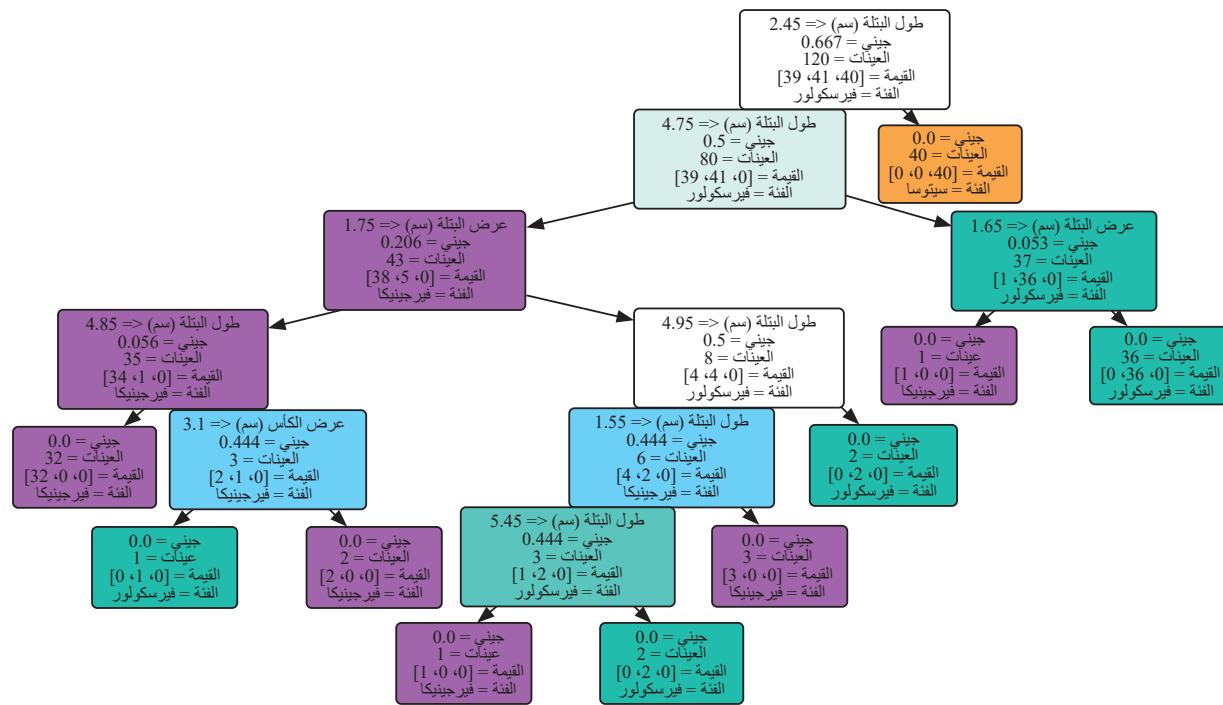
# Predict the labels of the test set
y_pred = dtree.predict(X_test)

# Calculate the general accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy}')
```

```
# Plot the decision tree
plt.figure(figsize=(20,10))
tree.plot_tree(dtree, filled=True, feature_names=iris.feature_names,
               class_names=iris.target_names, rounded=True)
plt.show()
```

الشكل 21.34 كود بايثون، يوضح استخدام مصنف شجرة القرار.





الشكل 21.35 شجرة القرار لمجموعة بيانات السوسن التي توضح كل قرار تم اتخاذه في الشجرة لتحديد فئة السوسن ذات الصلة.

شجرة القرار المنتجة لمجموعة بيانات السوسن تستخدم البيانات لتحديد عدة مسارات مختلفة. يُحدد كل مسار بواسطة مجموعة من قيم البيانات المختلفة، التي تساعد على تحديد احتمالية كون السوسن من فئة معينة (إماسيتوزا، فيرسيكور وإمافيرجكا). إذا أخذت الآن عينة جديدة من زهرة السوسن، واستخدمت قياسات البيانات المناسبة، يجب أن تكون قادرًا على اتباع المسار المناسب وتحديد فئتها الأكثر احتمالاً.

طرق بايزي (Bayesian Methods)

هذه مجموعة من الأساليب الإحصائية، تعتمد على نظرية بايز ، التي تصف احتمال حدوث حدث بناءً على المعرفة السابقة بالأحداث السابقة التي قد تكون ذات صلة.

يمكن تلخيصها باستخدام ثلاثة مصطلحات:

P – يشير إلى الاحتمال اللاحق (احتمال وقوع حدث)

H – يشير إلى الفرضية

E – يشير إلى الدليل

هذا يؤدي إلى عدد من الصيغ الرياضية المختلفة - على سبيل المثال، $P(H|E)$ هو الاحتمال اللاحق للفرضية H ، بالنظر إلى الدليل E .

قد يحدث تطبيق بسيط لطرق بايزي في قطاع الصحة. تخيل طبيبًا يحاول تشخيص مرض مريضه بناءً على مجموعة من الأعراض.

- من المرجح أن يستخدم الطبيب خبرته ومعرفته لتحديد احتمالية الإصابة بأمراض مختلفة قد تسبب أعراض المريض. سيكون هذا هو (الاحتمال المسبق) والفرضية التي يبدأ بها الطبيب.
- بمجرد إبداء ملاحظات جديدة حول الأعراض، ومقارنة احتمالية رؤية هذه الأعراض مع كل من الأمراض التي نظروا فيها في الأصل، يمكنهم جمع أدلة جديدة.

المصطلح الرئيس

التجميع – أحد أنواع مشكلات التعلم دون إشراف، فالهدف المطلوب بلوغه هو فحص نقاط البيانات المقدمة وتحديد التجميع الطبيعي أو "المجموعات" داخل مجموعة البيانات.

تطبيق واستخدام النماذج لمشكلات التجميع

على عكس مشكلة التعلم الخاضع للإشراف، ستحتاج الخوارزمية المختارة لمشكلات **التجميع** إلى تحديد الأنماط أو أوجه التشابه من مجموعة من نقاط البيانات، والتي تمثل بشكل أساسي الميزات المحددة التي تمثل بشكل أفضل كل مجموعة طبيعية (محتملة). هذا يعني أن التدريب باستخدام أمثلة معنونة ليس مطلوباً فعلياً، وهذا يساعد عادةً على تقليل التكاليف.

تشمل الأنواع الشائعة من المشكلات الواقعية التي يمكن أن يساعد التجميع على حلها:

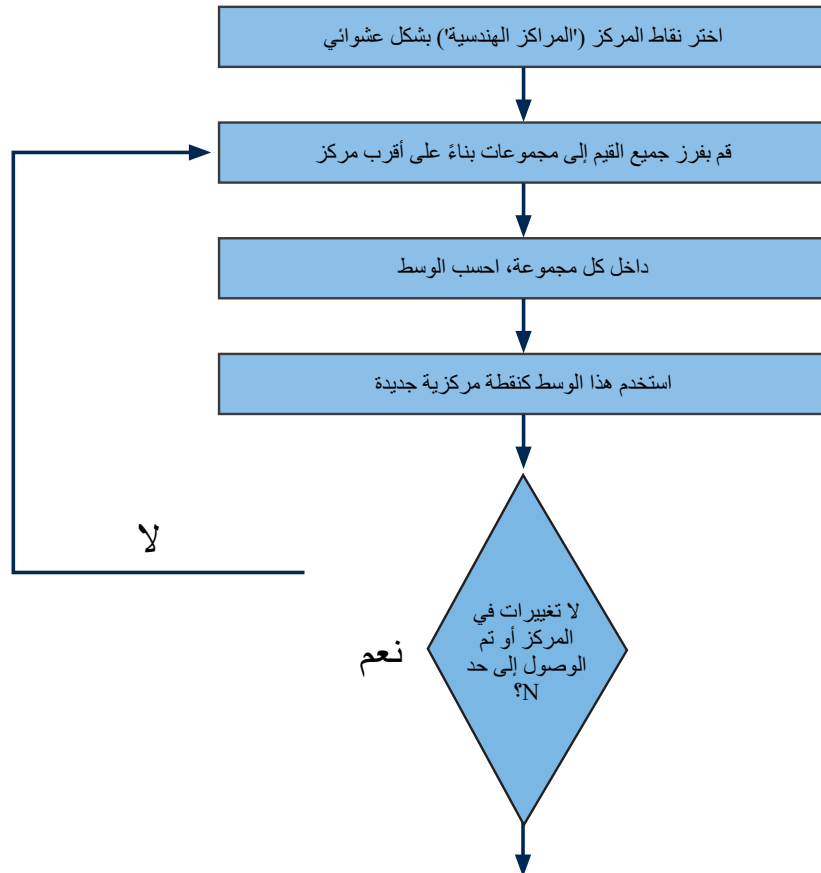
- تحليل البيانات الاستكشافي – العثور على الأنماط أو العلاقات المخفية في البيانات.
- التسويق المخصص للعملاء – تقسيم العملاء إلى مجموعات مختلفة بناءً على عادات التسوق أو التركيبة السكانية وجعل عروضهم أكثر استهدافاً وأكثر جاذبية.
- اكتشاف الانحرافات - العثور على نقاط البيانات غير العادية والتي تتحرف بشكل كبير عن المجموعات أو الأنماط المحددة.

K-means

K-means هي واحدة من خوارزميات التجميع الأكثر شيوعاً المستخدمة لتنفيذ هذا النوع من المهام. والتي تعمل بشكل أساسي باستخدام خوارزمية سهلة الفهم نسبياً. تمثل خوارزمية K ببساطة عدد المجموعات التي يشتبه في وجودها في البداية. عند تنفيذه، يمثل N الحد الأقصى لعدد التكرارات التي سيتم تنفيذها. يتم تمثيل هذه الخوارزمية بشكل أفضل في شكل مخطط انسيابي، كما هو موضح في الشكل 21.36.

المهارات

- المهارات المعرفية: العمليات
والاستراتيجيات المعرفية:
- التفكير الناقد
 - حل المشكلات
- الإبداع:
- الابتكار



الشكل 21.36 مخطط انسيابي لخوارزمية K-means.

مثال عملي

المثال في الشكل 21.37 يستخدم بايثون لتوضيح عملية K-means هذه. إنه يعيد النظر في مجموعة بيانات السوسن التي تمت مشاهدتها سابقاً ويستخدم تقنيات التجميع لمعرفة ما إذا كان ممكناً استخدام القيم المختارة لتحديد أنواع السوسن المختلفة. الناتج البياني للمجموعات المحددة موضح في الشكل 21.38.

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.datasets import load_iris
from sklearn.cluster import KMeans

# Load Iris dataset
iris = load_iris()
X = iris.data

# Define the number of clusters (K)
k = 3

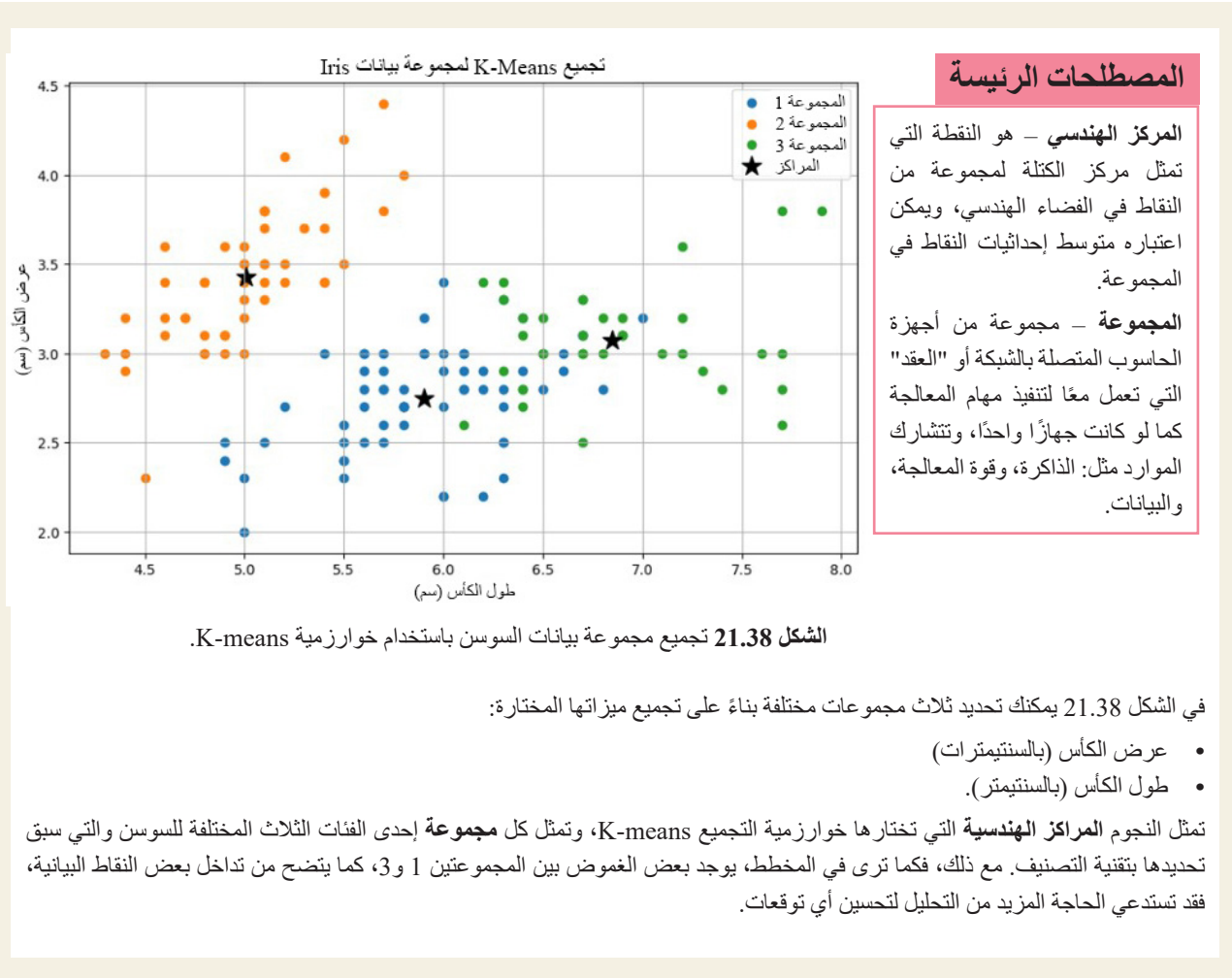
# Initialize and fit K-Means model
kmeans = KMeans(n_clusters=k, random_state=42)
kmeans.fit(X)

# Get cluster centroids and labels
centroids = kmeans.cluster_centers_
labels = kmeans.labels_

# Visualize the clusters
plt.figure(figsize=(10, 6))
for i in range(k):
    cluster_points = X[labels == i]
    plt.scatter(cluster_points[:, 0], cluster_points[:, 1], label=f'Cluster {i+1}')

plt.scatter(centroids[:, 0], centroids[:, 1], marker='*', s=200,
            color='black', label='Centroids')
plt.xlabel(iris.feature_names[0])
plt.ylabel(iris.feature_names[1])
plt.title('K-Means Clustering of Iris Dataset')
plt.legend()
plt.grid(True)
plt.show()
```

الشكل 21.37 كود بايثون يوضح عملية K-means.



تحليل المكونات الرئيسية (PCA)

يرمز PCA إلى تحليل المكونات الرئيسية، وهو نوع شائع من **تقليل الأبعاد الخطية** الإحصائية الذي يستخدم التعلم دون إشراف. تقليل الأبعاد الخطية هو خوارزمية تهدف إلى تقليل عدد ميزات البيانات مع الاحتفاظ بأكبر قدر ممكن من المعلومات الأساسية، والأنماط، والعلاقات.

يشار إلى المخرجات من تقليل الأبعاد الخطية على أنها مكونات رئيسية، ومن هنا جاء اسم تحليل المكونات الرئيسية (PCA). عند تقليله بشكل صحيح، يحتوي المكون الأساسي على معلومات أكثر أهمية من الثاني، والثاني أكثر أهمية من الثالث وهكذا. مع دمج كل ميزة، تصبح الأهمية النسبية والوزن أكبر، لذلك فإن تقليل الميزات من أربعة إلى اثنتين (إذا تم بشكل صحيح) لا يُفقد أي بيانات مهمة وقد يكون في الواقع يُكسب بعض البيانات.

يعد تقليل الأبعاد أمراً مهماً، لأنه كلما زاد عدد ميزات البيانات المستخدمة، زادت التباديل الممكنة، وزادت صعوبة معالجة البيانات بسرعة وكفاءة.

المصطلح الرئيس

تقليل الأبعاد الخطية – هو خوارزمية تهدف إلى تقليل عدد ميزات البيانات مع الاحتفاظ بأكبر قدر ممكن من المعلومات الأساسية، والأنماط، والعلاقات.

مثال عملي

يوضح الشكل 21.39 مثالاً على استخدام تحليل المكونات الرئيسية (PCA) في بايثون لتقليل أربعة ميزات من مجموعة بيانات السوسن (طول الكأس، عرض الكأس، طول البتلة، عرض البتلة) ليصبحا مكونين رئيسيين. يظهر الإخراج الناتج في الشكل 21.40.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.decomposition import PCA

# Load the Iris dataset
iris = load_iris()
X = iris.data
y = iris.target

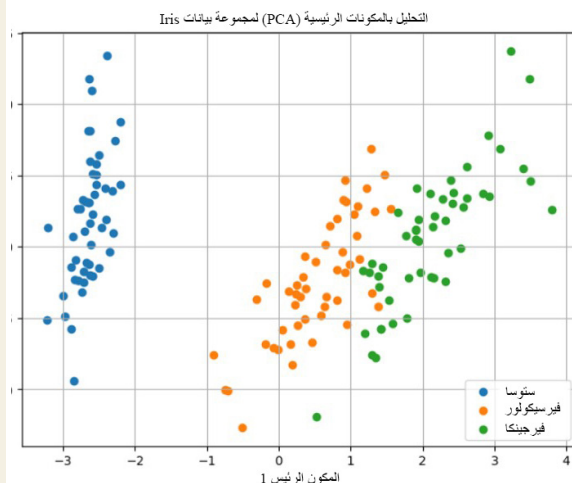
# Apply PCA to reduce dimensionality to 2 principal components
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

# Plot the data in the reduced-dimensional space
plt.figure(figsize=(8, 6))
for i in range(len(np.unique(y))):
    plt.scatter(X_pca[y == i, 0], X_pca[y == i, 1],
                label=iris.target_names[i])

plt.title('PCA of Iris Dataset')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.legend()
plt.grid(True)
plt.show()
```

الشكل 21.39 يُستخدم تحليل المكونات الرئيسية (PCA) في بايثون لتقليل أربع ميزات ليصبحا مكونين رئيسيين.

دون تقليل عدد الميزات من أربعة إلى اثنين، كان من الصعب رسم البيانات في فضاء ثنائي الأبعاد.



الشكل 21.40 نتائج تحليل المكونات الرئيسية (PCA) مع النقاط التي تتعلق بكل فئة حددت بصرياً باستخدام ألوان مختلفة.

المهارات

المهارات المعرفية: الإبداع:

• الإبداع

مهارات التواصل الشخصي: أخلاقيات

العمل/الضمير:

• المسؤولية

تطبيق واستخدام نماذج للكشف عن الانحرافات

كما تمت مناقشته، يمكن استخدام نماذج الذكاء الاصطناعي لاكتشاف الحالات الشاذة في البيانات. قد تتضمن أمثلة الحالات الشاذة البريد الإلكتروني العشوائي بدلاً من البريد الإلكتروني العادي، أو (في الطب) شذوذ الأنسجة الخبيث وليس الحميد.

يمكن استخدام العديد من خوارزميات التعلم الآلي تحت الإشراف والتعلم الآلي دون إشراف لأداء هذا النوع من المهام، لكن ربما يكون الجيران الأقرب (kNN) هم النهج الأكثر بساطة وفعالية للاستخدام.

الجيران الأقرب (kNN)

تعمل هذه الخوارزمية عن طريق حساب المسافة بين نقطة البيانات وأقرب جيرانها k في مجموعة بيانات التدريب. إذا كانت النقطة البيانية أبعد بكثير من جيرانها الأقربين وفقاً لحد مفضل، يتم تصنيفها على أنها انحرافاً.

يمكن أن يختلف الحد الفعلي المستخدم، ويتم تحديده من خلال مزيج من المعرفة الإحصائية ومعرفة المجال.

مثال عملي

المثال البرمجي في بابثون في الشكل 21.41 ينشئ بيانات اصطناعية، بما في ذلك بعض البيانات التي يمكن اعتبارها انحرافات. ثم يحاول ملاءمة نموذج kNN فقط لنقاط البيانات العادية. ثم يتم حساب المسافات لجميع النقاط إلى أقرب نقاط k . أي نقاط تكون فوق نسبة 95% تعتبر انحرافاً. يظهر الإخراج الناتج في الشكل 21.42.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neighbors import NearestNeighbors

# Generate synthetic data (normal and anomaly)
np.random.seed(42)
normal_data = np.random.randn(100, 2) # Normal data points
anomaly_data = np.random.uniform(low=-5, high=5, size=(10, 2)) # Anomaly data points

# Concatenate normal and anomaly data
data = np.vstack((normal_data, anomaly_data))

# Fit kNN model
k = 5 # Number of neighbors
knn_model = NearestNeighbors(n_neighbors=k)
knn_model.fit(normal_data) # Fit to normal data only

# Calculate distances to k nearest neighbors for all data points
distances, indices = knn_model.kneighbors(data)
```

الشكل 21.41 (متابعة)

```
# Define threshold for anomaly detection
threshold = np.percentile(distances[:, -1], 95) # 95th percentile of
distances to k nearest neighbors

# Identify anomalies
anomalies = data[distances[:, -1] > threshold]

# Plot data points and anomalies
plt.figure(figsize=(8, 6))
plt.scatter(normal_data[:, 0], normal_data[:, 1], label='Normal Data')
plt.scatter(anomaly_data[:, 0], anomaly_data[:, 1], color='red',
label='Anomaly Data')
plt.scatter(anomalies[:, 0], anomalies[:, 1], color='orange', label='Detected Anomalies')
plt.title('Anomaly Detection with kNN Algorithm')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.legend()
plt.grid(True)
plt.show()

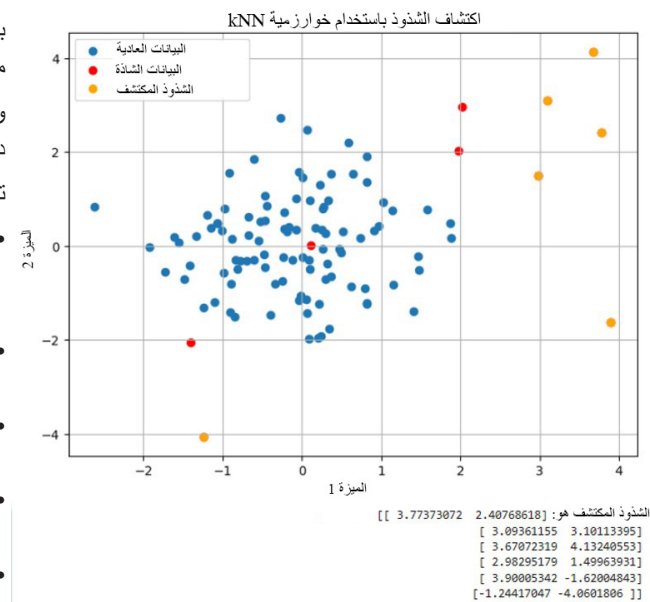
print("Detected anomalies are:", anomalies)
```

الشكل 21.41 (متابعة) كود بايثون لأقرب الجيران باستخدام بيانات اصطناعية.

كما ترى، هذه العملية ليست مثالية - في حين تم تحديد النقاط البعيدة بشكل صحيح، لم يتم تحديد بعضها الأقرب إلى المركز (أي هربت من الاكتشاف).

وبالتالي، سيكون من الضروري تحسين النموذج بشكل أكبر (وزيادة دقته العامة) قبل أن يمكن استخدامه بشكل موثوق في بيئة الإنتاج. تشمل التقنيات الشائعة لتحسين الدقة:

- هندسة الميزات - استخدام المعرفة بالمجال (الفهم الخاص بالموضوع) لتحسين اختيار الميزات المستخدمة في النموذج.
- استخدام خوارزمية أو نموذج مختلف لاكتشاف التأثير في دقة الكشف.
- ضبط النموذج عن طريق تعديل النسب أو الحدود (على سبيل المثال، 95 في المائة المستخدمة في هذا المثال).
- استخدام التحليل الزمني (المعتمد على الوقت) لتطبيق خوارزميات مثل المتوسطات المتحركة لتسوية البيانات بمرور الوقت.
- استخدام واحدة أو أكثر من هذه التقنيات معاً، وتكرارها عبر العديد من الاختبارات. يمكن أن يؤدي ذلك إلى تعزيز دقة (وبالتالي فعالية) اكتشاف الحالات الشاذة لتطبيق معين.



الشكل 21.42 اكتشاف الانحراف باستخدام خوارزمية kNN مع الإشارة بوضوح إلى النقاط الشاذة.

المهارات

- مهارات التواصل الشخصي: الانفتاح الفكري:
- الاهتمام الفكري والفضول

استخدامات وفوائد عيوب الشبكات العصبية

- هناك العديد من الاستخدامات المختلفة للشبكات العصبية. تشمل أربعة تطبيقات رئيسية لهذه التقنية ما يأتي:
- تعرف الأنماط (على سبيل المثال، معالجة الكلام والصورة واللغة الطبيعية (NLP))
- أنظمة التوصية (على سبيل المثال، منتجات التجارة الإلكترونية ومنصات البث ومنصات وسائل التواصل الاجتماعي)
- اكتشاف حالات الانحراف
- أنظمة التحكم المعقدة في الوقت الفعلي (على سبيل المثال، الروبوتات الصناعية والسيارات ذاتية القيادة وما إلى ذلك).

الفوائد الخمس الأساسية لاستخدام الشبكات العصبية موضحة بوضوح أدناه.

- القدرة على التعلم من كميات هائلة من البيانات دون برمجة صريحة. وهذا يُمكنها من اكتشاف الأنماط والروابط المخفية في البيانات.
- إمكانية إجراء معالجة متوازنة للبيانات. هذا يجعلها فعالة للغاية من الناحية الحسابية وقادرة على التوسع بسرعة.
- القدرة على التكيف، لا سيما عند العمل مع علاقات البيانات المعقدة وغير الخطية.
- القدرة على استخراج الميزات الجديدة تلقائيًا من البيانات. هذا يقلل الهندسة اليدوية للميزات التي يجب أن يقوم بها عالم البيانات.
- إمكانية تعميمها جيدًا على البيانات الجديدة. هذا يعني أن نموذجهم يعمل بدقة تامة على البيانات التي لم يتم التدريب عليها.

ومع ذلك، فإن الشبكات العصبية تمثل بعض التحديات التشغيلية.

- شدة التعقيد، ما يجعل من الصعب تفسيرها وبالتالي تحقيق الذكاء الاصطناعي القابل للتفسير (XAI).
 - استهلاك قدر كبير من موارد الحوسبة (على سبيل المثال، GPU وحدات معالجة الرسومات) والذاكرة والأجهزة المتخصصة الأخرى).
 - احتمالية طلبها كميات هائلة من البيانات المعنونة للتدريب الأولي، وهو ما سيكون غالبًا نشاطًا مكلفًا.
 - بالإضافة إلى ذلك، قد لا تتوفر كميات كافية من البيانات ذات الجودة المناسبة بسهولة.
 - تتعرض في بعض الأحيان لتمام الملاءمة ما يمكن أن يمثل مشكلة عند معالجة بيانات جديدة.
- لذلك، من الواضح أن هناك مشكلة تتعلق بالتوازن عند التفكير في الشبكات العصبية. إنها توفر فرصًا رائعة للذكاء الاصطناعي، ولكنها تمثل أيضًا العديد من التحديات الصعبة في الإعداد والتدريب وقابلية الشرح. ومن الجدير بالذكر أيضًا أن الشبكات العصبية يمكن أن تكون عرضة تحديدًا للهجمات العدائية.
- في بعض الأحيان يتم تنفيذ الهجمات العدائية ببساطة لتعطيل العمليات، وأحيانًا لخلق نتائج أكثر ملاءمة للمهاجم، وأحيانًا لتحقيق ميزة تجارية.

الطبيعة الغامضة لعملية اتخاذ القرار في الشبكة العصبية ووجود مثل هذه الثغرات يمكن أن يؤدي إلى مخاوف بشأن العدالة والتحيز والأخلاقيات والتأثيرات الاجتماعية. هذا يمكن أن يؤدي إلى تآكل الثقة العامة والثقة بشكل عام. ومصدر قلق خاص بالنظر إلى الاستخدام المتزايد لهذه الأنظمة في قطاعات التمويل والرعاية الصحية والعدالة الجنائية.

المصطلح الرئيس

الهجمات العدائية – تقع هذه الهجمات عندما يصنع المخربون مدخلات ضارة يمكن أن 'تسمم' بيانات تدريب الشبكة ما ينجم عنه عدم صحة التنبؤات التي يقدمها النموذج.

المهارات

المهارات المعرفية: العمليات
والإستراتيجيات المعرفية:

- التفكير الناقد
- حل المشكلات
- اتخاذ القرار

ج2 أدوات تطوير حلول الذكاء الاصطناعي

إن مشهد أدوات حلول الذكاء الاصطناعي واسع ومعقد على حد سواء، ويضم مجموعة واسعة من المنصات التجارية والمفتوحة المصدر والتطبيقات والمكتبات والأطر، وتقدم مجموعة من الخيارات لأولئك الذين يرغبون في حل مشكلات معقدة في العالم الحقيقي.

تحديد الأدوات المناسبة وتطبيقها لتطوير حلول الذكاء الاصطناعي

أصبح اختيار الأدوات لتطوير حلول الذكاء الاصطناعي أسهل نظرًا إلى أن معظم أدوات الذكاء الاصطناعي ترتبط مباشرة بالمرحلة المختلفة لعملية تطوير الذكاء الاصطناعي/مخططات الذكاء الاصطناعي التي تمت مناقشتها سابقًا.

جمع البيانات وتصنيفها

تتضمن هذه العملية جمع وتصنيف البيانات (التي قد تكون مجمعة من مصادر متعددة) لإنشاء مجموعة بيانات يمكن استخدامها بواسطة نماذج التعلم الآلي.

Labelbox تعد أحد الأمثلة على أداة لهذا النوع من التطبيقات.

استكشاف البيانات وتمثيلها

تتضمن هذه المرحلة تحليل مجموعة البيانات المجمعة لإجراء التجارب واكتساب الأفكار. تشمل الأدوات النموذجية لهذه المرحلة ما يأتي:

- دفتر جويتر (Jupyter Notebook)
- سيبورن (Seaborn)
- الباندا (Pandas)
- تابلو® (Tableau)
- ماتبلوتليب (Matplotlib)

التدريب النموذجي والاختبار والتقييم

تتضمن هذه المرحلة تدريب النموذج باستخدام الميزات المحددة من مجموعة البيانات وخوارزمية التعلم الآلي المحددة والمعلمات الفائقة المرتبطة بها. تشمل الأدوات النموذجية:

- سايكيت - ليرن (scikit-learn)
- بايتورتش® (PyTorch)
- تينسرفلو (TensorFlow)
- كيراس (Keras)

نشر النموذج

بمجرد تدريب النموذج بنجاح إلى مستوى مقبول من الدقة، يتم نشره في بيئات الإنتاج حيث يمكنه عمل تنبؤات بشأن البيانات الجديدة. من الممارسات الشائعة توفير واجهة برمجة تطبيقات قائمة على الويب للخدمات والتطبيقات الأخرى للتفاعل معها مباشرة. تشمل الأدوات النموذجية:

- دوكر® (Docker)
- سيلدون® (Seldon)
- كوبرنيتيس (Kubernetes)
- فاست إيه بي أي (FastAPI)
- خدمة تينسورفلو (TensorFlow Serving)

بحث

اختر أداة واحدة من كل مرحلة من مراحل عملية تطوير الذكاء الاصطناعي واكتشف:

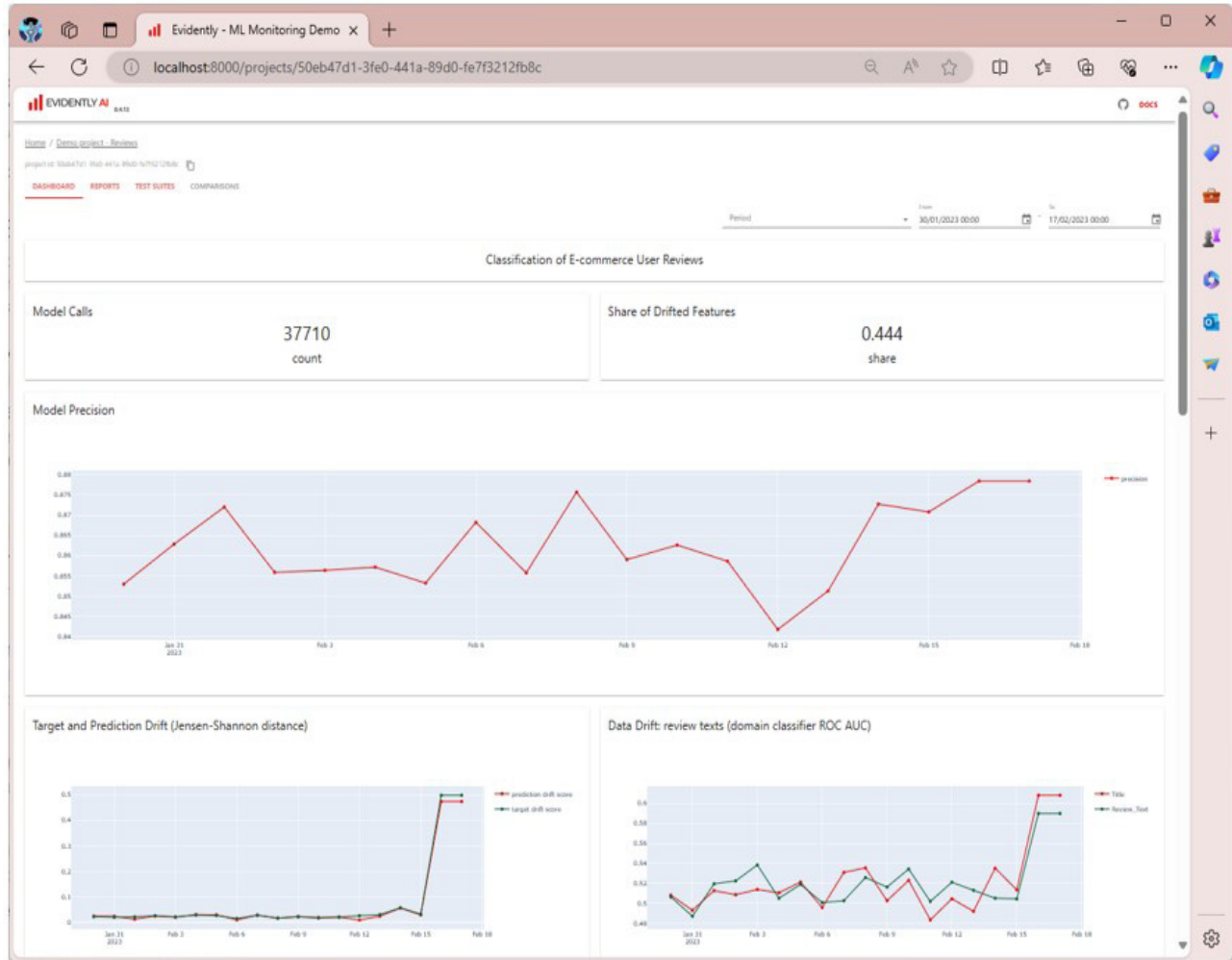
- ما أنظمة التشغيل التي تعمل عليها (على سبيل المثال، Linux، Mac، Microsoft Windows، OS X وغيرها)
- ما المهام التي يؤديها فعليًا
- كيف يمكن استخدامها (على سبيل المثال، يدويًا أو برمجياً)
- تكلفة البرنامج (أي ما إذا كان مفتوح المصدر، ترخيص تجاري، وما إلى ذلك)
- مدى تكاملها مع الأدوات الأخرى (على سبيل المثال، Matplotlib مع scikit-learn، وما إلى ذلك).

تطبيق الذكاء الاصطناعي القابل للتفسير (XAI) واستخدام التصور

مجال آخر للتركيز على الأدوات هو الذكاء الاصطناعي القابل للتفسير (XAI). يعد استخدام الوسائل المرئية وأدوات لوحة التحكم طريقة مثالية لتقديم النتائج وعمليات الذكاء الاصطناعي بطريقة شفافة وذات مغزى. يوجد العديد من المنتجات التجارية مفتوحة المصدر التي تسمح لعالم البيانات بما يأتي:

- تحديد المقاييس الشائعة (على سبيل المثال، الدقة العامة، التغييرات في جودة النموذج، وما إلى ذلك)
- تسجيل هذه المقاييس بانتظام
- إنشاء مكون مرئي (على سبيل المثال، مخطط أو رسم بياني، وما إلى ذلك) لهذه المقاييس
- إنشاء لوحة معلومات تحتوي على عدد من المكونات المرئية المختلفة.

من الواضح أن (<https://docs.evidentlyai.com>) هي أداة مفتوحة المصدر مصممة خصيصًا لمراقبة نماذج التعلم الآلي وتحليلها في أثناء تشغيلها في الإنتاج. بالإضافة إلى ذلك، تدعم Evidently AI تنبيهات قابلة للتخصيص يمكنها تفعيل الإشعارات عندما تنخفض المقاييس عن (أو تتجاوز) مستويات التسامح المحددة.



الشكل 21.43 لوحة تحكم مستندة إلى الويب من Evidently AI مع مكونات تعتمد على الرسوم البيانية لمشروع نموذجي.

موضوعات ذات صلة

الوحدة 4: البرمجة تركز على عدد من لغات البرمجة المختلفة، بما في ذلك بايثون. يجب أن تساعد هذه الوحدة على تعزيز فهمك لتركيبات تلك اللغة. يجب أن يبرز الاختلافات البنيوية مع اللغات الأخرى في قطاع تكنولوجيا المعلومات.

برمجة الحاسب ولغات البرمجة النصية لحلول كتابة الكود

يمكن استخدام العديد من لغات البرمجة المختلفة لتطوير وبرمجة الخوارزميات التي تشكل جوهر معظم حلول الذكاء الاصطناعي. وتشمل هذه اللغات بايثون ولبسب وآر وجافا.

بايثون

بايثون هي واحدة من أكثر اللغات شعبية في العالم وغالبًا ما تشكل مقدمة المتعلم للبرمجة. بدلاً من تفضيل نهج واحد، فإنه في الواقع مزيج من نماذج البرمجة المختلفة، بما في ذلك البرمجة الإجرائية والوظيفية والبرمجة الموجهة للكائنات.

إنها لغة مفسرة، إذ يُترجم أولاً كل توجيه وارد في التعليمات إلى بايت كود بايثون ثم إلى كود الآلة المستهدف لينفذ بوحدة المعالجة المركزية الأصلية. على الرغم من أن هذا النهج لا يجعلها أسرع لغة في الاستخدام، إلا أن طبيعتها ذات الأغراض العامة مناسبة للعديد من الأنشطة المختلفة، بما في ذلك علم البيانات والعمل على الذكاء الاصطناعي.

كما أنها مناسبة بشكل استثنائي للتعامل مع البيانات. بفضل طبيعتها القابلة للتوسعة وقدرتها على استيراد حزم إضافية (العديد منها مكتوب بلغة C عالية السرعة)، يمكن لبايثون توفير أدوات وأطر لحل التحديات الرياضية والإحصائية والخوارزمية المعقدة.

تتضمن الحزم المعروفة (مجموعة الوحدات) ما يأتي.

- **نامباي (NumPy):** حزمة أساسية للحوسبة العددية في بايثون تستخدمها العديد من اللغات الأخرى. توفر دعمًا للمصفوفات الكبيرة متعددة الأبعاد، إلى جانب مجموعة من الدوال الرياضية للعمل عليها بكفاءة.
- **الباندا (Pandas):** مكتبة أساسية وقوية لتحليل البيانات ومعالجتها لبايثون. إنها توفر هياكل بيانات مثل **DataFrames** و **Series** لتسهيل التعامل مع البيانات المنظمة وتحليلها.
- **ماتبلوتليب (Matplotlib):** مكتبة رسم تستخدم لإنشاء تصورات ثابتة وتفاعلية ومتحركة في بايثون.

مثال عملي

يقوم برنامج بايثون الموضح في الشكل 21.44 بحساب وإخراج متوسط قائمة قيم الأعداد الصحيحة (العدد الصحيح) - وهي عملية إحصائية شائعة. يظهر الناتج في الشكل 21.45.

```
import statistics

numbers = [5, 4, 3, 2, 1]
mean_average = statistics.mean(numbers)
print("Mean Average:", mean_average)
```

الشكل 21.44 استخراج بايثون لحساب متوسط قائمة من خمسة أعداد صحيحة.

Mean Average: 3

الشكل 21.45 توضح مخرجات الكود أن المتوسط الحسابي للأرقام 1، 2، 3، 4 و 5 هو 3.

تستخدم بايثون الرموز للإشارة إلى أنواع مختلفة من هياكل البيانات – على سبيل المثال، يُستخدم الرمز [للإشارة إلى القائمة. ستلاحظ أنه من الضروري استيراد دالة 'mean' من وحدة تسمى 'statistics'. على الرغم من أن هذه الوحدة جزء من مكتبة بايثون القياسية، إلا أنها ليست دالة (مضمنة)، ما يعني أنه يجب استيرادها يدويًا قبل الاستخدام.

يمكن تنزيل مجموعة واسعة من الحزم والوحدات الأخرى باستخدام مدير حزم بايثون 'pip' من مستودع الحزم عبر الإنترنت: pypi.org.

جافا

من المثير للدهشة أن جافا التي تم تطويرها في عام 1995، هي في الواقع أصغر من بايثون، التي ظهرت لأول مرة في عام 1991. منذ ذلك الحين، شهدت كلتا لغتي البرمجة تطورًا كبيرًا. يُنظر إلى جافا على أنها لغة برمجة قائمة على الفئة وموجهة للكائنات (OOP).

نظرًا لتجميع جزء من كود المصدر الخاص به إلى بايت كود جافا، من الناحية النظرية، من الممكن كتابة البرامج مرة واحدة وتشغيلها في أي مكان: وهو مفهوم يسمى WORA. يتم تنفيذ كود جافا فعليًا على آلة جافا الافتراضية (أو JVM) وهي متوفرة لمعظم بنى الحواسيب. هذا يجعل لغة جافا قابلة للنقل بشكل كبير وخيارًا شائعًا للمشاريع المتعلقة بالذكاء الاصطناعي.

على عكس اللغات مثل بايثون التي تكون ذات طابع ديناميكي، فإن جافا ذات طابع قوي وثابت، ما يعني أساسًا أنه بمجرد إعلان المتغير بشكل رسمي، لا يمكن تعيين أي قيم لا تتطابق مع نوع بياناته في أثناء مدة تنفيذه.

مثال عملي

يوضح الشكل 21.46 المشكلة نفسها من مثال بايثون، والتي حُلّت باستخدام لغة جافا. النتيجة موضحة في الشكل 21.47.

```
import java.util.Arrays;
import java.util.List;

public class Main {
    public static void main(String[] args) {
        List<Integer> numbers = Arrays.asList(1, 2, 3, 4, 5);

        double sum = 0;
        for(Integer number : numbers) {
            sum += number;
        }
        double meanAverage = sum / numbers.size();

        System.out.println("Mean Average: " + meanAverage);
    }
}
```

الشكل 21.46 برنامج جافا لحساب متوسط قائمة من خمسة أعداد صحيحة.

Mean Average: 3.0

الشكل 21.47 نتيجة من كود جافا بحسب المتوسط الحسابي للأرقام 1، 2، 3، 4 و 5 على أنه 3.0.

كما ترى في المثال العملي، فإن حلول كتابة الكود في جافا يتطلب بعض الجهد أكثر من بايثون أو ليسب (انظر إلى قسم ليسب، الذي يتبع جزء R، أدناه)، إذ يستخدم فئة، ودالة، وكتل متعددة محاطة بأقواس معقوفة (طريقة لتجميع العبارات موروثة من لغة البرمجة C) وحلقة 'for' محسنة، ولكن هذه مقارنة بسيطة نسبيًا. فهي معقدة إلى حد ما وربما أكثر رهبة للمبتدئين مقارنة بلغة بايثون ما يفسر بالتأكيد نمو بايثون وارتفاع شعبيتها في السنوات العشر الأخيرة.

R

لغة R هي لغة عامة، على عكس بايثون، مُصممة خصيصًا للتعامل مع الحوسبة الإحصائية والرسوم، وهي بارعة تحديدًا في التعامل مع هياكل البيانات المسماة (على سبيل المثال، المصفوفات، والمصفوفات الثنائية، والمتجهات، وإطارات البيانات، والقوائم).

إحدى أعظم مزايا لغة R هي قدرتها على دعم العمليات المتجهة، وهي عمليات تجعل معالجة البيانات وتحليلها أمرًا بسيطًا للغاية.

وتجدر الإشارة إلى أن العديد من وحدات المعالجة المركزية تدعم التعليمات التي يمكنها تطبيق العملية نفسها في وقت واحد على أجزاء متعددة من البيانات (خاصةً قائمة القيم). تعمل هذه العمليات الموجهة على تقليل التكرارات إذ تُشغّل التعليمات بشكل متزامن، ما يؤدي إلى تقليل أوقات المعالجة بشكل كبير (وأكثر كفاءة).

مثال عملي

الشكل 21.48 يوضح أحد مشتقات R يقوم بحساب المتوسط الحسابي بالطريقة نفسها. يظهر الإخراج الناتج في الشكل 21.49. في هذا المثال، تحتوي R على دالة 'mean' مدمجة يمكن استخدامها فورًا في الكود دون الحاجة إلى استيرادها من مكتبة.

```
numbers <- c(1, 2, 3, 4, 5)
mean_average <- mean(numbers)
cat("Mean Average:", mean_average, "\n")
```

الشكل 21.48 أحد مشتقات R الذي يقوم بحساب المتوسط الحسابي لخمسة أرقام.

Mean Average: 3.0

الشكل 21.49 الناتج من برنامج R يوضح أن متوسط 1، 2، 3، 4 و 5 هو 3.

ليسب (Lisp)

ليسب هو اختصار لمعالجة القوائم. إنها واحدة من أقدم لغات البرمجة عالية المستوى (الشبيهة باللغة الطبيعية)، حيث تم تطوير النسخة الأصلية في عام 1958. نتيجة لذلك، كان للغة ليسب تاريخ غني مع الذكاء الاصطناعي وتطوراتها لعدة عقود.

في الواقع، لغة ليسب هي في الحقيقة أكثر من مجرد لغة واحدة؛ فهي في الواقع مجموعة من اللغات. ليسب العادية هي واحدة من أكثر لغاتها شهرة، جنبًا إلى جنب مع كلوجر.

تتمثل إحدى نقاط قوتها الخاصة في قدرتها على إجراء العمليات الحسابية الرمزية، ما يمكنها من التلاعب بهذه الرموز بطريقة يمكن لعلماء الرياضيات تعرفها بسهولة باعتبارها تعبيرات جبرية. إن هذه الخاصية، إلى جانب دعمها للدوال العودية، تجعل اللغة مناسبة بشكل خاص للذكاء الاصطناعي، وإثبات النظريات، ومعالجة اللغة الطبيعية.

تسمح الدالة العودية، وهي ميزة في العديد من لغات البرمجة، للوظائف بالاتصال بنفسها لمعالجة أنواع معينة من الخوارزميات (على سبيل المثال، العوامل ومتتالية فيبوناتشي والبحث الثنائي).

مثال عملي

يوضح الشكل 21.50 برنامج ليسب يقوم بحساب المتوسط الحسابي بالطريقة نفسها. يظهر الناتج في الشكل 21.51.

```
(defun mean-average (numbers)
  (/ (reduce #'+ numbers) (length numbers)))

(setq numbers '(1 2 3 4 5))

(format t "Mean Average: ~A" (mean-average numbers))
```

الشكل 21.50 برنامج ليسب يقوم بحساب المتوسط الحسابي لخمس أرقام.

Mean Average: 3.0

الشكل 21.51 ناتج من ليسب يوضح أن متوسط الأرقام 1، 2، 3، 4 و 5 هو 3.

يمكن أن تكون صياغة لغة ليسب مربكة للوهلة الأولى، نظرًا لاستخدامها الأقواس للإشارة إلى كل من الكود وهياكل البيانات. في هذا المثال، يتم حساب متوسط المتوسط يدويًا (عن طريق حساب مجموع جميع العناصر في القائمة ثم تقسيمها على طولها). يتم تنفيذ العمل الفعلي من خلال دالة (المتوسط) التي تم الإعلان عنها، ويتم استدعاؤها من داخل وظيفة (التنسيق)، والتي توجه متوسط الدالة الذي تم إرجاعه إلى الإخراج القياسي (عادةً، جهاز العرض الخاص بك).

المكتبات والأطر

المكتبات والأطر عبارة عن طرود قابلة لإعادة الاستخدام من أكواد خارجية (عادةً) تساعد مطوري البرامج في المهام الشائعة. وعلى الرغم من استخدام هذين المصطلحين غالبًا بالتبادل، إلا أن هناك تمييزًا دقيقًا بينهما. بشكل عام، فإن الطريقة التي يتم بها دمجها في المشاريع هي التي تساعد على تمييزها.

- باستخدام مكتبة، يمكن لمطور البرامج أو عالم البيانات اختيار عناصر الكود التي يرغبون في استخدامها، كما أن بنية الحل تخصصهم أيضًا.
- باستخدام إطار العمل، يتم توفير الهيكل بشكل عام بواسطة البرنامج ويقوم مطور البرامج أو عالم البيانات بإضافة رمز إليه - الهيكل ليس تصميمهم الخاص.

الجدول 21.10 مكتبة مقابل إطار عمل، يوضح الخصائص الرئيسية لكل نوع

مكتبة	إطار العمل
<ul style="list-style-type: none"> • مجموعة من الفئات، والدوال، والثوابت، وما إلى ذلك. • يقوم المطور بتضمين هذه العناصر بحسب الحاجة • يستدعي المطور هذه العناصر بحسب الحاجة • يتحكم المطور في بنية الحل البرمجي 	<ul style="list-style-type: none"> • يوفر هيكلًا يمكن للمطور بناء مشروعه عليه • يتم ربط كود المطور واستدعاؤه بحسب الحاجة من الإطار • يحدد الإطار إلى حد كبير هيكل المشروع

تشمل المكتبات والأطر المشتركة التي تهتم بأبحاث ومطوري الذكاء الاصطناعي ما يأتي.

Scikit-learn

Scikit-learn هي مكتبة مفتوحة المصدر شهيرة لبايثون مبنية على وحدات بايثون الموجودة مثل NumPy و SciPy و matplotlib. تشتهر ببساطتها وسهولة استخدامها، ولهذا السبب تحظى بشعبية لدى طلاب علوم البيانات. وتدعم مجموعة كبيرة من خوارزميات التعلم الآلي تحت الإشراف والتعلم الآلي دون إشراف وهي خيار مثالي لاستكشاف البيانات وتحليلها. ستري استخدام scikit-learn في كل من الأوساط الأكاديمية والصناعية ببساطة لأنها قوية ومتعددة الاستخدامات.

TensorFlow

TensorFlow هي مكتبة مفتوحة المصدر تم تطويرها بواسطة Google. يتم استخدامها للحساب العددي والتعلم الآلي (ML). وتساعد على بناء شبكات عصبية تكتشف وتفك رموز الأنماط والارتباطات. تُستخدم بشكل أساسي مع بايثون و C++ (للسرعة)، وتدعم واجهة برمجة التطبيقات (API) للغات Go و Java و JavaScript و Swift.

OpenNN

OpenNN تعني الشبكات العصبية المفتوحة (المكتبة). هذه مكتبة برامج تمت كتابتها للغة ++C وهي مصممة لتنفيذ الشبكة العصبية، وهو مفهوم أساسي للتعلم العميق. يتم استخدامها بشكل أساسي لتطوير النماذج التنبؤية وتتمتع بسمعة طيبة لأدائها العالي وتعدد استخداماتها.

OpenCV

OpenCV® تعني رؤية الحاسب مفتوحة المصدر (المكتبة). طورتها شركة Intel في البداية، كما تدعمها الأنظمة الأساسية (على سبيل المثال، من Windows و Linux و macOS X). وتدعم لغات برمجة متعددة بما في ذلك ++C و بايثون و Java، ما يوفر مجموعة شاملة من الخوارزميات والأدوات لرؤية الحاسب والتعلم الآلي.

تم تصميم OpenCV لمساعدة علماء البيانات والمطورين والباحثين على إنشاء تطبيقات يمكنها معالجة البيانات المرئية وفهمها. تشمل تطبيقات OpenCV الاستخدام في أنظمة CCTV (تلفزيون الدائرة المغلقة) للكشف عن الحركات المشبوهة والسيارات ذاتية القيادة وأنظمة تمييز الوجوه.

Weka

Weka تعني بيئة Waikato لتحليل المعرفة وقد تم تطويرها في جامعة Waikato في نيوزيلندا. إنها مكتبة برامج تعلم آلي مفتوحة المصدر مكتوبة بلغة Java. توفر أدوات لتحليل البيانات والنمذجة التنبؤية والتصور. كما أنها أداة تمهيدية جيدة وقادرة على إنشاء نماذج معقدة. تتضمن واجهة المستخدم الرسومية سهولة الاستخدام (GUI) التي تتيح إمكانية إنشاء أنواع نماذج مختلفة دون الحاجة إلى كتابة أي كود. وتسفيد من التوثيق الجيد ودعم المجتمع.

Apache Hadoop

Apache Hadoop هو إطار مفتوح المصدر تم إنشاؤه بواسطة Apache وهو مصمم لتخزين ومعالجة البيانات الضخمة عبر المجموعات. يتمتع بسمعة طيبة بسبب قابليته للتطوير والمرونة في مواجهة الأعطال. يستخدم أباتشي هادوب HDFS - نظام الملفات الموزع Hadoop. وهو نظام تخزين قابل للتطوير ويتحمل الأخطاء بدرجة كبيرة ويمكنه تقسيم البيانات عبر أجهزة متعددة. كما يستخدم نموذج MapReduce الشهير الذي يعالج مجموعات البيانات الكبيرة باستخدام خوارزمية موزعة. يقوم نظام آخر يُعرف باسم YARN (مفاوض موارد آخر) بجدولة وإدارة الموارد في المجموعة.

المعالجة العنقودية (Cluster Processing)

تعد المعالجة العنقودية (باستخدام شبكة من أجهزة الحاسب أو العقد المتصلة) مفيدة بشكل خاص لتطبيقات الذكاء الاصطناعي والتطوير. إنه نهج يمكن توسيع نطاقه أو تقليله بسرعة اعتمادًا على الأحمال الحالية، دون الحاجة إلى تعديل الأكواد أو النماذج أو البيانات الأساسية. توزيع المهام يعني أيضًا أن هذا النهج

متسامح جداً مع الأخطاء. هذا لأن فشل العقد الفردية ليس كارثياً، حيث يمكن للأخريين متابعة المعالجة دون توقف أو فقدان الخدمة.

استخدام الأدوات التي توفرها بيئات التطوير المتكاملة وأدوات تطوير الأكواد ذات الصلة

عند تطوير أنظمة الذكاء الاصطناعي، يتعين على علماء البيانات استكشاف البيانات وإجراء العديد من التجارب المختلفة لإثبات أفكارهم ومفاهيمهم.

على الرغم من أنه يمكن استخدام بيئات التطوير المتكاملة التقليدية (على سبيل المثال، PyCharm و Microsoft VS Code وما إلى ذلك)، إلا أنها ليست أفضل الخيارات لتتبع التقدم أو إنشاء سجل سردي للخطوات المتخذة. هذا هو المكان الذي تثبت فيه تطبيقات مثل JupyterLab و Jupyter Notebook أنها أصول قيمة.

Jupyter Notebook هو تطبيق قائم على الويب يسمح لعالم البيانات بإنشاء ومشاركة المستندات التي يمكن أن تحتوي على تجارب تعلم الآلة القيمة. يمكن أن يحتوي كل دفتر ملاحظات على مزيج من الخلايا. تحتوي كل خلية من هذه الخلايا على ما يأتي.

- سرد وملاحظات مكتوبة بلغة Markdown. Markdown هي لغة ترميز خفيفة الوزن (دون علامات) تُستخدم لتنسيق مستندات النص العادي، وغالباً ما تستخدم لملفات (README) والوثائق والمدونات وما إلى ذلك. يوضح الشكل 21.53 لغة Markdown ومخرجاتها. كما ترى، فإن تعلمها أسهل بكثير من تعلم HTML.
- كود البرنامج القابل للتنفيذ (على سبيل المثال، بايثون، R، وما إلى ذلك).
- مخرجات تفاعلية في شكل صور (على سبيل المثال، مخططات ورسوم بيانية)، ومقاطع فيديو، و LaTeX (نظام تنضيد يستخدم لإنشاء المستندات، وهو شائع لدى الأشخاص الذين يقومون بالبحث العلمي أو الأكاديمي أو الكتابة الفنية أو النشر)، و HTML المخصص (لغة ترميز النص التشعبي) - لغة تستخدم لإنشاء محتوى صفحة الويب.

يمكن النقر فوق كل خلية ثم تحريرها أو حذفها أو حتى إعادة تسلسلها. بمجرد إعادة تشغيل الخلية، يتم تحديث الخلية الناتجة (على سبيل المثال، إخراج قيمة متغيرة أو إعادة رسم المخطط الناتج).

في أي نقطة، يمكن نقل الخلايا لأعلى أو لأسفل دفتر الملاحظات، وتغيير الترتيب الفعلي لتنفيذ الأكواد. ويمكن حفظ دفاتر الملاحظات هذه محلياً أو وضعها في VCS (أنظمة التحكم في الإصدار) للتعاون أو حتى مشاركتها عبر الإنترنت عبر البريد الإلكتروني والمنشديات وما إلى ذلك.

التنسيق الأساسي لملفات "ipynb" الناتجة هو JSON، لذا فهي قابلة للنقل بشكل كبير عبر الأنظمة والتطبيقات المختلفة.

هذا المستوى من التفاعل والوصول الفوري إلى النتائج والقدرة على تعديل الكود (وترتيبه) يجعل Jupyter Notebook أداة شائعة بشكل استثنائي ومستخدمة عالمياً لبدء المشاريع المتعلقة بالذكاء الاصطناعي.

Load the Iris dataset from scikit-learn



الشكل 21.52 تحميل مجموعة بيانات زهرة السوسن

مشروع الذكاء الاصطناعي الخاص بي

صُمم هذا المشروع لتقديم رؤى حول....

بدء النشاط

تثبيت المشروع سهل...

ثلاثة أشياء تحتاج إلى معرفتها

- هذا المشروع مكتوب بلغة بايثون 3
- هذا المشروع موثق جيدًا
- لا تتردد في تقديم الملاحظات

مشروع الذكاء الاصطناعي الخاص بي

صُمم هذا المشروع لتقديم رؤى حول....

بدء النشاط

تثبيت المشاريع *سهل*...

ثلاثة أشياء تحتاج إلى معرفتها

* المشروع مكتوب بلغة *بايثون 3*

* المشروع موثق جيدًا

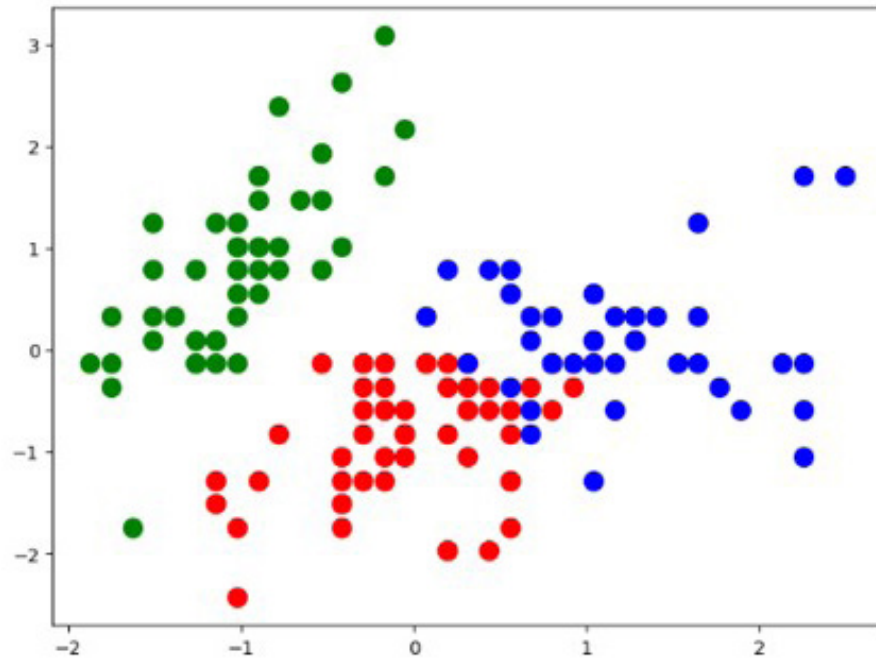
* لا تتردد في تقديم الملاحظات

الشكل 21.53 لغة Markdown ومخرجاتها المقدمة (مع الكود على اليمين والنتائج على اليسار).

```
[20]: iris = datasets.load_iris()
X = iris.data
y = iris.target
```



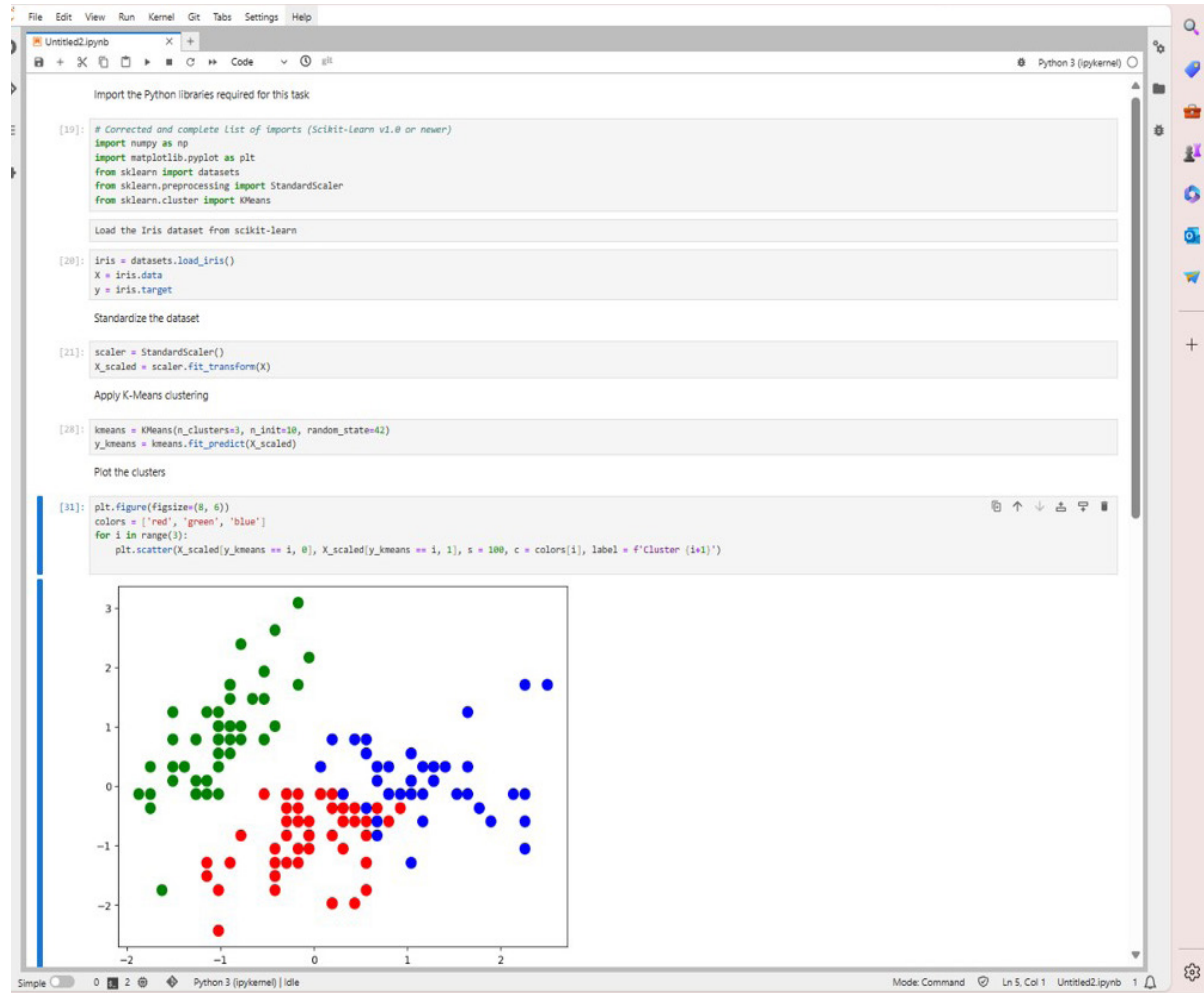
الشكل 21.54 كود قابل للتنفيذ.



الشكل 21.55 الناتج.

مثال عملي

يوضح المثال الآتي استخدام دفتر Jupyter لاستكشاف تصنيف زهرة السوسن باستخدام خوارزمية التجميع K-means الخاصة بلغة scikit-learn.



الشكل 21.56 استخدام خوارزمية التجميع K-means في مكتبة scikit-learn لعرض الكود والمخرجات.

إبراز الأكواد البرمجية والقواعد النحوية

على الرغم من أن أدوات تطوير الأكواد القديمة كانت أحادية اللون بشكل صارم (تُعرض بصريًا باللونين الأسود والأبيض فقط)، فإن بيئات التطوير الحديثة (بما في ذلك Jupyter Notebook و Microsoft VS Code، وما إلى ذلك) ومحركات النصوص الشهيرة (على سبيل المثال: Nano، وما إلى ذلك) تدعم جميعًا تمييز بناء الجملة القابل للتخصيص. هناك العديد من الأسباب الجيدة لاستخدام تمييز الصيغة.

- 1 يساعد المستخدم على التنقل في الكود.
- 2 يعمل على تحسين قابلية قراءة الكود.
- 3 يلبي احتياجات المستخدمين ذوي الإعاقات البصرية.
- 4 يحسن فهم الكود.
- 5 يساعد على تحديد الأخطاء في بناء الجملة أو المنطق الخاص بالكود.
- 6 يقلل إجهاد العين.

المصطلح الرئيس

تمييز بناء الجملة – قدرة الأداة على تحليل لغة البرمجة وتمثيل الكلمات الرئيسية، والدوال، والعوامل، والرموز ذات الألوان المختلفة على الشاشة.

مثال عملي

يظهر الكود في الشكل 21.57 إبراز كود للغة بايثون. الشكل 21.58 يوضح ما تشير إليه الألوان المختلفة.

```
# ارسم المجموعات
plt.figure(figsize=(8, 6))
colors = ['red', 'green', 'blue']
for i in range(3):
    plt.scatter(X_scaled[y_kmeans == i, 0], X_scaled[y_kmeans == i, 1], s =
100, c = colors[i], label = 'Cluster {i+1}')
```

الشكل 21.57 إبراز كود بلغة بايثون، تم إنجازه أوتوماتيكياً من خلال البرنامج المستخدم.

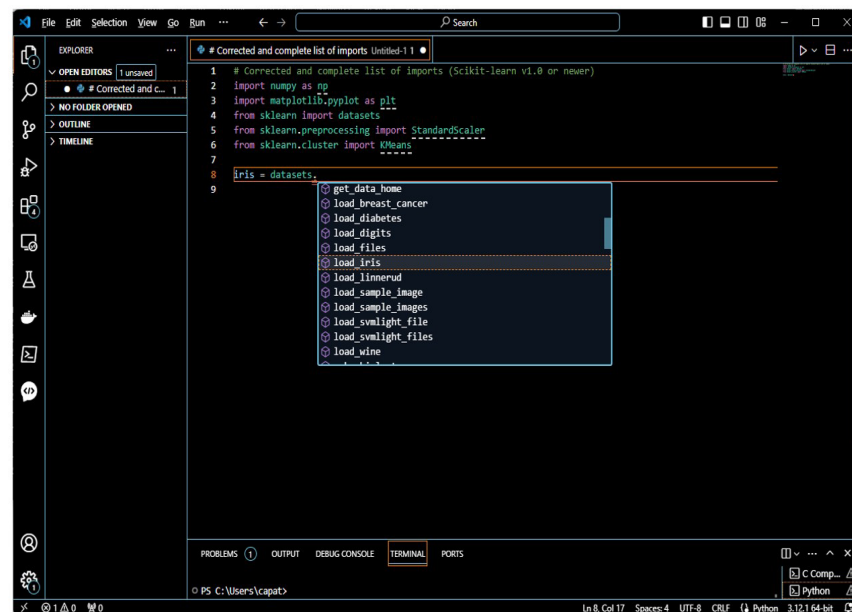
أخضر	تعليق بايثون
أزرق	متغير أو مرجع بايثون
أصفر	دالة بايثون المدمجة
برتقالي	نصوص/سلاسل بايثون
أحمر	الكلمات المفتاحية في بايثون
وردي	الأعداد الصحيحة في بايثون
أبيض	الكائنات والأساليب في بايثون
أزرق داكن	الأحرف الخاصة في بايثون

الشكل 21.58 الألوان محددة في المثال أعلاه.

تجدر الإشارة إلى أن اللغات المختلفة تستخدم مفاتيح تلوين مختلفة وحتى خيارات لغة بايثون الموضحة في الشكل قد يتم تعديلها لتناسب التفضيلات الشخصية لكل متعلم.

تنفيذ الأكواد البرمجية

ميزة مفيدة في بيئات التطوير المتكاملة الحديثة (مثل PyCharm و Microsoft VS Code وما إلى ذلك) هي مفهوم التنفيذ التلقائي للكود.



الشكل 21.59 يوضح كود Microsoft VS ميزة تنفيذ الأكواد البرمجية الخاصة به.

هذه ميزة تقدم بشكل أساسي اقتراحات للمطور حول ما يمكنه كتابته بعد ذلك لإكمال سطر من كود البرنامج. وتقوم بذلك من خلال النظر إلى عنصر الكود الذي يتم التركيز عليه. ثم تعرض قائمة بالخيارات الممكنة التي يمكن للمطور الاختيار من بينها. يقلل بوضوح هذا النوع من الميزات العدد المحتمل للأخطاء التي يمكن أن يرتكبها المطور، ونتيجة لذلك، يجعل إنشاء الكود أسرع (ومضموناً).

قد تحتوي بعض بيانات التطوير المتكاملة (IDEs) أيضاً على وظائف محسنة توفرها الإضافات القابلة للتثبيت (مثل عميل متصفح الويب مثل Microsoft Edge أو Google Chrome). النوع الشائع هو رمز (co-pilot)، وهو ملحق يستخدم الذكاء الاصطناعي لتقديم اقتراحات حول كيفية حل مهام البرمجة الشائعة. كما هو الحال مع جميع المحتويات التي يتم إنشاؤها بواسطة الذكاء الاصطناعي، يجب فحص المحتوى بدقة قبل الاستخدام، حيث يمكن أن تختلف جودة النتائج بين الجيدة وغير القابلة للعمل بشكل عام (أو تحتوي على أخطاء).

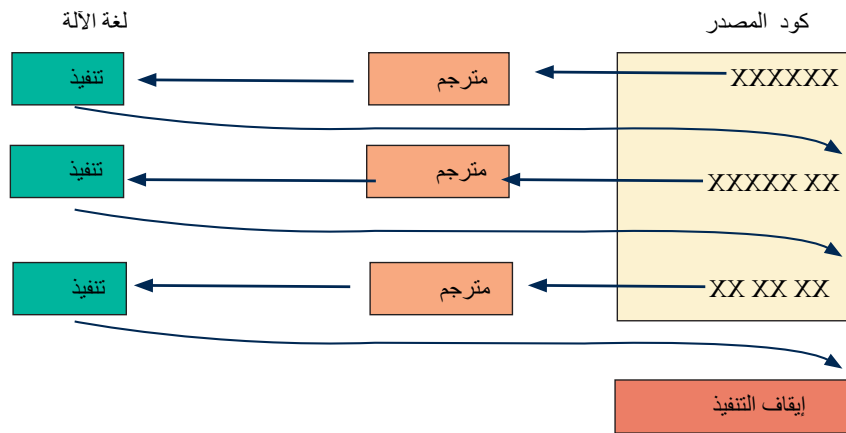
المترجم الشفوي / المترجم الإلكتروني

في النهاية، يمكن لوحدة المعالجة المركزية (CPU) تنفيذ تلك الأكواد الآلية الثنائية فقط التي تفهمها تماماً. يجب أولاً ترجمة أي كود برنامج تكتبه، سواء كان بايثون أو R أو جافا، وما إلى ذلك، إلى كود الجهاز قبل تشغيله.

هناك أشكال مختلفة للترجمة الممكنة. ربما يكون الشكل الأكثر شيوعاً للترجمة هما التفسير والتجميع (كما هو موضح أدناه). لكل منها مزاياها وعيوبها. قد تستخدم بعض اللغات، في الواقع، مزيجاً من كلا الأسلوبين.

التفسير

المترجم هو ببساطة برنامج للمساعدة يأخذ رمز المصدر (كود البرنامج المكتوب) سطرًا واحدًا في كل مرة، ويترجمه إلى لغة الآلة وينفذه.



الشكل 21.60 تفسير كود البرنامج 'المصدر' الذي يوضح كيفية تفسير سطور الكود بشكل فردي قبل التنفيذ.

إذا واجه سطرًا لا يمكنه تفسيره بنجاح، فإنه (يتعطل) ويبلغ عن خطأ في بناء الجملة.

وإلا، تستمر العملية حتى يتم تفسير جميع الأسطر وتنفيذها بنجاح.

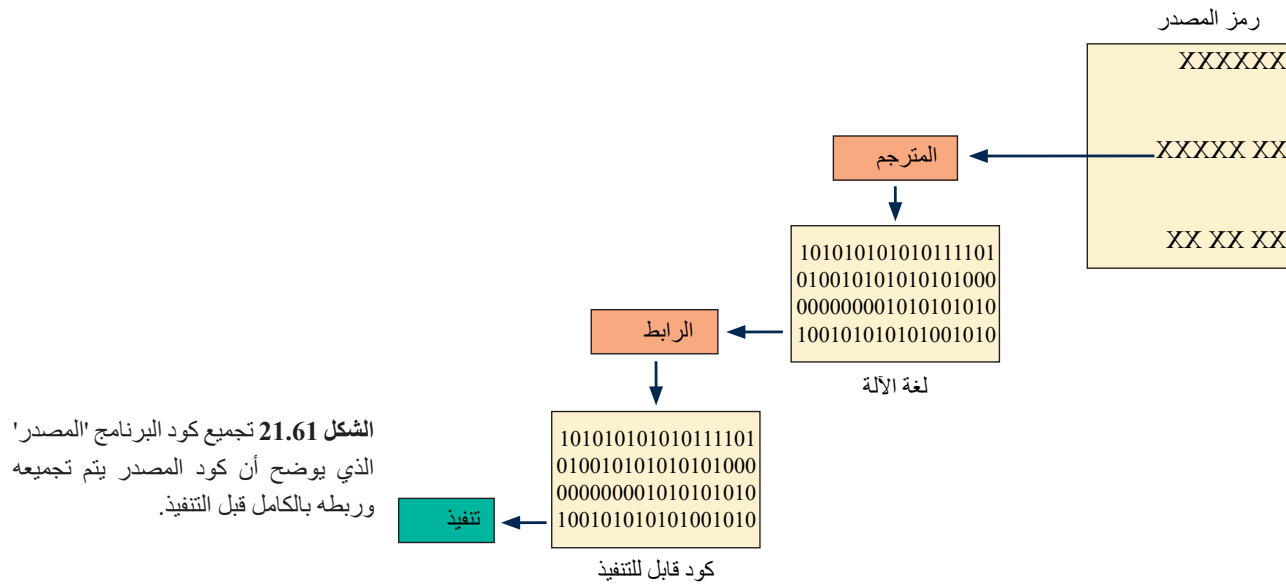
غالبًا ما تكون عملية التفسير بطيئة (نسبيًا) ولديها (مقارنة بالتجميع) فرصة ضئيلة للتحسين (أي لاتخاذ أي اختصارات لتوفير الوقت تعمل على تحسين وقت التنفيذ أو كفاءة الذاكرة).

يُعتبر بايثون، ليسب R لغات مفسرة بشكل عام، على الرغم من أن الأولى تُترجم فعليًا إلى بايت كود بايثون أولاً (في المرحلة الوسيطة) قبل أن ينفذها مفسر بايثون – ما يسمح بإدخال مستوى من التحسين الأساسي.

تجميع

في التجميع، يتم استخدام زوج من الأدوات – المترجم والموصل – لترجمة كود البرنامج إلى لغة الآلة لوحدة المعالجة المركزية المستهدفة وإنشاء ملف تنفيذي مستقل.

الميزة الأساسية لهذه التقنية هي أن المترجم يمكنه 'رؤية' ومعالجة رمز المصدر للبرنامج بأكمله دفعةً واحدة بدلاً من سطر بسطر. ونتيجةً لذلك، فلديها فرصة لتحقيق تحسين كبير في الكود الناتج، بمجرد إنتاج الكود القابل للتنفيذ، يمكن استخدامه وتوزيعه على الآخرين دون الكشف عن كود المصدر الأصلي.



الشكل 21.61 تجميع كود البرنامج 'المصدر' الذي يوضح أن كود المصدر يتم تجميعه وربطه بالكامل قبل التنفيذ.

المهارات

المهارات المعرفية: العمليات
والإستراتيجيات المعرفية:

- حل المشكلات
- مهارات التواصل الشخصي: أخلاقيات العمل/الضمير:
- المسؤولية
- الإنتاجية

المصطلح الرئيس

تصحيح الأخطاء – عملية تحديد الأخطاء أو العيوب (غالبًا ما يشار إليها بالأخطاء) في رمز مصدر البرنامج وإصلاحها بنجاح.

تجدر الإشارة إلى أنه في حالة مواجهة أي أخطاء في بناء الجملة في أثناء الترجمة، تتوقف العملية (دون إنتاج ملف تنفيذي) ويجب تصحيح الرموز البرمجية بنجاح قبل محاولة إعادة التحويل البرمجي. قارن هذا بالكود المفسر الذي ينفذ (سطرًا بسطر) حتى يحدث خطأ.

اللغات المترجمة الشائعة هي فورتران وجافا، على الرغم من أنه في حالة جافا يتم ترجمة الكود إلى جافا بايت كود (مجموعة أخرى من التعليمات المحايدة للمعالج) ثم يتم تفسيرها بواسطة آلة جافا الافتراضية (JVM) التي تعمل على جهاز الحاسب المستهدف.

وتجدر الإشارة أيضًا إلى أن بعض لغات البرمجة قد تتم مواجهتها في أشكال مفسرة أو مجمعة. ليسب مثال جيد على ذلك.

تصحيح الأخطاء

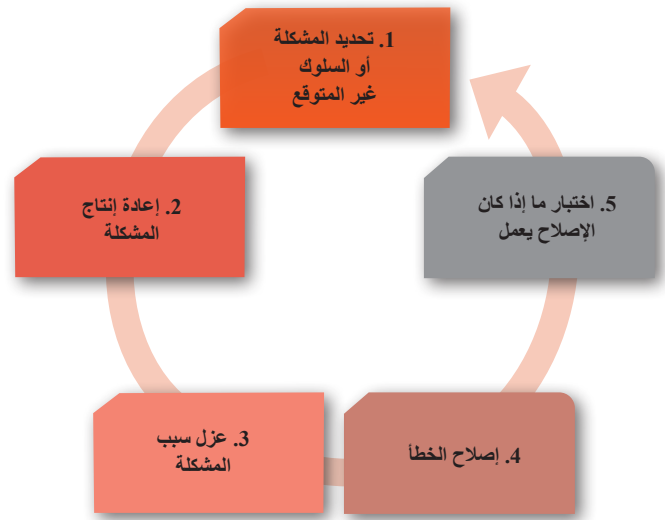
غالبًا ما يكون العثور على "الأخطاء" وتصحيحها (تصحيح الأخطاء) أمرًا صعبًا في البداية، لكنه عادة ما يصبح أسهل بمجرد أن تصبح لغة البرمجة المستخدمة مألوفة. قد يصعب للغاية تحديد موقع بعض الأخطاء وربما يصعب حلها، فغالبًا ما تتطلب هذه العملية قدرًا كبيرًا من الصبر والانضباط والاهتمام بالتفاصيل، ويمكن أن تعود وجهة النظر المختلفة التي يقدمها الزملاء بفائدة كبيرة في تحديد الخطأ وإصلاحه. تصحيح الأخطاء هو في الأساس عملية من خمس مراحل (كما هو موضح في الشكل 21.62).

حتى بعد إجراء الإصلاحات بنجاح، قد تستمر المشكلات الجديدة في الظهور بمرور الوقت. من الممكن أيضًا أن تؤدي أي إضافات جديدة إلى الكود المصدري إلى حدوث أخطاء (جديدة). هذا يعني أن عملية التصحيح بأكملها يمكن أن تدور بشكل حتمي.

غالبًا ما تحتوي بيانات التطوير المتكاملة الحديثة على مجموعة شاملة من أدوات التصحيح. تشمل الأمثلة الشائعة ما يأتي.

- التتبع – يسمح للمطورين بمتابعة تنفيذ الكود الخاص بهم (ببطء ومنهجية) سطرًا بسطر، ومراقبة أي الأسطر من الكود يتم تنفيذها فعليًا.
- المراقبة/الفحص – يسمح للمطورين بمراقبة تغير القيم في المتغيرات في أثناء التنفيذ.
- نقاط التوقف – تتيح للمطورين تحديد الأماكن التي يجب أن يتوقف فيها الكود عن التنفيذ العادي حتى يمكن تتبع الخط سطرًا بسطر وفحص المتغيرات.

كما هو الحال في جميع المشاريع الموجهة نحو البرمجيات، من الضروري لمشاريع الذكاء الاصطناعي ألا تؤثر الأخطاء أبدًا في مراحل التدريب أو الاختبار أو الإنتاج.



الشكل 21.62 دورة التصحيح التي توضح عملية الخطوات الخمس.

يتم تجميع بعض أخطاء الأكواد البرمجية ولكنها تعطي مخرجات غير صحيحة. لماذا يحدث ذلك؟

وقفة للتفكير

تلميح فكر في سطور الأكواد البرمجية حيث توجد التحديدات والتكرارات.

توسيع الأفق ما الذي يجب فعله لاكتشاف هذه الأخطاء؟ اكتب خمس اقتراحات.

منصة إدارة الأكواد

مشاريع الذكاء الاصطناعي كبيرة جدًا بحيث لا يمكن لمطور واحد فقط العمل عليها. للسماح للعديد من المطورين بالمساهمة في مشروع، من المحتم استخدام نوع ما من منصات إدارة الكود.

تسمح منصة إدارة الرموز البرمجية لفريق من المطورين بما يأتي:

- تخزين قاعدة الشفرة الخاصة بهم بطريقة آمنة ومضمونة
- إدارة وتأمين قاعدة الأكواد البرمجية الخاصة بهم (أو المستودع)
- التعاون بطريقة منهجية مع قاعدة البيانات
- تتبع جميع التغييرات التي تم إجراؤها على قاعدة الأكواد البرمجية، استنادًا إلى التاريخ/الوقت والغرض والتأليف (توفير شكل من أشكال التحكم في الإصدار)
- دعم فروع متعددة من التطوير الموازي والتجريبي
- مراقبة المشكلات والأخطاء وطلبات الميزات – ما يسمح للفرق بتنظيم وترتيب أولويات أعباء عملهم في مدة زمنية معينة (أو سابق).

منصة إدارة الأكواد الشهيرة هي **Git**، وحتى وقت كتابة هذا الكتاب، ما تزال **Git** تُعتبر نظام التحكم في الإصدارات (VCS) الأكثر شهرة في العالم.

يتوفر واجهات وخدمات ويب إضافية لـ **Git** عن طريق منصات مثل: **GitLab** و **GitHub** وما إلى ذلك، وتسمح هذه الأمور بتنصيب **Git** محليًا على أجهزة المطورين. وهذا بدوره يسمح للمطورين بدفع (الترامات) الكود إلى مستودع بعيد لقاعدة الأكواد البرمجية الخاصة بهم وسحب التغييرات منه.

يتم تقسيم أي مشكلة في **Git** إلى ثلاث مناطق منفصلة:

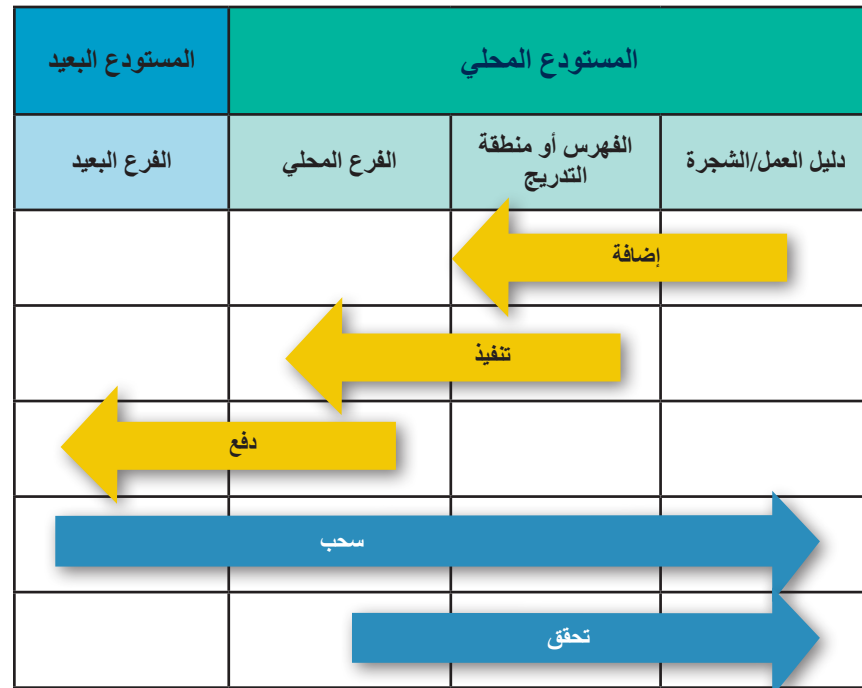
- دليل العمل/الشجرة
- الفهرس أو منطقة التدرج
- مستودع **Git**.

توجد هذه المناطق عادةً على حاسوب المطور.

المصطلح الرئيس

Git – نظام تحكم في الإصدارات (VCS) أنشاه لينوس تورفالدس (مبتكر نواة لينكس) في عام 2005 وصدر بوصفه أحد البرامج مفتوحة المصدر بموجب رخصة **GNU GPL v2**.

عندما يبدأ المطور، يعتبر دليل العمل (نظيفًا). عندما يقوم المطور بإضافة الأصول أو حذفها أو تعديلها، تتغير حالة الشجرة لدليل العمل. يتم تعقب هذه التغييرات (يعرف Git عنها) أو يتم إلغاء تتبعها (لا يعرف Git عنها).



الشكل 21.63 عرض لوظائف Git الأساسية عبر مناطق مختلفة من المستودعات، سواء كانت محلية أو بعيدة.

يؤدي استخدام الأمر 'git add' إلى نقل الملف المعدل إلى منطقة التدرج. يتم استخدام الأمر (git commit) لتنفيذ التغييرات المتعقبة على مستودع Git المحلي. من أفضل الممارسات تضمين تعليق وصفي مناسب إلى جانب الالتزام الذي يتعلق بالغرض منه.

بمجرد إجراء التغييرات على المستودع المحلي، يمكن استخدام أمر (git push) آخر لمشاركة/تحميل التغييرات إلى المستودع البعيد. بمجرد قبول تغييرات المطور، يمكن للآخرين جلب هذه التغييرات أو سحبها إلى أجهزتهم المحلية، مع تحديث نسخهم من قاعدة الأكواد البرمجية.

تعارضات الدمج شائعة. تحدث هذه عادةً عندما يقوم العديد من المطورين بإجراء تغييرات متناقضة على الأجزاء نفسها من قاعدة الأكواد البرمجية. عادةً ما يقوم مالك قاعدة الأكواد البرمجية بحل هذه المشكلات. ومع ذلك، لتجنبها تمامًا، يُنصح المطورون بالالتزام (قليلاً وفي كثير من الأحيان).

قد تحتوي المشاريع أيضًا على فروع متعددة – اعتبرها بمنزلة تفرعات مختلفة (غالبًا ما تكون تجريبية) للفرع "الأساسي" للتطوير. قد يتم دمج هذه الفروع لاحقًا أو تجاهلها، اعتمادًا على نجاحها.

فكر ملياً

فكر في برامج الحاسب التي كتبتها، ثم قمت بتصحيحها ثم ربما أعدت كتابتها. هل سبق لك أن أردت العودة إلى إصدار سابق لأن شيئاً ما كان يعمل لم يعد يعمل؟ هل تحفظ برامجك المعدلة باسم جديد، أم بالاسم ورقم الإصدار نفسه (مثل V2، أو 2.1)؟ عندما تعمل في المجال، من المتوقع أن تتمكن من استعادة الإصدارات السابقة من الأكواد البرمجية التي كتبتها. ابدأ التفكير الآن في كيفية استخدام التحكم في الإصدارات في أنشطة البرمجة المستقبلية.

الحوسبة السحابية لتوفير موارد قابلة للتطوير لتخزين ومعالجة للبيانات

يمكن إجراء مشروع الذكاء الاصطناعي وتجارب التعلم الآلي الأساسية بنجاح داخل الحدود المادية للمنظمة (أي 'في الموقع'). ومع ذلك، فمن الشائع أن تبحث المؤسسات عن بيئات عبر الإنترنت يمكن فيها بناء أنظمة الذكاء الاصطناعي الخاصة بها واختبارها ونشرها. هذا يسمح لهم بالاستفادة من منصات الحوسبة السحابية ذات التكلفة الإجمالية للملكية (TCO) المنخفضة والمرنة مثل:

- خدمات أمازون ويب (AWS)
- منصة جوجل السحابية (GCP)
- Microsoft Azure®.

الميزة الأساسية التي توفرها هذه المنصات هي قابلية التوسع السريع – أفقيًا (إضافة المزيد من الحوسبة) وعموديًا (حوسبة ذات موارد أفضل، أي وحدة معالجة مركزية/وحدة معالجة رسومات أسرع، والمزيد من ذاكرة الوصول العشوائي، وما إلى ذلك). تتناسب هذه القابلية للتوسع مع الطبيعة الديناميكية لطلبات خدمات الذكاء الاصطناعي المتزايدة.

بالإضافة إلى ذلك، يوفر معظم موفري السحابة أدوات وخدمات محددة للذكاء الاصطناعي والتعلم الآلي يمكن استخدامها معًا لتصميم وبناء ونشر حلول كاملة.

Microsoft's Azure AI

هذه مجموعة من الخدمات المعتمدة على السحابة والتي تتضمن ما يأتي.

- تضيف خدمات Azure المعرفية الرؤية واللغة والكلام واتخاذ القرار إلى التطبيقات.
- يتم استخدام التعلم الآلي من Azure لتطوير نماذج التعلم الآلي وتدريبها ونشرها في نطاق واسع.
- تُستخدم خدمات الروبوتات في Azure لبناء ونشر وإدارة مساعدات روبوتية ذكية يمكنها التفاعل مع المستخدمين عبر قنوات اتصال مختلفة (على سبيل المثال، الدردشة، البريد الإلكتروني، وما إلى ذلك).

منصة جوجل للذكاء الاصطناعي

هذه مجموعة احترافية من الأدوات لتطوير التعلم الآلي ونشره وتتضمن ما يأتي.

- يتم استخدام دفاتر منصة الذكاء الاصطناعي مع مكتبات وأدوات التعلم الآلي المتكاملة.
- تدريب منصة الذكاء الاصطناعي هي خدمة تساعد في تدريب نماذج التعلم الآلي في نطاق واسع.
- تنبؤات منصة الذكاء الاصطناعي هي خدمة مُدارة لنشر وتقديم نماذج التعلم الآلي المختلفة. لديها القدرة على تتبع ومراقبة النماذج بدقة.
- حلول الذكاء الاصطناعي المبنية مسبقًا هي نماذج أو أطر أو تطبيقات جاهزة للاستخدام مقدمة من Google.

أمازون سيج ميكرو

هذه خدمة AWS مُدارة بالكامل وتوفر مجموعة الأدوات الشاملة الآتية.

- يتم استخدام مثيلات Jupyter Notebook مع مكتبات وأدوات التعلم الآلي المتكاملة.
- الخوارزميات المحسنة المضمنة هي مجموعة من الخوارزميات المضمنة والمحسنة للغاية التي تحل مشكلات التعلم الآلي الشائعة.
- يوفر تدريب النماذج تدريبًا فعالًا ويضبط النماذج، ما يبسط خطوات العمل مع الخوارزميات ومجموعات البيانات المختلفة.
- يوفر نشر النموذج نقلًا سهلًا للنماذج المدربة إلى بيئة الإنتاج (أي بيئة 'حية' ومواجهة للعملاء)، ويدعم إصدارات مختلفة، وخيارات التوسع التلقائي، وما إلى ذلك.
- توفر مراقبة النماذج وإدارتها أدوات تمنح المطورين القدرة على تتبع المقاييس الرئيسية مثل الدقة والانحراف.

الأهم من ذلك، أن جميع هذه الأدوات والخدمات تتكامل بشكل جيد مع خدمات السحابة الأخرى مثل تخزين البيانات، استضافة الويب، الأمان، التسجيل والمراقبة. وبعبارة أخرى، فإنها توفر حزمة ذكاء اصطناعي كاملة جدًا تسمح بإستراتيجية "الدفع أولاً بأول" (أي أن العميل يدفع فقط مقابل الحوسبة والخدمات التي يستخدمها فعليًا، مع توسيع نطاقها وتقليصها بحسب الحاجة، بدلاً من دفع مبلغ ثابت مقدماً).

المصطلح الرئيس

الحوسبة – هي مصطلح يستخدمه مقدمو السحابة للإشارة إلى الموارد والإمكانات اللازمة لمعالجة المهام أو إدارة أعباء العمل في بيئة تعتمد على السحابة، وعادةً ما يتضمن قوة المعالجة (وحدة المعالجة المركزية)، والذاكرة (ذاكرة الوصول العشوائي)، وتخزين القرص وقدرات الشبكة.

C.P.5, C.M3, BC.D2

تمرين تقييمي 21.3

أنت تعمل لصالح مدير مشروع الذكاء الاصطناعي الذي طلب منك تطوير حل ذكاء اصطناعي لتلبية احتياجات العميل المحددة.

باستخدام البيانات التي تم الحصول عليها في دراسة الحالة والتي أعدت في التمرين التقييمي 21.2، نفذ المهمة الآتية لتطوير حل الذكاء الاصطناعي:

استفسرت سلطات إدارة المطار عما إذا كان من الممكن تحليل البيانات وتدريب نموذج الذكاء الاصطناعي بنجاح بحيث يمكن التنبؤ بالتأخيرات المحتملة في المستقبل.

ستحتاج إلى:

- تطوير حل الذكاء الاصطناعي باستخدام لغة برمجة وأدوات حوسبة مناسبة
- اختبار حل الذكاء الاصطناعي وتحسينه
- تقييم فعالية حل الذكاء الاصطناعي

التخطيط

- ضع خطة للمهمة. ضع قائمة بكل ما ستحتاج إلى القيام به وضع جدول زمني لموعد إنجاز العمل.

التنفيذ

- تصميم حل الذكاء الاصطناعي وتنفيذه.
- اختبار الحل وتحسينه.
- اعرض حلك على معلم أو متعلم آخر.

المراجعة

- علق على الحل بشكل مناسب.
- قم بإنشاء دليل قصير يشرح كيفية عمل الحل. قم بتضمين نسخة من الكود.



فكر في المستقبل

أزها خريجة تعمل كعالمة بيانات متدربة في قطاع الصيدلة. وهي متخصصة في مشاريع الذكاء الاصطناعي التي تهدف إلى دفع الابتكار وتحسين نتائج المرضى ورفع الكفاءة العامة عبر المراحل المختلفة لتطوير الأدوية وتحليل السوق.

إنها تحقق ذلك من خلال تنظيم وتحليل مجموعات كبيرة من بيانات السجلات الصحية للمرضى المجهولين ونتائج التجارب الدوائية والبيانات الجينومية. الهدف من هذا العمل هو تحديد الأنماط والاتجاهات والرؤى التي يمكن أن توجه عمليات صنع القرار في شركتها في تطوير الأدوية والطب الشخصي.

تتطلب وظيفتها مزيجاً من علوم البيانات والبيولوجيا، إلى جانب فهم قوي للرياضيات والإحصاء. كما هو متوقع، لديها أيضاً فهم جيد لقطاعها ووعي بالمعايير التنظيمية والتشريعات المحلية التي تحكمه.

تستمتع بتعلم التقنيات الخوارزمية الجديدة، وعلى الرغم من أنها ليست مطور برامج، يمكنها إنشاء تجارب وتصورات في بايثون باستخدام أدواتها المفضلة: دفتر جويتر.

عندما تحتاج إلى مساعدة لنشر خدمات الذكاء الاصطناعي، فإنها تتعاون مع زملائها في قسم عمليات التطوير. ونتيجة لذلك، اهتمت بشدة بالحوسبة السحابية ومنتجاتها وخدماتها المتاحة.

خلال الأشهر الاثني عشر الماضية، عملت على العديد من المشاريع المثيرة، بما في ذلك مهمة مبتكرة تهدف إلى استخدام خوارزميات التعلم الآلي لتصنيف المرضى بناءً على علاماتهم الجينية وخصائص المرض. تم القيام بذلك لتحسين الكفاءة والفعالية السريرية لتجربة دواء جديد.

تركيز مهاراتك

العمل متدرجاً في علوم البيانات، لا سيما في الذكاء الاصطناعي، هو دور وظيفي رائع، خاصةً أنك ستعمل على الأرجح على مشاريع ذات تأثير إيجابي في من حولك.

ستكون بالتأكيد شخصاً يستمتع بالعمل مع الأرقام والإحصاءات وستكون أحد هؤلاء الأشخاص الذين يهتمون دائماً بالتفاصيل للتأكد من أن نتائجك هي أفضل ما يمكن.

مهارات البرمجة – ستحتاج إلى أن التحلي بمهارات تمنحك الثقة عند التعامل مع عدد من لغات البرمجة وأطر الذكاء الاصطناعي وعند العمل مع تدفقات البيانات الديناميكية في الوقت الفعلي.

الاختبار - يجب عليك التأكد من أن لديك مهارات اختبار جيدة حتى تتمكن من اختبار النماذج وإجراء تحسينات على الخوارزميات بحسب الحاجة، بناءً على نتائج الاختبار الخاصة بك.

مهارات البيانات – بصفتك متدرجاً في علم البيانات، ستقوم في البداية بدعم الموظفين الأكثر خبرة، وتنظيف البيانات ومعالجتها مسبقاً وإجراء بعض التحليلات الأولية. كل هذه المهارات ضرورية إذا كنت ترغب في التقدم في هذا المجال.

التواصل – يحتاج علماء البيانات إلى التحلي بالبراعة في التواصل، لأن شرح نتائج التحليل قد يكون صعباً، فعند العمل في هذا الدور، يجب أن تتمتع بالقدرة على استخدام مجموعة هائلة من المخططات والرسوم البيانية للمساعدة على دعم نتائجك. ستكون هذه القدرة ضرورية عند مشاركة البيانات مع أشخاص لديهم فهم أقل للتحليلات، ويجب أن تمارس مهارات الاتصال المكتوبة والشفوية، لا سيما للعروض التقديمية (لأن ذلك سيحسن ثقتك وفعاليتك).