ORACLE

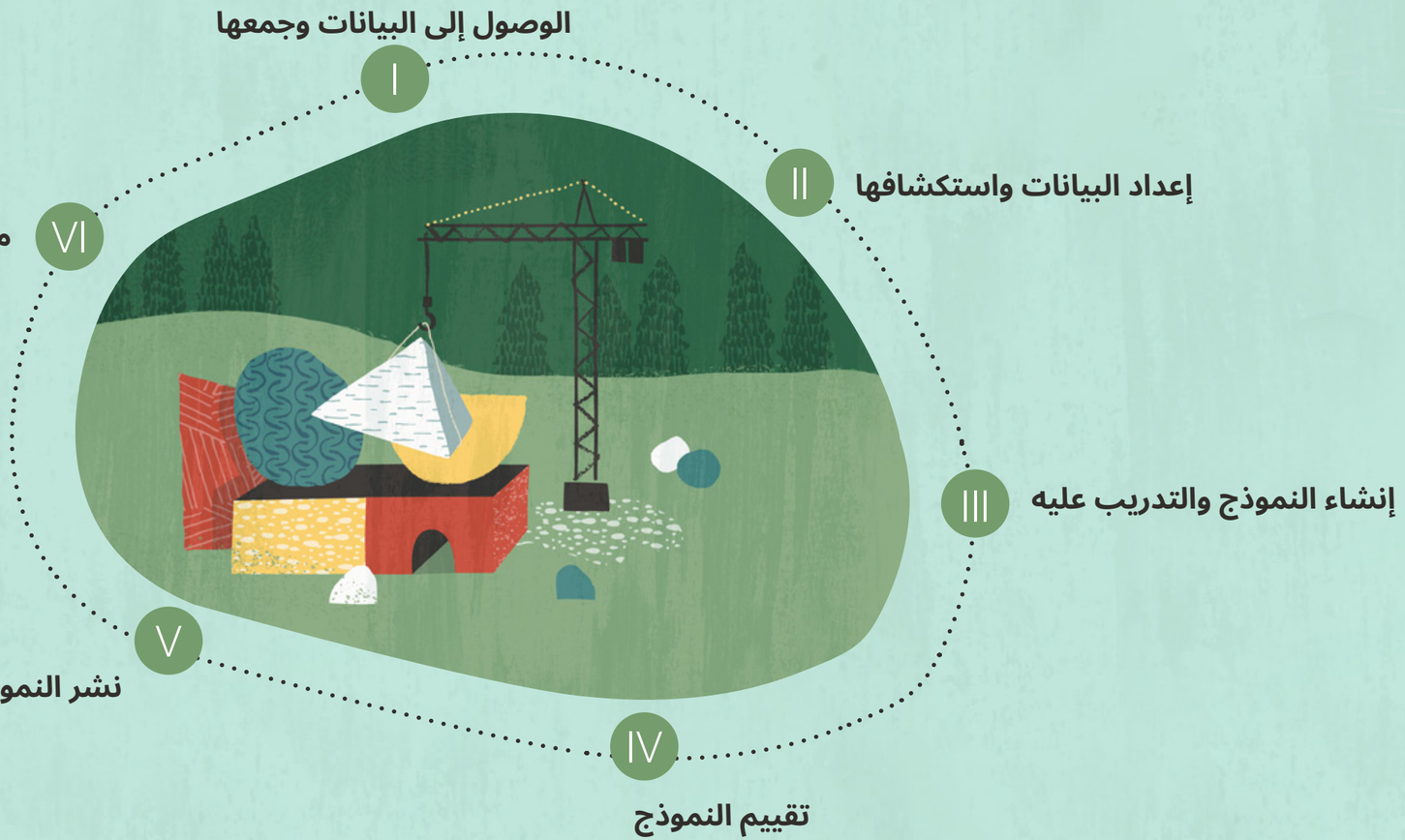
دورة حياة نماذج التعلم الآلي

مقدمة

التعلم الآلي هو مجال تستثمر فيه المؤسسات بشكل متزايد أو تحدده كمجال محتمل للنمو. ثمة أسباب عديدة تدفع المؤسسات إلى الاستثمار في التعلم الآلي، بدءًا من القدرة على الاستفادة من البيانات للعثور على معلومات معمقة حول عملائها لجعل العمليات أكثر كفاءة. في هذا الكتاب، نقوم بتقسيم كيفية إنشاء نماذج التعلم الآلي في ست خطوات: الوصول إلى البيانات وجمعها، وإعداد البيانات واستكشافها، وإنشاء النماذج والتدريب عليها، وتقييم النماذج، ونشرها ومراقبتها.

يُعدّإنشاء نموذج تعلم آلي عملية تكرارية. يتم تكرار العديد من الخطوات اللازمة لإنشاء نموذج تعلم آلي وتعديلها حتى أداء النموذج. تتطلب هذه يشعر علماء البيانات بالرضا عن العملية قدرًا كبيرًا من استكشاف البيانات والتمثيل المرئي والتجارب حيث يجب استكشاف كل خطوة وتعديلها ومراجعتها بشكل مستقل.

خطوات إنشاء نماذج التعلم الآلي



I. الوصول إلى البيانات وجمعها

إن الخطوة الأولى نحو حل مشكلة التعلم الآلي هي الوصول إلى البيانات. عادةً ما يحصل علماء البيانات على البيانات لمشاكل الأعمال التي يتعاملون معها من خلال الاستعلام عن قواعد البيانات التي تخزن شركاتهم بياناتها فيها. بالإضافة إلى ذلك، هناك قيمة كبيرة في مجموعات البيانات غير المنظمة التي لا تتلاءم بشكل جيد مع قاعدة البيانات العلائقية (مثل السجلات والنصوص الأولية والصور ومقاطع الفيديو، وما إلى ذلك). تتم معالجة مجموعات البيانات هذه بكثافة عبر مسارات استخراج وتحويل وتحميل (ETL) مكتوبة بواسطة مهندسي البيانات وعلماء البيانات. توجد مجموعات البيانات هذه في مستودع بيانات أو في قاعدة بيانات (إما علائقية أم لا) .

عندما لا تتوفر البيانات اللازمة لدى علماء البيانات لحل مشاكلهم، يمكنهم الحصول عليها عن طريق استخلاصها من المواقع الإلكترونية أو شرائها من مزودي البيانات أو جمعها من الدراسات الاستقصائية، وبيانات مسار التنقل، والمستشعرات، والكاميرات، إلخ.

II. إعداد البيانات واستكشافها

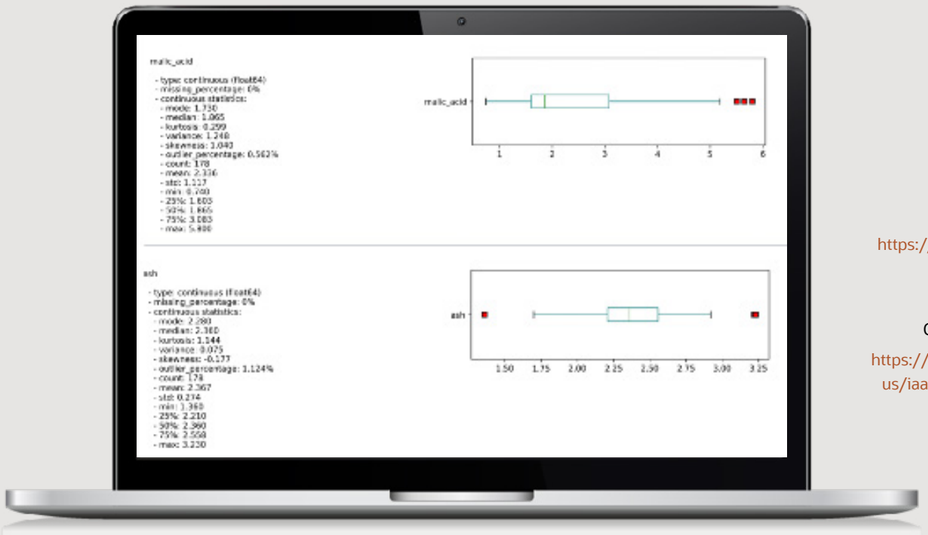
بعد الحصول على البيانات، يتعين على علماء البيانات إعداد البيانات الأولية، واستكشاف البيانات وتصورها وتحويلها، وربما تكرار الخطوات حتى تكون البيانات جاهزة للاستخدام في إنشاء النماذج. يشتمل إعداد البيانات على تنقيح البيانات الأولية ومعالجتها قبل التحليل. قبل إنشاء أي نموذج تعلم آلي، يتعين على علماء البيانات فهم البيانات المتاحة. قد تكون البيانات الأولية فوضوية أو مكررة أو غير دقيقة. يقوم علماء البيانات باستكشاف البيانات المتاحة لهم، ثم يقومون بتنقيحها من خلال تحديد البيانات الفاسدة وغير الدقيقة وغير المكتملة واستبدالها أو حذفها .

بالإضافة إلى ذلك، يتعين على علماء البيانات تحديد ما إذا كانت البيانات تحتوي على تسميات أم لا.

على سبيل المثال، إذا كان لديك سلسلة من الصور وترغب في تطوير نموذج اكتشاف لتحديد ما إذا كانت هناك سيارة في الصورة، فيجب أن يكون لديك مجموعة من الصور تسمي ما إذا كانت هناك سيارة بها، ومن المرجح أن تحتاج مربعات محددة حول السيارات في الصور. إذا كانت الصور تفتقر إلى التسميات، فسيتعين على علماء البيانات تسميتها. هناك أدوات مفتوحة المصدر إضافة إلى موردين تجاريين يوفرون منصات لتسمية البيانات، بالإضافة إلى أشخاص يهتمون بوضع التسميات يمكن توظيفهم للمهمة .

بعد تنقيح البيانات، يستكشف علماء البيانات الميزات (أو المتغيرات) في مجموعة البيانات الخاصة بهم، ويحددون أي علاقة بين تحويلات الميزات.

ملخص الإحصائيات والتمثيل المرئي للميزات في مجموعة بيانات تضم ثلاثة أنواع من النبيذ وميزات كل نبيذ.



المصدر: https:// scikit-learn.org/ Scikit Learn/، تم إجراء التمثيل المرئي باستخدام علم بيانات Oracle Cloud Infrastructure https:// docs.cloud.oracle.com/en-us/iaas/ data-science/user/data-science.htm

هناك أدوات مختلفة يمكن لعلماء البيانات استخدامها لتحليل البيانات الاستكشافية في المكتبات المفتوحة المصدر ومنصات التحليلات/علوم البيانات. تعد الأداة التي تقوم بإجراء التحليل الإحصائي لمجموعة البيانات وتكوين تمثيلات مرئية للبيانات لإنشاء مخططات للميزات مفيدة في هذه الخطوة.

من المهم معرفة أنواع الميزات الموجودة في مجموعة البيانات. يمكن أن تكون الميزات رقمية، والتي يمكن أن تكون نقطة عائمة أو عددًا صحيحًا. تشتمل الميزات التصنيفية على عدد محدود من القيم المحتملة، وعادةً ما يتم تعيين البيانات إلى مجموعات. على سبيل المثال، إذا كانت لديك مجموعة بيانات من دراسة استقصائية تستهدف العملاء، فإن جنس المستجيب (ذكر أو أنثى) هو ميزة تصنيفية. الميزات الترتيبية عبارة عن ميزة تصنيفية ذات ترتيب أو مقياس معين. على سبيل المثال، الأجوبة على رضا العملاء: هناك ترتيب معين لكل من راض جدًا، راضٍ ، غير مبالٍ ، غير راضٍ ، وغير راضٍ بتاتًا. يمكنك تحويل هذا الترتيب إلى مقياس عدد صحيح (5> -1). بعد تحديد نوع الميزات الموجودة، تتمحور الخطوة التالية حول الحصول على توزيع قيم لكل ميزة والحصول على إحصائيات موجزة لكل منها. سيساعد ذلك في الإجابة عن الأسئلة التالية حول مجموعة البيانات:

- هل يتم تغيير مجموعة البيانات نحو مجموعة من القيم أو مجموعة فرعية من الفئا ت؟
- ما هي قيم الحد الأدنى والحد الأقصى والمتوسط والمعتدل والوضع للميزة؟
- هل هناك قيم مفقودة أو قيم غير صالحة مثل صفر ؟
- إذا كان الأمر كذلك، فكم عددها ؟

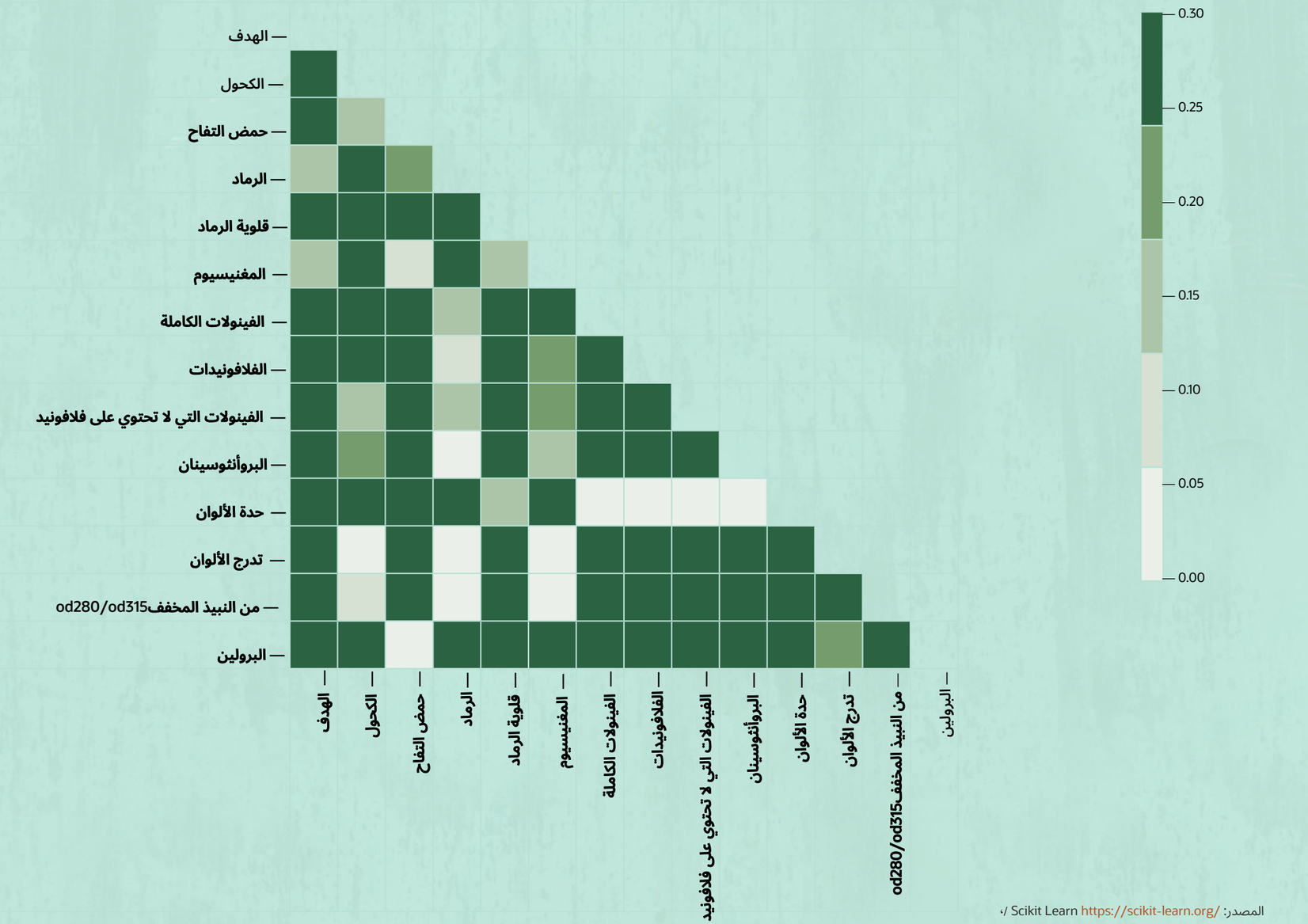
هل توجد قيم شاذة في مجموعة البيانات؟ خلال خطوة استكشاف البيانات، من المفيد تخطيط الميزات ومقارنتها مع بعضها البعض أيضًا لتحديد الأنماط في مجموعة البيانات. ويساعد ذلك في تحديد الحاجة إلى تحويل البيانات. بعض الأسئلة التي تحتاج إلى الإجابة عنها هي :

- كيف تتعامل مع القيم المفقودة؟ هل تريد ملء القيم وإذا كان الأمر كذلك، فما هو النهج الذي تخطط لاعتماده لملئه للقيمة المفقودة؟ تتضمن بعض النُهُج أخذ القيمة المتوسطة والمعتدلة والوضع وقيمة الإدخال التقريبية ومتوسط قيم الإدخال التقريبية.
- كيف يمكنك التعامل مع القيم الشاذة؟
- هل بعض ميزاتك مرتبطة ببعضها البعض ؟
- هل تحتاج إلى تطبيع مجموعة البيانات أو إجراء بعض التحولات الأخرى لإعادة توزيع البيانات (مثل تحويل السجلات) ؟
- ما هو نهجك تجاه القائمة الطويلة من القيم التصنيفية؟ هل تستخدمها كما هي أو تجمعها بطريقة مجدية أو تتجاهل مجموعة فرعية منها تمامًا ؟

خلال خطوة استكشاف البيانات، يمكنك تحديد الأنماط في مجموعة البيانات للحصول على أفكار حول كيفية تطوير ميزات جديدة تمثل مجموعة البيانات بشكل أفضل. ويعرف هذا باسم هندسة الميزات. على سبيل المثال، إذا كان لديك مجموعة بيانات حركة المرور لعدد السيارات التي تمر عبر تقاطع رئيسي في كل ساعة، فقد ترغب في إنشاء ميزة جديدة تصنف الساعة إلى أجزاء مختلفة من اليوم، مثل الصباح الباكر، منتصف الصباح، في وقت مبكر بعد الظهر، وفي وقت متأخر من بعد الظهر، وخلال الليل .

بالنسبة إلى الميزات التصنيفية، غالبًا ما يكون من الضروري وجود ترميز ساخن واحد للميزة. تعني عبارة ترميز ساخن واحد تحويل ميزة تصنيفية إلى ميزات ثنائية، واحدة لكل فئة من الفئات. على سبيل المثال، افترض أن لديك مجموعة بيانات من العملاء، ولدينا ميزة تنص على أن العميل يأتي من: واشنطن وأوريجون وكاليفورنيا. سينتج ترميز ساخن واحد ميزتين ثنائيتين حيث إحدى الميزات هي ما إذا كان العميل من ولاية واشنطن أم لا، والميزة الثانية هي ما إذا كان العميل من أوريغون أم لا. من المفترض أنه إذا لم يكن العميل من واشنطن أو أوريغون، فسيكون من كاليفورنيا، لذلك لا توجد حاجة إلى ميزة ثالثة.

خريطة حرارية لمدى ارتباط الميزات ببعضها البعض، من مجموعة بيانات لثلاثة أنواع من النبيذ وميزات كل نوع نبيذ.



المصدر: Scikit Learn https://scikit-learn.org/، المرئي باستخدام علم بيانات Oracle Cloud Infrastructure https://docs.cloud.oracle.com/en-us/iaas/data-science/using/data-science.htm

III. إنشاء النموذج والتدريب عليه

يشتمل إنشاء النموذج على اختيار نماذج التعلم الآلي الصحيحة لحل المشاكل والميزات التي تدخل في النماذج. في الخطوة الأولى من إنشاء النموذج، يحتاج علماء البيانات إلى تحديد ما قد يكون نموذج التعلم الآلي المناسب لحل المشكلة. هناك نوعان رئيسيان من نماذج التعلم الآلي: الخاضع للإشراف وغير الخاضع للإشراف. يتضمن التعلم الخاضع للإشراف نمذجة مجموعة من بيانات الإدخال إلى إخراج أو تسمية. التصنيف والانحدار عبارة عن مشاكل تعلم خاضعة للإشراف. يتضمن التعلم غير الخاضع للإشراف نمذجة مجموعة من بيانات الإدخال بدون تسمية. على سبيل المثال، تعتبر تجزئة العملاء مشكلة تعلم غير خاضعة للإشراف. أنت لا تعرف مسبقًا الجزء الذي ينتمي إليه العميل. سيتم تعيين الجزء بواسطة النموذج .

يتم استخدام فئات مختلفة من نماذج التعلم الآلي لحل مشاكل التعلم غير الخاضعة والخاضعة للإشراف. عادةً ما يحاول علماء البيانات استخدام نماذج وخوارزميات متعددة وإنشاء مرشحين متعددين للنماذج. لا يعرف علماء البيانات مسبقًا النموذج الذي سيعمل بشكل أفضل على مجموعة البيانات، لذلك يقومون بتجربة العديد منها. أثناء التدريب على النموذج، قد يقوم عالم البيانات بتحديد الميزات، وهي عملية تحديد مجموعة فرعية فقط من الميزات كإدخال إلى نموذج التعلم الآلي. تتمثل ميزات الحد من عدد متغيرات الإدخال في تقليل التكلفة الحسابية للتدريب على النموذج، وجعل النموذج أكثر قابلية للتعميم وربما تحسين أداء النموذج.

خلال التدريب على النموذج، يتم تقسيم مجموعة البيانات إلى مجموعات تدريب واختبار. تُستخدم مجموعة بيانات التدريب للتدريب على النموذج، وتُستخدم مجموعة بيانات الاختبار لمعرفة مدى جودة أداء النموذج على البيانات التي لم يراها. ستتم مناقشة التقييم النموذجي بمزيد من التفصيل أدناه .

يعتبر ضبط المعلمات الفائقة للنموذج مهمة رئيسية في عملية التدريب على النموذج. تُعدّ النماذج خوارزميات، أما المعلمات الفائقة فهي المقابض التي يمكن لعالم البيانات توليفها لتحسين أداء النموذج. على سبيل المثال، يُشكّل عمق شجرة القرار معلمة فائقة .

يمكنك اختيار أن تكون شجرة القرار عميقة جًدا أو ضحلة جدًا. سيؤثر ذلك على التحيز والتباين في نموذجك. التحيز هو خطأ عدم كفاية أو خطأ عدم تسجيل العلاقة بين الميزات والإخراج. التباين هو الخطأ الناتج من الإفراط في التناسب حيث يؤدي النموذج أداءً جيدًا في مجموعة بيانات التدريب ولكنه لا يؤدي أداءً جيدًا بالنسبة إلى البيانات غير المرئية. يمكن أتمتة ضبط المعلمات الفائقة للنموذج جزئيًا، على الرغم من ضرورة إشراك علماء البيانات دائمًا في العملية .

هذا ويتعين على علماء البيانات تحديد نوع موارد الحوسبة التي يحتاجون إليها للتدريب على نماذجهم. يمكنك إعداد البيانات وتدريب النماذج محليًا على جهاز الكمبيوتر الخاص بك. ومع ذلك، اعتمادًا على كمية البيانات المطلوب إعدادها ثم استخدامها للتدريب على النموذج، قد لا يكون الكمبيوتر الخاص بك كافيًا. قد تضطر إلى نقل حمل العمل إلى السحابة حيث يمكنك الوصول إلى مجموعة أوسع من موارد الحوسبة بما في ذلك وحدات معالجة الرسومات (GPU).

يمكن التدريب على بعض النماذج بشكل أسرع على الأجهزة المتخصصة (على سبيل المثال، تدريب الشبكات العصبية/ نماذج الشبكة العصبية العميقة على وحدات معالجة الرسومات). يمكنك أيضًا استكشاف بيئات التدريب الموزعة التي يمكن أن تسرع العملية، تحديدًا عندما لا يمكن أن تتلاءم كمية البيانات في ذاكرة أكبر آلة متوفرة، من خلال تقسيم البيانات وتوزيعها عبر أجهزة متعددة، أو عندما تريد تدريب مرشحين متعددين في الوقت نفسه على آلات منفصلة.

التعلم الآلي التلقائي

حظي التعلم الآلي التلقائي باهتمام كبير خلال السنوات القليلة الماضية بسبب الوعد بالقدرة على جعل التعلم الآلي أكثر قابلية للوصول إلى جمهور أكبر. يختصر التعلم الآلي التلقائي بـ AutoML. وهو يقوم بأتمتة عملية تحديد الميزة وتحديد النموذج/الخوارزمية وضبط المعلمات الفائقة. إنها ميزة مشتركة بين كل المنصات الرئيسية لعلوم البيانات. يمكن للمستخدمين تغذية مجموعة بيانات إلى التعلم الآلي التلقائي، وستقوم بتدريب العديد من نماذج التعلم الآلي، وضبط المعلمات الفائقة لهذه النماذج، وتقييم أدائها مقابل بعضها البعض.

يمكن للتعلم الآلي التلقائي تحسين إنتاجية علماء البيانات من خلال أتمتة عملية التدريب. وهوا يسمح أيضًا لمحللي البيانات ومطوريها بإنشاء نماذج تعلم آلي من دون تعديل كل جانب من جوانب عملية التدريب النموذجية التي تأتي مع خبرة علوم البيانات. تدعم معظم إمكانيات التعلم الآلي التلقائي البيانات الجدولية لمشاكل التصنيف والانحدار بينما يتوفر لدى الآخرين عروض أكثر تقدمًا تدعم الصور والبيانات النصية، بالإضافة إلى التنبؤ بالسلسلة الزمنية .

يكمن العيب في التعلم الآلي التلقائي، أو أي نموذج معقد، في أنه قد يبدو مثل حل الصندوق الأسود بحيث يصعب على المستخدمين فهم كيفية وصول النماذج إلى التوقعات. يجب أن يتطلع المستخدمون إلى عرض قابلية شرح النموذج لنظام التعلم الآلي التلقائي لمعرفة القدرة الموجودة لمساعدة المستخدمين على تفسير النماذج وفهم كيفية وصول النماذج المحددة إلى التوقعات.

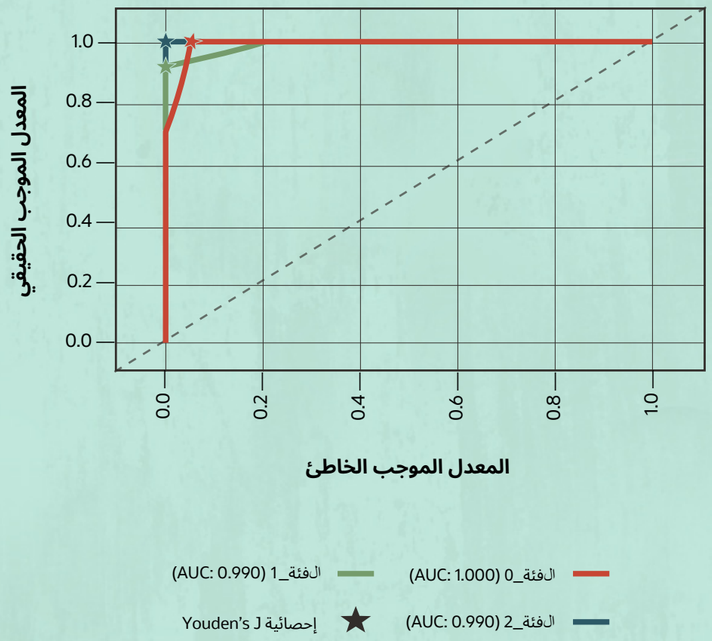
تفسيرات نموذجية تقع عادةً في التفسير العالمي والشرح المحلي. الشرح العام هو فهم السلوك العام لنموذج التعلم الآلي ككل. ويشمل ذلك شرح مدى أهمية كل ميزة في المساهمة في التوقعات النموذجية. يوفر الشرح المحلي فهمًا لسبب قيام نموذج التعلم الآلي بتوقع معين لنموذج بيانات واحد. على سبيل المثال، لماذا تتوقع خوارزمية الكشف عن الاحتيال معاملة معينة على أنها احتيالية؟

IV. تقييم النموذج

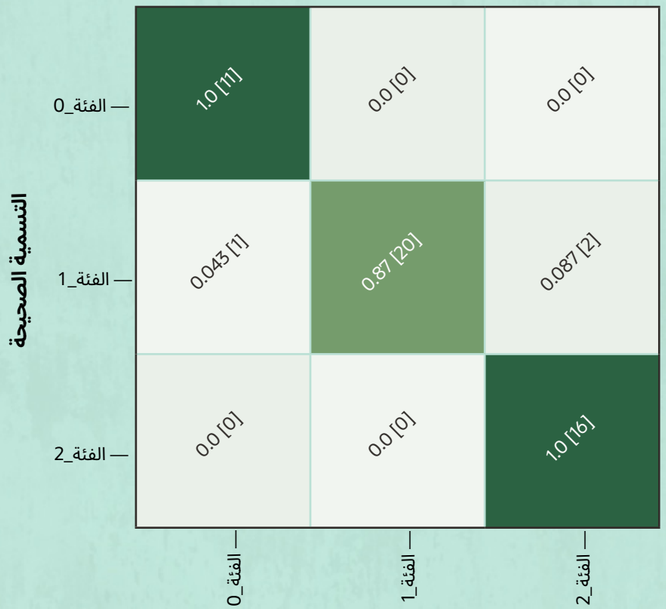
هناك العديد من الأدوات المفتوحة المصدر التي تساعد علماء البيانات على حساب المقاييس لتقييم نماذج التعلم الآلي ومساعدتهم في تصور المقاييس)على سبيل المثال، منحنى AUC-ROC والحصول على الرسوم البيانية ورفعها(. عند تقييم نماذج التعلم الآلي، يحتاج علماء البيانات إلى تحديد المقاييس المهمة لمشكلة الشركة التي يحاولون حلها.

بالنسبة إلى مشاكل التصنيف، يمكن للمرء استخدام الدقة لتقييم النموذج، ولكن في بعض الأحيان قد لا يكون الخيار الأمثل للقياس. إذا كانت المشكلة تنطوي على اكتشاف ما إذا كان شخص ما يعاني من مرض نادر، فقد يكون القياس الأفضل هو عدد الأشخاص المصابين بالمرض الذين يتم تشخيصهم بدقة مقسوم على كل الأشخاص المصابين بالمرض. في هذه الحالة، سيكون من المفيد النظر إلى مصفوفة الارتباك التي تبين عدد النتائج الإيجابية الحقيقية، والسلبية الحقيقية، والإيجابية الخاطئة، والسلبية الخاطئة، وحساب الدقة والاستدعاء. بالنسبة إلى مشاكل الانحدار، يمكنك استخدام مقاييس مثل خطأ الجذر المتوسط المربع أو متوسط الخطأ المطلق أو حساب معامل التحديد 2r. بالنسبة إلى المشاكل غير الخاضعة للإشراف، تعتبر مجموعة من المجموعات ذات التماسك العالي داخل المجموعات والفصل بين المجموعات مثالية. ويمكن قياس ذلك بمقاييس مثل درجة الصورة الظلية ومعامل كالينسكي ـ هاراباس.

منحنى خصائص تشغيل جهاز الاستقبال (ROC curve) للتصنيف المتعدد الطبقات الذي تم إنشاؤه للنتائج من نموذج غابات عشوائي حول التوقع بنوع النبيذ من مجموعة بيانات تحتوي على ثلاثة أنواع من النبيذ وميزات كل نبيذ.



مصفوفة الارتباك للتصنيف متعدد الطبقات للنتائج من نموذج غابات عشوائي حول توقع نوع النبيذ على أساس ميزات النبيذ، من مجموعة بيانات تحتوي على ثلاثة أنواع من النبيذ وميزات كل نبيذ.



المصدر: مكتبة Scikit Learn /https://scikit-learn.org، التمثيل المرئي الذي يتم تنفيذه باستخدام علم بيانات Oracle Cloud Infrastructure https://docs.cloud.oracle.com/en-us/iaas/data-science/using/data-science.htm

٧. نشر النموذج

بعد اكتمال عمليات التدريب والتقييم النموذجية، يتم حفظ أفضل نماذج المرشحين. عادةً ما يتم حفظ النماذج بتنسيقات Pickle وONNX وPMML. ووفقًا للأهداف، قد يعمل علماء البيانات على مشكلة في التعلم الآلي لإثبات المفهوم أو التجربة أو نشره إلى الإنتاج. يستهلك نشر النموذج التوقعات التي أجراها نموذج التعلم الآلي بطريقة ما. وعلى الأرجح، يجب أيضًا نشر مسار تحويلات البيانات. عادةً ما يعمل علماء البيانات مع المهندسين على نشر النماذج.

وتبعًا لكيفية الاستفادة من التوقعات، يمكنك النشر للاستفادة من التوقعات في مجموعات أو في الوقت الفعلي. بالنسبة إلى الاستفادة في مجموعات، يمكن جدولة التوقعات (على سبيل المثال، كل ساعة، كل يوم). ويمكن تخزين التوقعات بعد ذلك في قاعدة بيانات واستهلاكها في تطبيقات أخرى. عادةً، يكون حجم البيانات التي تقوم بمعالجتها أكبر من حجم التوقع في الوقت الفعلي. ومن الأمثلة على حالة الاستخدام أن تقوم بتشغيل موقع تجارة إلكترونية وتريد إرسال بريد إلكتروني أسبوعي إلى العملاء حول المنتجات الموصى بها لهم بناءً على المشتريات السابقة. يمكن جدولة نماذج التعلم الآلي للتشغيل مسبقًا .

بالنسبة إلى الاستفادة في الوقت الفعلي، تقوم الحالة المحفزة ببدء عملية استخدام النموذج المستمر لخدمة التوقع. على سبيل المثال، يتطلب تحديد ما إذا كانت المعاملة احتيالية أم لا عند بدء الدفع أن تقوم بالتوقع في الوقت الفعلي. يتعين عليك التفكير في مدى سرعة خدمة التوقعات (مللي ثانية، ثوانٍ؟)، وحجم الطلب على الخدمة، وحجم البيانات لتشغيل التوقعات.

يشكل الحد من زمن الوصول لخدمة التوقعات أهمية كبيرة. يمكنك تحسين زمن وصول الخدمة باستخدام نموذج أصغر في الحجم، باستخدام المسرع مثل وحدة معالجة الرسومات (GPU) وتحسين كيفية استرجاع الميزات المرتبطة بالكيان للتوقع في الوقت الفعلي (على سبيل المثال، إذا كنت توصي بمنتجات لمستخدم أثناء قيام المستخدم باستعراض موقع، فإن التحسينات التي طرأت على كيفية الحصول على معلومات عن عمليات الشراء السابقة للمستخدم يمكن أن تحسن زمن الوصول.)

هناك أدوات وعروض منصات سحابية مختلفة لنشر النموذج مثل منصات الوظائف كخدمة (FaaS)، والنشر المدار بشكل كامل للنماذج مثل نقاط انتهاء HTTP وDIY مع flask أو Django في نظام تنظيم حاوية مثل k8 وdocker swarm، إلخ .

VI. مراقبة النماذج

تُعد مراقبة النماذج خطوة صعبة تنساها أحيانًا المؤسسات التي لا تمتلك مبادرات متقدمة للتعلم الآلي وعلم البيانات. تتطلب إعادة التدريب على النماذج وإعادة توزيعها وقتًا من فريق علوم البيانات والهندسة وموارد الحوسبة. وتساعد مراقبة النماذج الفريق على تحديد ما إذا كان من الضروري إعادة التدريب على النموذج وإعادة توزيعه ومتى يكون ذلك ضروريًا. ويمكن تقسيم الرصد النموذجي إلى عنصرين: الرصد السريع/الإحصائي للأداء النموذجي ورصد العمليا ت.

بعد نشر النماذج، انخفضت المقاييس التي تم من خلالها قياس النماذج والتدريب عليها في الإنتاج. وذلك لأن البيانات غير ثابتة. يمكن أن يظهر عدم الثبات بطرق عديدة: يمكن أن تأخذ الميزات في بيانات الإنتاج قيمًا خارج النطاق في مجموع بيانات التدريب؛ وقد يكون هناك انحراف بطيء في توزيع القيم، إلخ.

بسبب تدهور النموذج، يجب مراقبة جودة النماذج لتحديد ما إذا كان بالإمكان إعادة تدريب النموذج وإعادة النشر ومتى يتم ذلك. في بعض الأحيان، لا يمكن الحصول فورًا على دقة التوقع من البيانات المباشرة التي تنتقل إلى نظام الإنتاج.
على سبيل المثال، قد يستغرق الأمر بعض الوقت قبل أن تتمكن من تحديد ما إذا كان نموذج توقع تبديل موفر الخدمة

أو نموذج اكتشاف احتيال يوفر توقعًا دقيقًا. ومع ذلك، من الممكن النظر في إحصاءات وتوزيع بيانات التدريب مقارنة

بالبيانات المباشرة وكذلك مقارنة توزيع التوقعات النموذجية مع التدريب والبيانات المباشرة. على سبيل المثال، إذا كنت تتعامل مع نموذج تبديل العملاء، فيمكنك مقارنة ميزات العملاء المستخدمة لتدريب النموذج الخاص بك بميزات العملاء في نظام الإنتاج. أيضًا، يمكنك الاطلاع على النسبة المئوية للعملاء المتوقع أن يتم تبديلهم في عينة التدريب مقارنة بالإنتاج المباشر.

ستتطلب مراقبة العمليات لنظام التعلم الآلي شراكة بين علماء البيانات وفريق الهندسة. تتضمن الأمور التي يجب مراقبتها زمن الوصول إلى الخدمة واستخدام الذاكرة/وحدة المعالجة المركزية والإنتاجية وموثوقية النظام. هذا ويلزم إعداد السجلات والمقاييس للتتبع والمراقبة. تحتوي السجلات على سجلات الأحداث، إلى جانب وقت حدوثها. ويمكن استخدامها للتحقيق في حوادث محددة ومعرفة سبب الحادث. Kibana عبارة عن أداة مفتوحة المصدر تستخدم في البحث عن السجلات وعرضها. تقوم القياسات بقياس استخدام نظام التعلم الآلي وسلوكه. Promethus وGrafana هما أداتان لمراقبة المقاييس.

الخاتمة

نأمل أن يكون هذا دليلًا مفيدًا حول الخطوات اللازمة لإنشاء نموذج تعلم آلي. من المهم أن نتذكر أن التعلم الآلي عملية متكررة للغاية، وسيتم تكرار الخطوات المبينة في هذا الكتاب وتحسينها عدة مرات.

هناك العديد من الموارد المتوفرة التي تتعمق أكثر في كل خطوة من الخطوات التي يغطيها هذا الكتاب، ويمكنك معرفة المزيد عنها في أثناء اتخاذ القرارات المتعلقة بإستراتيجية علوم البيانات في مؤسستك. إذا كنت مستعدًا لبدء مختبرات عملية بحيث يمكنك التجربة Oracle، تقدم من خلال إنشاء نماذج علوم البيانات الخاصة بك.

شركة Oracle

المقر الرئيسي العالمي

Oracle Parkway, Redwood Shores, CA 94065, USA 500

للاستفسارات على نطاق العالم

الهاتف + ORACLE1.1.800 + 1.650.506.7000

الفاكس + 1.650.506.7200

oracle.com

تواصل معنا

يرجى الاتصال بالرقم ORACLE1.1.800 أو تفضل بزيارة الموقع oracle.com. خارج أمريكا الشمالية، ابحث عن مكتبك المحلي على oracle.com/contact.

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

الكاتب

ويندي ييب، عالمة في البيانات.

حقوق النشر © 2020، Oracle و/أو الشركات التابعة لها. كل الحقوق محفوظة. يتم توفير هذا المستند لأغراض إعلامية فقط، وتخضع محتوياته للتغيير من دون إشعار. لا نضمن خلوه من الأخطاء، كما أنه لا يخضع لأي ضمانات أو شروط أخرى، سواء تم التعبير عنها شفويًا أم ضمنيًا في القانون، بما في ذلك الضمانات الضمنية وشروط قابلية التسويق أو الملاءمة لغرض معين. ونخلي مسؤوليتنا على وجه التحديد عن أي مسؤولية في ما يتعلق بهذا المستند، ولا تتشكل به أي التزامات تعاقدية إما بشكل مباشر أو غير مباشر. يُحظر إعادة إنتاج هذا المستند أو نقله بأي شكل أو بأي وسيلة، سواء الإلكترونية أو ميكانيكية، لأي غرض، بدون إذن كتابي مسبق منا.

تُعد Oracle وJava من العلامات التجارية المسجلة لشركة Oracle و/أو للشركات التابعة لها. قد تكون الأسماء الأخرى علامات تجارية تخص أصحابها.

تعتبر كل من Intel وIntel Xeon علامات تجارية أو علامات تجارية مسجلة لشركة Intel. تُستخدم كل علامات SPARC التجارية بموجب ترخيص وهي علامات تجارية أو علامات تجارية مسجلة لشركة SPARC International, Inc، إن AMD وOpteron وشعار AMD وشعار AMD Opteron هي علامات تجارية أو علامات تجارية مسجلة لشركة Advanced Micro Devices. UNIX عبارة عن علامة تجارية مسجلة لـ The Open Group 05.10.19.