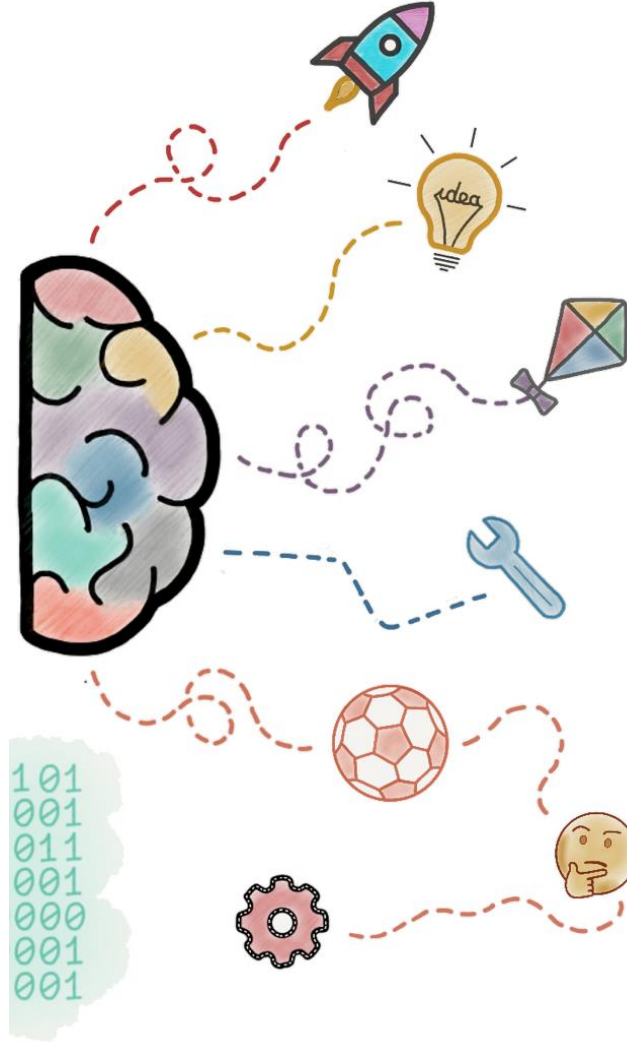


الخطوات الرئيسية

1. ضع اطاراً عاماً للمشكلة، ثم انظر إلى الصورة الكبيرة
2. احصل على البيانات
3. استكشف هذه البيانات واستخلص الرؤى
4. حضر هذه البيانات لعرض أنماطها الأساسية بشكل أفضل لخوارزميات تعلم الآلة
5. جرب عدة نماذج وخوارزميات وأعد قائمة مختصرة للنماذج الأفضل
6. اضبط متغيرات النموذج ثم ادمجه في النظام الكامل
7. اعرض النموذج والنظام
8. أطلق، وراقب النموذج ثم اعمل على صيانتها باستمرار

تعلم الآلة للجميع



قائمة خطوات مشروع تعلم الآلة

سترشدك قائمة الخطوات هذه إلى العمل المنظم على مشاريع تعلم الآلة التي تعمل عليها. وهنا الخطوات الثمانية الرئيسية، التي يمكنك إعادة ضبطها وتعديلها لتناسب واحتياجاتك واحتياجات مشروعك

إعداد

م. محمد رضا كاتي
@morikapt

تصميم

محمد هيثم كاتي
@heysem_k



1- ضع إطاراً عاماً للمشكلة، ثم انظر الكبيرة

حاول أن تسأل الأسئلة التشخيصية لفريق العمل، ثم اجمع الأجوبة وقم بالتحقق منها فذلك ما يجعل منك عالم بيانات متميز

☐ تحديد الهدف من النموذج بمصطلحات العمل

☐ كيف سيتم استخدام هذا النموذج؟

☐ ماهي الحلول الحالية؟ أو الحلول الملتوية المستخدمة حالياً؟

☐ كيف يمكن تأطير هذه المشكلة؟ (تعليم بمعلم/تعليم بدون معلم، ...)

☐ كيف سيتم قياس أداء النموذج؟

☐ هل يتماشى مقياس الأداء مع أهداف العمل؟

☐ ما هو الحد الأدنى للدقة المطلوبة اللازمة للوصول إلى أهداف العمل؟

☐ ماهي المشاكل المماثلة؟ هي يمكنك إعادة استخدام الخبرات أو الأدوات السابقة؟

☐ هل هناك خير في هذا المجال؟

☐ كيف كانت هذه المشكلة تحل يدوياً سابقاً؟

☐ ضع قائمة بكافة الفرضيات أو الإجابات التي قمت بها أنت أو فريقك للإجابة على هذه الأسئلة وغيرها

☐ حاول التحقق من هذه الفرضيات أو الإجابات إن أمكن



2- احصل على البيانات

حاول أتمتة خطواتك بأكبر قدر ممكن حتى تتمكن من الحصول على بيانات جديدة بسهولة

☐ ضع قائمة بالبيانات التي تحتاجها والمقدار الذي تحتاجه

☐ ابحث ووثق من أين يمكنك الحصول على هذه البيانات

☐ تحقق من توفر المساحة الكافية لاستيعاب مجموعات البيانات تلك

☐ تحقق من القيود القانونية للوصول إلى البيانات

☐ تأكد من حصولك على إذن وصول للبيانات لمدة المشروع

☐ قم بإنشاء مساحة عمل جديدة مع مساحة تخزين كافية

☐ احصل على البيانات

☐ حول البيانات إلى صيغة يمكنك معالجتها بسهولة (دون تغيير البيانات نفسها)

☐ تأكد من حذف المعلومات الحساسة أو حمايتها (كإخفاء الهوية مثلاً)

☐ تحقق من حجم ونوع البيانات (سلاسل زمنية، عينات، مواقع جغرافية، إلخ)

☐ قم باقتطاع مجموعة الاختبار، وضعها جانباً ولا تنظر إليها أبداً (لا تتطفل عليها)



3- استكشف هذه البيانات

واستخلص الرؤى

حاول أن تحصل على آراء الخبراء في مجال العمل، ستلعب الدور الأهم في نجاح المشروع

☐ أنشئ نسخة من البيانات لأغراض التجربة والاستكشاف (يمكن أن تكون عينة من البيانات إذا كانت ذات حجم كبير، بحيث يمكن التحكم بها بسهولة)

☐ تنظيف البيانات

- قم بإصلاح أو إزالة القيم المتطرفة (اختياري)
- املئ القيمة المفقودة (بصفر، أو المتوسط، أو الوسيط...) أو احذف السطور أو الأعمدة الخاصة بها

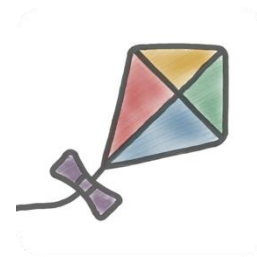
☐ اختر الميزات أو الأعمدة (اختياري):

- احذف الميزات التي لا توفر معلومات مهمة للمهمة التي تعمل عليها

☐ قم بهندسة الميزات إذا اقتضت الحاجة

- قطع القيمة المستمرة إلى فئات
- فكك الميزات أو الأعمدة إلى مركبات أبسط (فئات أو أصناف، التاريخ والوقت، إلخ)
- أضف بعض التحويلات الواعدة كميزات جديدة \sqrt{x} ، $\log x$ ، x^2
- جمع بعض الميزات بميزة واحدة جديدة

☐ تحجيم الميزات: مثل التوحيد normalization أو تطبيعها standardization



4- تحضير البيانات

- اعمل على نسخة جانبية من البيانات، وحافظ على النسخة الأصلية سليمة
- اكتب اجراء منفصل لكل تحويل من تحويلات البيانات التي تقوم بتطبيقها على الأعمدة، وذلك للأسباب التالية:
 - حتى تتمكن من تحضير البيانات بسهولة في المرة التالية التي تحصل فيها على نسخة جديدة من البيانات
 - يمكنك تطبيق هذه التحويلات في المشاريع المستقبلية
 - لتحضير مجموعة بيانات الاختبار وتنظيفها
 - لتحضير وتنظيف البيانات الجديدة عندما ترفع المشروع على البيئة الحية
 - لتسهيل التعامل مع خيارات تحضير البيانات التي قمت بإنشائها على أنها جزء من معاملات خوارزمية التعلم hyperparameters

☐ أنشئ دفتر ملاحظات Jupyter للاحتفاظ بسجل استكشاف البيانات

☐ ادرس الميزات أو الأعمدة وخصائصها الاسم

- النوع (أصناف أو فئات، رقمي أو عائم، محدد أو غير محدد، نصي، مهيكّل أو غير مهيكّل، إلخ)

- نسبة القيم المفقودة في كل ميزة أو عمود
- الضوضاء وأنواعها (عشوائية، قيم متطرفة، أخطاء تقريب، إلخ)
- حدد أهمية الميزات أو الأعمدة للمهمة المطلوبة

- حدد نوع التوزيع الاحتمالي لكل عمود أو ميزة (طبيعي، موحد، لوغاريتمي، إلخ)

☐ إذا كانت المهمة من نوع التعلم بمعلم فقم بتحديد عمود الهدف أو العنوان

☐ اعرض البيانات بشكل مرئي

☐ ادرس الترابط بين اعمدة البيانات الموجودة لديك

☐ ادرس كيف يمكن حل هذه المشكلة يدوياً

☐ حدد التحويلات التي يمكن تطبيقها على الأعمدة لتساهم في أداء أفضل للمهمة.

☐ حدد البيانات الاضافية التي يمكن أن تساهم في أداء أفضل للمهمة

☐ وثق كل ما تعلمته



5- حضر قائمة مختصرة للنماذج الواعدة

- قد يكون مفيداً أن تأخذ عينة من مجموعة بيانات التدريب حتى تتمكن من تدريب عدة نماذج مختلفة في وقت معقول هذا إذا كانت البيانات ضخمة (لكن انتبه أن هذا قد يقلل من دقة النماذج الكبيرة مثل الشبكات العصبية والغابات العشوائية)
- حاول أتمتة هذه الخطوات بأكبر قدر ممكن

☐ درّب العديد من النماذج السريعة والصغيرة من فئات مختلفة من الخوارزميات (خطية، Naive Bayes، SVM، غابة عشوائية، شبكة عصبية، إلخ) باستخدام القيمة الافتراضية لبارامترات خوارزميات التعلم hyperparameters



6- اضبط متغيرات النظام

☐ قم بقياس ومقارنة أداء كل نموذج مع الآخر

- استخدم التحقق المتقاطع Cross-Validation لحساب متوسط الخطأ والانحراف المعياري على n مرة من عمليات التحقق

☐ حلل أهم المتغيرات لكل خوارزمية

☐ حلل أنواع الأخطاء التي تحدثها كل خوارزمية على حدى

- حدد البيانات التي يستعملها الخبير البشري لتجنب مثل هذه الأخطاء

☐ قم بجولة سريعة على هندسة وتحديد الميزات المستخدمة

☐ كرر الخطوات السابقة من 1 حتى 5 مرة أو مرتين

☐ ضع قائمة بأفضل 3 إلى 5 نماذج واعدة، مفضلاً النماذج التي ترتكب أنواعاً مختلفة من الأخطاء

- سترغب باستخدام كافة بيانات التدريب في هذه الخطوة، خاصة عندما تتقدم نحو الضبط الدقيق النهائي للنموذج
- حاول أتمتة أكبر قدر ممكن من الخطوات كما هو الحال دائماً

☐ اضبط بارامترات الخوارزمية Hyperparameters باستخدام التحقق المتقاطع Cross-Validation

○ تعامل مع خيارات تحويل البيانات التي قمت بكتابتها في إجراءات منفصلة، كأنها بارامترات خوارزمية التعلم Hyperparameters، خاصة عندما لا تكون متأكد منها (على سبيل المثال هي يجب استبدال القيم المفقودة بصفر؟ أم بالمتوسط؟ أو أن تقوم بحذف سطورها بشكل كامل؟)

○ ابدء بالبحث العشوائي عن قيم بارامترات خوارزمية التعلم Hyperparameter إذا كان عددهم قليل، لكن ذلك لن يفيد إذا كان عدد هذه البارامترات كبير



8- الإطلاق!

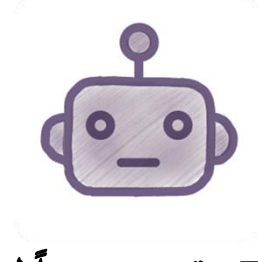
وصل النظام بنموذج تعلم الآلة، وطبق التحويلات على تدفق البيانات القادم من البيئة الحية

☐ جهاز الحل الخاص بك للبيئة الحية (وصل المدخلات من البيئة الحية، وقم بكتابة وحدات الاختبار، إلخ)

☐ اكتب كود لمراقبة أداء النموذج بشكل مستمر ویرسل تنبيه عند هبوط الأداء تحت حد معين

- احذر من تدهور الأداء البطيء، لأن النماذج تميل إلى الهرم والقدم من تطور البيانات
- راقب جودة المدخلات على النموذج، لأنه جهاز استشعار معطل قد يرسل قيماً عشوائية، وفريق آخر تعتمد على خرج كمدخلات قد يبدأ بإعطاء بيانات قديمة لسبب ما

☐ أعد تدريب النموذج بشكل منتظم كلما تراكمت لديك بيانات جديدة (أتمت قدر الإمكان)



7- قدم عرضاً لحلك النهائي

جهاز شرائح جذابة لعرض الخطوات التي قمت بها، والنتائج التي توصلت لها

☐ وثق كافة الخطوات التي قمت بها

☐ أنشاء عرض تقديمي جميل

○ تأكد من عرض الصورة الكبيرة أولاً

☐ اشرح كيف يحقق حلك أهداف العمل

☐ لا تنسى تقديم النقاط التي أثارت اهتمامك برحلة عملك على المشروع

☐ تأكد من عرض النتائج الرئيسية التي توصلت لها من خلال رسومات جميلة وعبارات رنانة يسهل تذكرها

☐ جرب أساليب المجموعات فغالباً ما يؤدي الجمع بين أفضل النماذج لديك إلى أداء أفضل من تشغيلها بشكل فردي

☐ قم بقياس خطأ التعميم Generalization Error لأداء النموذج على مجموعة الاختبار فقط عندما تكون واثقاً من نموذجك النهائي.

لا تقم بتعديل نموذجك بعد قياس خطأ التعميم، لأن ذلك سيفرط بتخصيص Overfitting نموذجك على مجموعة الاختبار.

تعلم الآلة للجميع

