تعلم الآلة لتطبيقات المخاطرة العالية

العنوان: تعلم الآلة لتطبيقات المخاطرة العالية: الأساليب للذكاء الاصطناعي المسئول

المؤلفون: باتريك هال، جيمس كورتيس، و بارول باندي

الناشر: أورلي O›Rleily

تاريخ النشر: ٢٠٢٢

عدد الصفحات: ٨٩ صفحة

اشتمل هذا الكتاب علي ٣ فصول تتمثل في التالي:

١.حوكمة النموذ المعاصر

- الالتزام القانوني الأساسي

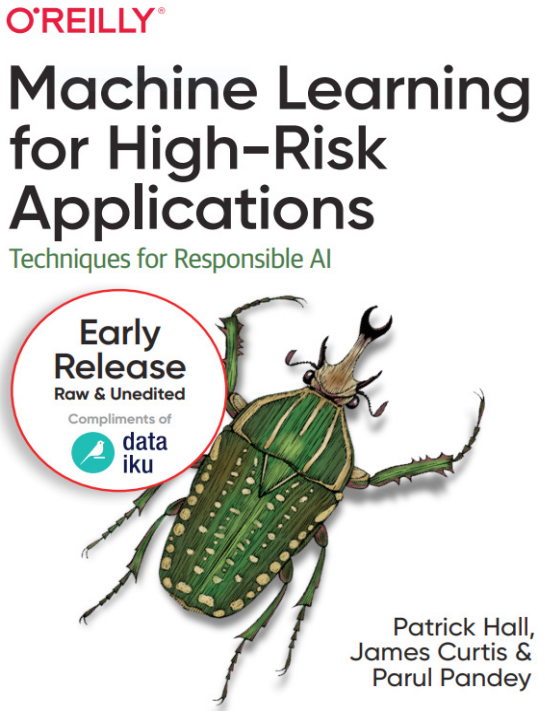
- الكفاءات التنظيمي والثقافية للذكاء الاصطناعي المسئول

- العمليات التنظيمية للذكاء الاصطناعي المسئول

٢. تصحيح أخطاء نظم التعلم الآلي للسلامة والأداء

- التدريب

- تصحيح أخطاء النموذج



- النشر

٣.خصوصية وأمن البيانات لتعلم الآلة

- أساسيات الأمن

- هجمات تعلم الآلة

- اهتمامات أمن الذكاء الاصطناعي العام

- التدابير المضادة

- دراسة حالة: هجمات التهرب في العالم الحقيقي

في الوقت الحالي، تعلم الآلة هو أكثر فروع

الانضباط قابلية للتطبيق تجاريا للذكاء الاصطناعي. نظم تعلم الالـة تستخدم لاتخاذ قـرارات عالية المخاطر في التوظيف، الكفالة، الإفراج المشروط، الإقراض وكثير من التطبيقات عالية المخاطرة الأخري في جميع أنحاء الاقتصاديات العالمية. وفي بيئة الشركات، نظم تعلم الآلة تستخدم في كل أجزاء التنظيم من المنتجات التى تواجه المستخدم إلي تقييم الموظف المختص، آلية المكتب الخلفي، وأكثر من ذلك.

في الواقع، العقد الماضي جاء معه اعتمادا واسعا لأساليب تعلم الآلة؛ لكنها أثبتت أيضا أن تعلم الآلة يعرض مخاطر لعمليتها وللمستهلكين أيضا. حيث أنه لسوء الحظ، مثل كل التكنولوجيات الأخري تقريبا، تعلم الآلة يمكن أن يفشل سواء عن طريق سوء الاستخدام غير المتعمد أو من خلال الأساءة المتعمدة. واعتبارا من الآن، صار هناك أكثر من ألف تقرير متاح عن التمييز الخوارزمي، انتهاك خصوصية البيانات، التدريب علي انتهاكات أمن البيانات وغيرها من الحوادث الضارة. مثل هذه المخاطر يجب أن تخفف من قبل المنظمات وعامة الناس حتى يمكن إدراك افوائد الحقيقية لهذه التكنولوجيات.

وبالفعل يتطلب هذا الفعل من قبل الممارسين اثناء التنظيم الذي يهدف هذا الكتاب الإلتزام به ، والذي بدأ يتشكل فعليا، علي الرغم من أن ممارسة تعلم الآلة ما زالت غير منظمة لحد كبير، كما يفتقر إلي معايير مهنية مقبولة. مما يعني أنه يعود الأمر علي الأفراد لحد كبير لتحميل أنفسهم المسئولية عن العالم الحقيقي. وعلي هذا الأساس، تعلم الآلة للتطبيقات علية المخاطرة سوف تسلح الممارسين معفهم قوي لعمليات الحوكمة النموذجية وطريقة جديدة لاستخدام أدوات لغة بايثون Python المشتركة لتدريب النماذج القابلة للتفسير وتصحيح الأخطاء منأجل قضايا الأداء، السلامة، الإنصاف، الأمن والخصوصية.

وقد تفرع هذا الكتاب لثلاث محاورين رئسية، المحور الأول يناقش القضايا من منظور تطبيق عملي مع اشتراطت النظرية عند الضرورة. ويستعرض حوكمة النموذج المعاص المرتبط بالالتزام القانوني الأساسي؛ الكفاءات التنظيمية والثقافية للذكاء الاصطناعي المسئول؛ مع دراسة حالة الموتت بواسطة المركبات المستقلة ذاتية القيادة. أما المحور الثاني الرئيسي عن تصحيح أخطاء نظم التعلم الآلي للسلامة والأداء المتفرع للتدريب؛ تصحيح أخطاء النموذج؛ والنشر. ويتضمن أمثلة لغة بايثون Python التي تخاطب المجالات المحددة في المحور الأول. أما المحور الثالث والأخير عن خصوصية وأمن البيانات لتعلم الآلة فقد ناقش موضوعات أساسيات الأمن؛ هجمات تعلم الآلة؛ اهتمامات أمن الذكاء الاصطناعي العام؛ مع استعراض دراسة حالة عن هجمات التهرب في العالم الحقيقي.