محاضرة 4: تعلّم مُراقَب وتعلّم غير مُراقَب

تمهيد

هناك طريقان رئيسيان لكي "يتعلّم" الحاسوب من البيانات. الطريق الأول يعتمد على أمثلة تحمل **إجابة صحيحة معروفة** مسبقًا؛ نعلّم النموذج أن يربط بين صفات المدخلات ونتيجة نهائية محدّدة. هذا هو **التعلّم المُراقَب**، وهو مناسب عندما نريد **تصنيفًا** (هل هذه الرسالة شكوى أم لا؟) أو **تنبؤًا** (من الأكثر عُرضة للتسرّب المدرسي؟). الطريق الثاني يعمل عندما **لا نملك إجابات جاهزة**، فنطلب من النموذج أن يبحث عن **بِنى خفية** في البيانات: مجموعات متشابهة، أبعاد أساسية تختصر التعقيد، أو نقاط شاذّة تستحق الانتباه. هذا هو **التعلّم غير المُراقَب**، وهو مناسب عندما نريد **استكشافًا وتنظيمًا وفهمًا أوّليًا** قبل وضع قرارات نهائية (Russell & Norvig, 2020; Mitchell, 2019). في علم الاجتماع، اختيار الطريق ليس تقنيًا فحسب؛ بل **منهجي وأخلاقي** أيضًا، لأنه يؤثر على من يستفيد ومن قد يتضرّر.(Crawford, 2021)

(1ما هو التعلّم المُراقَب؟ (لغة بسيطة وأمثلة قريبة(

التعلّم المُراقَب يشبه تدريب طالبٍ على حلّ مسائل ومعه **مفاتيح الإجابة** .لدينا جدول بيانات يحتوي على **مدخلات** (مثل العمر، المسافة إلى المدرسة، مستوى الدخل، درجات سابقة...) ومع كل صفّ **نتيجة صحيحة** نعرفها سلفًا (تسرّب/لم يتسرّب، أو شكوى/غير شكوى). نُري النموذج كثيرًا من هذه الأزواج، فيتعلّم نمطًا عامًا يساعده على **التوقّع** عندما يرى صفًا جديدًا دون إجابة.(James, Witten, Hastie, & Tibshirani, 2013)

أمثلة اجتماعية شائعة:

- **التنبؤ بالتسرّب المدرسي :**الهدف هو كشف المعرضين للخطر مبكرًا لتوجيه الدعم (Lazer et al., 2009).
- **تصنيف رسائل المواطنين :**فرز الرسائل الواردة إلى فئات مثل "شكوى خدمية"، "اقتراح"، "ملاحظة" لتسريع الاستجابة.
- **التنبؤ بالبطالة طويلة الأمد :**استخدام خصائص تعليمية وجغرافية لتحديد من يحتاج برامج تأهيل مركّزة.

نقاط قوة :يصلح لاتخاذ قرارات **عملية مباشرة** عندما تكون "الإجابة الصحيحة" موثوقة ومناسبة اجتماعيًا.
قيود ومخاطر:

- **جودة "الإجابات الصحيحة" نفسها :**إن كانت مُتحيّزة أو ضعيفة الدقة سننقل الانحياز إلى النموذج (O'Neil, 2016).
- **عدم توازن الفئات :**إذا كانت حالات التسرّب نادرة، قد يحقق النموذج "دقة" عالية لأنه يتجاهل القلّة المهمّة (He & Garcia, 2009).
- **تسرّب المعلومات :**إدخال مُلخّصات أو متوسّطات محسوبة من كامل البيانات قبل التقسيم بين التدريب والاختبار يزوّد النموذج بما لا يحقّ له رؤيته ويجعل التقييم مضلّلًا.(Arlot & Celisse, 2010)

(2كيف نقيم التعلّم المُراقَب بطريقة مسؤولة اجتماعيًا؟

الأرقام العامة لا تكفي وحدها. نحتاج قراءة **تفصيلية** مرتبطة بالغرض الاجتماعي.

أ. مصفوفة الأخطاء بلغة عادية

عند اتخاذ قرار "نعم/لا" (مثل: يحتاج تدخّل/لا يحتاج)، لدينا أربعة احتمالات بسيطة:

- اكتشفنا حالة تستحق التدخّل فعلًا) **صحيح إيجابي**.(
- لم نكتشف حالة تستحق التدخّل) **سلبي كاذب**(، وهو خطأ يفوّت حقًا.
- اعتبرنا شخصًا بحاجة للتدخّل وهو ليس كذلك) **إيجابي كاذب**(، وهو خطأ قد يستهلك موارد أو يضع وصمًا.
- تجاهلنا من لا يحتاج التدخّل أصلًا) **صحيح سلبي**.(اختيار الحدود بين "نعم/لا" يجب أن **يعكس أولوياتك الاجتماعية** :أحيانًا نفضّل التقاط معظم المحتاجين حتى لو زادت بعض الإنذارات الكاذبة.(Lipton, 2018)

ب. مؤشرات مفهومة

- **الاستدعاء (الحساسية) :**كم نسبة المستحقين الذين التقطناهم؟ مهمّ عندما يكلّفنا فقدان شخص مستحق ثمنًا اجتماعيًا كبيرًا.
- **جودة الإشارة (الدقّة الإيجابية) :**من الذين قلنا إنهم يحتاجون تدخّلًا، كم منهم فعلًا يحتاج؟ مهم عندما تكون الموارد محدودة.

- **التوازن بينهما (درجة توفيقية مثل "إف-1") :**مقياس يلخّص التوازن عندما نهتم بالاثنين معًا (Sokolova & Lapalme, 2009).
- **المعايرة :**عندما يقول النموذج "احتمال 70%"، هل يعني ذلك تقريبًا سبعة من كل عشرة فعلًا؟ المعايرة الجيدة تسهّل وضع سياسات عادلة.(Niculescu-Mizil & Caruana, 2005)
- **عدالة الأداء بين الفئات :**هل تختلف المؤشرات السابقة بوضوح بين فئات النوع الاجتماعي أو المناطق أو الوضع الاقتصادي؟ إن حدث ذلك، نراجع التصميم أو نجري تصحيحات واعية (Hardt, Price, & Srebro, 2016; Suresh & Guttag, 2021).

ج. منحنى الحساسية مقابل الإنذارات الكاذبة

بدل رقم واحد، يمكننا تغيير حدّ القرار ورسم العلاقة بين التقاط المستحقين وزيادة الإنذارات الكاذبة، ثم اختيار نقطة مناسبة لغرضنا .(Fawcett, 2006) هذه الطريقة توضّح **المقايضة** بدل إخفائها.

(3ما هو التعلّم غير المُراقَب؟ (استكشافٌ قبل القرار(

عندما لا نملك "إجابات صحيحة"، نستخدم أدوات تنظر إلى **التشابه والهيكل الداخلي** للبيانات. الهدف هو **الفهم والتنظيم** :ما المجموعات المتشابهة؟ ما الأبعاد القليلة التي تُلخّص معظم التنوّع؟ هل هناك حالات شاذّة تستحق فحصًا يدويًا؟

أدوات رئيسية بلغة مبسّطة:

- **التجميع إلى عناقيد (عنقدة) :**تجميع الأفراد أو الأحياء أو المدارس في مجموعات متشابهة داخليًا ومختلفة خارجيًا—مثل تقسيم أحياء المدينة إلى أنماط سكنية وخدمية مختلفة (Kaufman & Rousseeuw, 2009).
- **اختزال الأبعاد (تحليل المكوّنات الرئيسية) :**تبسيط عدد المتغيّرات إلى محاور قليلة تلخّص أغلب التغيّر، ما يساعد على الرسم والفهم الأولي.(Jolliffe & Cadima, 2016)
- **نمذجة الموضوعات في النصوص :**استخراج موضوعات متكرّرة من مجموع مقابلات أو منشورات لفهم الثيمات السائدة، مع مراجعة بشرية دقيقة للمعنى والسياق.(Blei, 2012)

- **رصد الشذوذ :**تحديد حالات "مختلفة جدًّا" عن الباقي، قد تشير إلى أخطاء أو تجارب نادرة تستحق دراسة متعمّقة.

نقاط قوة :تمهّد الطريق للفهم، تقترح فرضيات جديدة، تكشف بنى غير متوقعة. **قيود ومخاطر :**النتائج **اقتراحات تنظيمية** لا "حقيقة نهائية". قد تُغري بتثبيت **صور نمطية** حول مجموعات ("هذا الحيّ من النوع كذا") إذا لم نراجعها اجتماعيًا وأخلاقيًا.(Crawford, 2021)

(4كيف نقيم التعلّم غير المُراقَب دون "إجابة صحيحة"؟

- **مدى التماسك الداخلي والانفصال بين المجموعات :**هل أفراد كل مجموعة متقاربون، والمجموعات متباعدة منطقيًا؟ يوجد مؤشّر كلاسيكي يسمى **"درجة الظلّ** (Silhouette) **"**يقيس ذلك رقميًا، لكن يمكننا أيضًا فحص الأمثلة يدويًا لنتأكد أن التقسيم **يعني شيئًا اجتماعيًا** .(Rousseeuw, 1987)
- **اختبار حساسية عدد المجموعات :**نجرّب أعدادًا مختلفة ونراقب متى تتوقّف الفائدة عن التحسّن ملحوظًا ("طريقة الكوع")، ثم نتحقّق من التفسير الاجتماعي لكل تقسيم.(Kaufman & Rousseeuw, 2009)
- **الاستقرار :**إذا أخذنا عيّنات مختلفة من البيانات، هل يعطينا التقسيم نتائج متشابهة؟ الاستقرار علامة على بنية حقيقية لا على "ضجيج" عشوائي.
- **تحقّق خارجي :**نقارن المجموعات بمتغيّرات لم تدخل في الحساب (مثل مؤشرات خدمات أو نتائج لاحقة). إن رأينا فروقًا منطقية وثابتة، يزيد ذلك ثقتنا.

(5متى أختار أي نهج؟ (دليل قرار بسيط(

- **لديك هدف قرار واضح ومخرجات تؤثّر على الناس الآن** (توجيه دعم، أولوية خدمة): استخدم **التعلّم المُراقَب** بشرط أن تكون "الإجابات الصحيحة" موثوقة وتراعي العدالة، ومع تقييم مفصّل وعرض مؤشرات لكل فئة.
- **لا تزال تفهم المشكلة وتستكشف بنية الظاهرة** (أنماط أحياء، شرائح طلبة، موضوعات في النصوص): ابدأ بـ **غير المُراقَب** لاكتشاف محاور أولية، ثم ارجع إلى الميدان أو الخبراء للتثبّت والشرح.

- **الوسيط الذكي :**كثيرًا ما نبدأ بغير المُراقَب لتكوين الحدس، ثم نبني باستخدام المُراقَب قرارًا مُعلنًا وشفافًا. المهم أن **لا نخلط بين الاستكشاف والقرار** :التجميع ليس تصنيفًا حُكميًا ولا أساسًا تلقائيًا للسياسات (Mitchell, 2019).

(6أخلاقيات ومزالق يجب التنبه لها في المسارين

- **عدالة "الإجابات الصحيحة :"**إذا كانت سجلات "المخالفات" أو "الشكاوى الموثّقة" نفسها متحيّزة تاريخيًا ضد فئة ما، سينقل المُراقَب هذا الظلم إلى المستقبل .(O'Neil, 2016) الحلّ: مراجعة بشرية للمصادر، وتعريفات منصفة لما نعدّه "نجاحًا" أو "فشلًا."

- **تثبيت الصور النمطية في التجميع :**عندما نسمي المجموعات بأسماء مبسّطة ("حيّ فقير—غير مندمج") نُخاطر بتجريد الناس من سياقهم. الحلّ: أسماء وصفية محايدة، ومراجعة مع أصحاب العلاقة قبل الاستخدام.

- **الخصوصية والسريّة :**لا ينبغي إدخال بيانات تعريفية حساسة في أدوات مساعدة أو مشاركتها خارج الضوابط الأخلاقية.(Floridi et al., 2018)

- **الشفافية :**وثائق علنية واضحة للبيانات (بطاقة بيانات) وللنموذج (بطاقة نموذج) توضّح الجمهور المقصود والحدود ومؤشرات الأداء لكل فئة.(Mitchell et al., 2019; Gebru et al., 2021)

(7حالتان دراسيتان مبسّطتان

الحالة الأولى — دعم مبكر ضد التسرّب

الهدف عملي وواضح: توجيه موارد دعم محدودة. نستخدم تعلّمًا مُراقَبًا مع تعريف دقيق للتسرّب من السجلات، ونقسّم البيانات إلى تدريب واختبار، ونعلن مؤشرات **الاستدعاء** و**جودة الإشارة** لكل فئة (جندر/منطقة)، ونضبط حدّ القرار بما يوازن بين التقاط المستحقين وإدارة الموارد. ننشر بطاقة نموذج تبيّن أين **لا** ينبغي استخدامه (Hardt et al., 2016; Lipton, 2018).

الحالة الثانية — فهم أنماط السكن والخدمات

الهدف استكشافي: ما أنماط الأحياء في مدينة كبيرة؟ نستخدم التجميع لاكتشاف مجموعات أحياء متشابهة في الكثافة السكانية، الأسعار، القرب من الخدمات. نقيم التماسك داخليًا ونسأل: هل لهذه المجموعات **معنى اجتماعي؟** ثم نعود

لأصحاب المصلحة المحليين لتعديل التعريفات قبل أن نقترح سياسات عامة (Kaufman & Rousseeuw, 2009; Crawford, 2021).

(8كيف نستفيد من أدوات الكتابة الذكية دون الوقوع في الفِخاخ؟

- لطلب **مساعدة تنظيميّة** (لا قرار نهائي): صِغ أمرًا واضحًا يطلب "جدولًا لمحاور أولية + أمثلة" أو "عناصر لروبرك تقييم"، ثم **راجِع يدويًا** أي ادّعاء واقعي قبل الاستشهاد.(Bender et al., 2021)
- لتوثيق عملك: احتفظ **بالأوامر والمخرجات** كملحق منهجي يشرح كيف وصلت إلى النتائج، وما الذي رفضته ولماذا.(Mitchell, 2019; Raji et al., 2020)

خاتمة

التعلّم المُراقَب مفيد عندما نعرف ما نبحث عنه ونحتاج قرارًا واضحًا وعادلاً الآن. التعلّم غير المُراقَب مفيد عندما نريد فهمًا أوليًا منظمًا لبنية الظاهرة قبل أن نُصدر حُكمًا. في الحالتين، **المنهجية والشفافية** هما الضمان: قياس منصف، بيانات ممثّلة، تقييم يراعي الفئات المختلفة، وتسميات لا ترسّخ صورًا نمطية. بهذه الروح، تصبح الخوارزميات **مساعدة للخبرة الاجتماعية** بدل أن تطغى عليها (Mitchell, 2019; Russell & Norvig, 2020; Suresh & Guttag, 2021).

قائمة المراجع

1. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79.
2. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *Proceedings of FAccT 2021*, 610–623.
3. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. Crawford, K. (2021). *Atlas of AI*. Yale University Press.
4. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

   Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework. *Minds and Machines, 28*(4), 689–707.

5. Gebru, T., Morgenstern, J., Vecchione, B., et al. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92.
6. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *NeurIPS 29*, 3315–3323.
7. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE TKDE, 21*(9), 1263–1284.
8. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
9. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review. *Philosophical Transactions A, 374*(2065), 20150202.
10. Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. Wiley.
11. Lipton, Z. C. (2018). The mythos of model interpretability. *Queue, 16*(3), 31–57. Lazer, D., Pentland, A., Adamic, L., et al. (2009). Computational social science. *Science, 323*(5915), 721–723.
12. Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
13. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *ICML*, 625–632.
14. O'Neil, C. (2016). *Weapons of Math Destruction*. Crown. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.
15. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
16. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.
17. Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the ML life cycle. *EAAMO '21*.