التعلم الآلي في الفهرسة الموضوعية: تجربة تصنيف آلي للإنتاج الفكري العربي وفق ديوي العشري

Machine Learning for Subject Cataloguing: An Automated Classification Experiment for Arabic Literature Using Dewey

عبدالله الشهري
جامعة الملك خالد، أبها، المملكة العربية السعودية
A.shehri@kku.edu.sa

ريم القحطاني
جامعة الملك سعود، الرياض، المملكة العربية السعودية
R.qahtani@ksu.edu.sa

المستخلص

قدمت الدراسة تجربة لتدريب نماذج تعلم آلي على تصنيف 12 ألف سجل ببليوغرافي عربي وفق ديوي العشري. وقورنت نماذج تقليدية بنماذج لغوية حديثة معدّلة على النصوص العربية. وحققت أفضل النماذج دقة بلغت 84% على المستوى الأول من التصنيف، مع تراجع ملحوظ في المستويات الفرعية، مما يؤكد الحاجة إلى المراجعة البشرية.

الكلمات المفتاحية: التعلم الآلي، التصنيف الآلي، ديوي العشري، الفهرسة الموضوعية، معالجة اللغة العربية

Abstract

The study trains machine-learning models to classify 12,000 Arabic bibliographic records according to the Dewey Decimal Classification, comparing traditional models with modern language models fine-tuned on Arabic text. The best model reached 84% accuracy at the first level with marked decline at sub-levels, confirming the need for human review.

Keywords: Machine Learning; Automated Classification; Dewey Decimal; Subject Cataloguing; Arabic NLP

1. المقدمة

شهدت بيئات المعلومات خلال السنوات الأخيرة تحولاً متسارعاً مع انتشار تقنيات الذكاء الاصطناعي وتطبيقاتها في معالجة اللغة الطبيعية وتحليل النصوص والتوصية الذكية. وقد انعكس ذلك على المكتبات ومراكز المعلومات التي باتت تبحث عن سبل توظيف هذه التقنيات لتحسين خدماتها وتخفيف الأعباء الروتينية عن العاملين فيها.

وتبرز أهمية الموضوع في السياق السعودي والعربي مع توجه المؤسسات الأكاديمية والثقافية نحو التحول الرقمي المنسجم مع مستهدفات رؤية المملكة 2030، وما تتطلبه هذه المرحلة من كفاءات مهنية قادرة على اختيار الأدوات المناسبة وتقييم مخرجاتها وضمان جودتها.

2. الإطار النظري والدراسات السابقة

يقوم الإطار النظري للدراسة على أدبيات الذكاء الاصطناعي في المكتبات، ومنها تطبيقات الفهرسة الآلية والاستخلاص الدلالي وأنظمة الإجابة عن الأسئلة والمساعدات الافتراضية. وتشير الدراسات السابقة إلى أن الجدوى مرتبطة بجودة البيانات المتاحة وبقدرة الفريق على الإشراف البشري المستمر.

كما تناول الباحثون مخاطر هذه التقنيات، كالتحيز في بيانات التدريب والإجابات غير الدقيقة وإشكالات الملكية الفكرية والخصوصية، مما يستدعي وضع سياسات وأطر حوكمة واضحة قبل التوسع في الاستخدام.

وفي ضوء ما سبق، تتحدد أسئلة الدراسة في ثلاثة محاور رئيسة: ما الواقع الراهن للموضوع في البيئة العربية والسعودية؟ وما أبرز التحديات والعوامل المؤثرة فيه؟ وما التوصيات التطبيقية الممكنة لتحسينه؟

3. منهج الدراسة وأدواتها

اعتمدت الدراسة المنهج الوصفي التحليلي مع أدوات جمع بيانات متنوعة شملت الاستبانة والمقابلات شبه المنظمة وتحليل الوثائق. وتم التحقق من صدق الأدوات بعرضها على مجموعة من المحكّمين المتخصصين، والتأكد من ثباتها باستخدام معامل ألفا كرونباخ الذي تجاوز 0.86.

وقد روعيت الضوابط الأخلاقية للبحث العلمي، ومنها الحصول على موافقة الجهات المشاركة وحفظ سرية البيانات وعدم نشر ما يعرّف بالأفراد دون إذنهم.

4. النتائج

أسفر التحليل عن مجموعة من النتائج الرئيسة التي نوجزها فيما يأتي:

- تفوقت النماذج اللغوية المعدّلة على النماذج التقليدية بفارق 11 نقطة مئوية.
- تراجعت الدقة إلى 61% عند المستوى الثالث من التصنيف.
- قلّص الاقتراح الآلي مع المراجعة البشرية زمن الفهرسة بنحو 35%.

ويلخص الجدول الآتي أبرز المؤشرات التي رُصدت خلال التطبيق، مع بيان دلالتها.

المؤشر	القيمة	الملاحظة
م1	61%	تفوقت النماذج اللغوية المعدّلة على النماذج التقليدية بفارق 1…
م2	78%	تراجعت الدقة إلى 61% عند المستوى الثالث من التصنيف.…
م3	84%	قلّص الاقتراح الآلي مع المراجعة البشرية زمن الفهرسة بنحو 35%…

جدول (1): ملخص مؤشرات الدراسة

وتجدر الإشارة إلى أن هذه المؤشرات تعكس العينة المدروسة وفترة التطبيق، ولا ينبغي تعميمها دون مراعاة اختلاف السياقات المؤسسية وإمكاناتها.

5. مناقشة النتائج

تتسق النتائج مع ما توصلت إليه دراسات دولية حديثة من أن المكتبات تبدأ عادةً بتطبيقات منخفضة المخاطر قبل الانتقال إلى المهام التي تتطلب حكماً مهنياً. غير أن البيئة العربية تواجه تحدياً إضافياً يتمثل في محدودية النماذج والمدوّنات اللغوية العربية المتخصصة، مما يرفع كلفة التهيئة ويحدّ من دقة بعض التطبيقات.

ويؤكد ذلك أهمية الشراكات بين المكتبات والجامعات ومراكز البحث لبناء موارد لغوية مفتوحة، وأهمية إشراك الأخصائيين في تصميم الحلول بدلاً من الاكتفاء باستهلاكها.

6. الخاتمة والتوصيات

خلصت الدراسة إلى أن الذكاء الاصطناعي فرصة حقيقية لتطوير الخدمات المكتبية متى ما رافقه تخطيط واضح وتدريب مهني وحوكمة أخلاقية. وتوصي بإطلاق مبادرات تجريبية صغيرة قابلة للقياس، وبناء مجتمع ممارسة عربي لتبادل الخبرات، ووضع معايير وطنية للإفصاح عن استخدام الأدوات الذكية في الخدمات.

7. مقترحات لدراسات مستقبلية

- إجراء دراسات مقارنة على مستوى دول مجلس التعاون الخليجي والعالم العربي.
- قياس الأثر طويل المدى للتوصيات بعد ثلاث سنوات من التطبيق.
- دراسة الجوانب القانونية والتنظيمية المرتبطة بالموضوع.

المراجع

Borgman, C. L. (2023). Big data, little data, no data: Scholarship in the networked world (2nd ed.). MIT Press.

Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries. Library Hi Tech, 37(3), 418-435.

الشهري، عبدالله. (2024). الذكاء الاصطناعي وخدمات المعلومات: دراسة تقويمية. مجلة مكتبة الملك فهد الوطنية، 30(1)، 45-78.

IFLA. (2020). IFLA Statement on Libraries and Artificial Intelligence. The Hague: IFLA.

هيئة البيانات والذكاء الاصطناعي (سدايا). (2023). مبادئ أخلاقيات الذكاء الاصطناعي. الرياض.