CONCEPT

أنظمة التوصية والتخصيص: التصفية التعاونية والتضمينات المتجهية وترتيب النتائج

أنظمة التوصية والتخصيص: التصفية التعاونية والتضمينات المتجهية وترتيب النتائج

مقدمة:

تحدد أنظمة التوصية والتخصيص ما يراه العميل تاليا في المتجر الإلكتروني ومنصة المحتوى المرئي وتطبيق البنك، لكن مؤسسات كثيرة ما زالت تعتمد على قوائم الأكثر رواجا أو على أدوات جاهزة من موردين لا يستطيع أحد تقييمها أو ضبطها أو تفسير نتائجها. وتزيد سجلات التفاعل المتناثرة والمستخدمون والمنتجات الجدد المشكلة تعقيدا. تدرب هذه الدورة من كور كونسبت علماء البيانات ومهندسي تعلم الآلة على بناء محركات التوصية وقياسها وتشغيلها بلغة Python، من التصفية التعاونية إلى الاسترجاع بنماذج البرجين وترتيب النتائج. ويعد المشاركون دفتر نموذج أولي لمحرك التوصية وتقرير التقييم غير المتصل على مجموعة بيانات تفاعل واقعية.

أهداف الدورة:

- صياغة مشكلة التخصيص في التجزئة أو الإعلام أو الخدمات المالية بوصفها مهمة استرجاع وترتيب لها هدف أعمال قابل للقياس
- تجهيز التقييمات الصريحة وسجلات التفاعل الضمني في مصفوفات مستخدم ومنتج متناثرة مع تقسيم زمني خال من تسرب البيانات
- بناء نماذج توصية قائمة على الجوار وتحليل المصفوفات والمحتوى والنماذج الهجينة بلغة Python ومقارنة سلوكها
- تقييم القوائم المرتبة دون اتصال بمقاييس precision@k recall@kg MAPg وNDCG إلى جانب مقاييس التغطية والجدة والتنوع
- تصميم تجربة حية ومسار خدمة منخفض زمن الاستجابة بفهارس التضمينات ومخازن السمات والتخزين المؤقت
- تقدير مخاطر البداية الباردة وتحيز الشهرة والخصوصية والإنصاف وتحديد ضوابطها قبل الإطلاق

الفئة المستهدفة:

- وظائف علم البيانات التي تنمذج سلوك العملاء وتتولى تجارب التخصيص
- وظائف هندسة تعلم الآلة التي تدرب نماذج الترتيب وتنشرها وتراقبها في بيئة التشغيل
- وظائف تحليلات المنتجات الرقمية التي تحدد أهداف التوصية وتقرأ نتائج التجارب
- وظائف تحليلات التسويق المسؤولة عن منطق العرض التالي الأنسب واستهداف البيع المتقاطع
- وظائف منصات البيانات التي تدير خطوط الأحداث ومخازن السمات وفهارس المتجهات المغذية لمحركات التوصية

محاور الدورة:

اليوم 1: حالات استخدام التوصية وبيانات التفاعل

- حالات استخدام التوصية: سلة التجزئة وبث المحتوى والعرض المصرفي التالي الأنسب
- التقييمات الصريحة مقابل الإشارات الضمنية: النقرات ومدة المكوث والمشتريات
- بناء مصفوفة المستخدم والمنتج المتناثرة بمكتبتي SciPy و pandas
- نماذج الأساس القائمة على الشهرة والحداثة مرجعا للمقارنة
- التقسيم الزمني لبيانات التدريب والاختبار وفحص تسرب البيانات في سجلات التفاعل

اليوم 2: بنى التصفية التعاونية والتصفية بناء على المحتوى والنماذج الهجينة

- التصفية بأقرب الجيران k-NN على مستوى المستخدم والمنتج بتشابه جيب التمام
- تحليل المصفوفات بطريقة المربعات الصغرى المتناوبة ALS للتغذية الراجعة الضمنية
- التصفية بناء على المحتوى بأوزان TF-IDF ومتجهات خصائص المنتج
- تصاميم التوصية الهجينة: الترجيح والتبديل ودمج السمات
- بنية سلسلة توليد المرشحين ثم التسجيل ثم إعادة الترتيب

اليوم 3: التضمينات المتجهية ونماذج التوصية العميقة وتقييم الترتيب

- تضمينات المستخدمين والمنتجات المتعلمة من تسلسلات التفاعل
- نظرة عامة على نماذج الاسترجاع ثنائية البرج وشبكات الترتيب العميقة
- دوال هدف تعلم الترتيب: النقطية والزوجية وعلى مستوى القائمة
- مقاييس الترتيب غير المتصلة: Precision@k Recall@k و MAP و NDCG
- مقاييس ما بعد الدقة: تغطية الكتالوج والجدة والمصادفة السارة

اليوم 4: البداية الباردة والتحيز والاختبار الحي والخدمة الفورية

- معالجة مشكلة البداية الباردة: الافتراضات المسبقة من الخصائص وأسئلة التهيئة وخوارزميات القائمة المتعددة الأذرع
- تحيز الشهرة وفقاعات التصفية وإعادة الترتيب للتنوع بطريقة Maximal Marginal Relevance
- تصميم اختبار A/B الحي: مقاييس الحماية وحجم العينة وأثر الجدة
- الخدمة الفورية: مخزن السمات وفهرس الجار الأقرب التقريبي والتخزين المؤقت للاستجابات
- التخصيص المحافظ على الخصوصية وتدقيق الإنصاف لفئات المزودين والمستخدمين

اليوم 5: مختبر بناء محرك التوصية وتقرير التقييم

- مختبر التجارة الإلكترونية: نموذج ALS للتغذية الراجعة الضمنية مع بديل هجين قائم على المحتوى
- مختبر كتالوج المحتوى الإعلامي: الاسترجاع بالتضمينات مع فهرس الجار الأقرب التقريبي
- لوحة مقارنة النماذج وفق NDCG والتغطية والتنوع
- خطة التجربة وورقة تصميم الخدمة للإطلاق في بيئة التشغيل
- مراجعة الزملاء لدفتر النموذج الأولي لمحرك التوصية وتقرير التقييم غير المتصل

المهارات التي ستكتسبها:

- صياغة مشكلات التوصية
- هندسة بيانات التفاعل
- نمذجة التصفية التعاونية
- الاسترجاع القائم على التضمينات
- تقييم الترتيب
- تصميم التجارب الحية
- خدمة النماذج منخفضة زمن الاستجابة
- التخصيص المسؤول

لماذا تحضر هذه الدورة:

- العودة بدفتر نموذج أولي لمحرك التوصية وتقرير تقييم غير متصل جرى بناؤهما ومقارنتهما في المختبرات
- معرفة عائلة النماذج المناسبة للسجلات المتناثرة أو البيانات الوصفية الغنية أو الكتالوجات الجديدة قبل تخصيص جهد هندسي
- اكتشاف النتائج غير المتصلة المضللة الناتجة عن تسرب البيانات أو تحيز الشهرة أو اختيار مقياس خاطئ قبل وصولها إلى اختبار A/B
- مقارنة تصاميم محركات التوصية مع ممارسين من فرق التجزئة والإعلام والاتصالات والمصارف

الخاتمة:

يكسب محرك التوصية الثقة حين تكون بياناته ونموذجه وتقييمه قابلة للفحص والتكرار. تنتقل الدورة من حالات الاستخدام وبيانات التفاعل، إلى طرق الجوار وتحليل المصفوفات والتصفية بناء على المحتوى والتصاميم الهجينة، ثم إلى التضمينات والاسترجاع ثنائي البرج ومقاييس الترتيب، فالبداية الباردة والتحيز والاختبار الحي والخدمة الفورية والخصوصية. ويخصص اليوم الأخير لمختبرات بناء عملية تنتهي بدفتر نموذج أولي لمحرك التوصية وتقرير تقييم غير متصل يمكن للفرق تطويرهما نحو بيئة التشغيل.