

📌 سلسلة معالجة اللغات الطبيعية (NLP Pipeline)

في هذا الدرس، نتعلم كيف يحول بايثون النص الذي يكتبه البشر إلى بيانات يفهمها الحاسوب. بالنسبة للأخصائي النفسي، هذه العملية تشبه تماماً "تفريغ المقابلات وتصنيفها"، لكنها تتم هنا آلياً وبسرعة فائقة لتوفير الجهد والوقت.

📦 مكتبة NLTK: الحقيبة الأدوات للأخصائي الرقمي

بايثون لا تأتي بكل الأدوات مدمجة، لذا نعلم على مكتبة NLTK (Natural Language Toolkit). يمكن تشبيه هذه المكتبة بحقيبة أدوات تحتوي على:

- **مقصات:** لتقطيع الجمل والكلمات.
- **مساحات:** لحذف الكلمات الزائدة والضجيج اللغوي.
- **قواميس:** لفهم أصول الكلمات وتصنيفها.

🧼 مرحلة التنظيف (Preprocessing): تحضير النص للتحليل

النص البشري يحتوي على "ضجيج" لغوي يعيق فهم الحاسوب للمعنى النفسي العميق، لذا نمر بثلاث خطوات أساسية:

1. **التقطيع (Tokenization):** تقسيم النص إلى وحدات صغيرة (كلمات) ليتمكن الحاسوب من عدها وتحليل كل منها على حدة.
2. **إزالة كلمات الوقف (Stop Words):** حذف الكلمات المتكررة التي لا تحمل قيمة تشخيصية (مثل: من، إلى، في) للتركيز فقط على الكلمات ذات الشحنة الانفعالية.
3. **التوحيد والنمذجة (Stemming):** رد الكلمات إلى أصولها؛ فكلما مثل (يذاكر، تذاكر، مذاكرة) تُجمع تحت أصل واحد (ذاكر) لمنع تشتت الإحصائيات.

📊 التحليل النوعي والبصري: عندما نتحدث البيانات

بمجرد تنظيف النص، نبدأ في استخراج المؤشرات النفسية عبر أداتين قويتين:

- **تحليل تكرار الكلمات:** اكتشاف "الموضوعات المهيمنة"؛ فإذا تكررت كلمات مثل (محبط، فشل، صعب) في رسائل الطلاب، فهذا مؤشر مباشر على حالتهم الوجدانية.
- **سحابة الكلمات (Word Cloud):** أداة بصرية تظهر فيها الكلمات الأكثر تكراراً بحجم أكبر. هي وسيلة مثالية لعرض المشكلات النفسية والتربوية بشكل سريع وقوي أمام غير المتخصصين كمديري المدارس.

💻 تطبيق عملي: تحليل ملاحظة تربوية

عند تحليل ملاحظة مثل: "أحمد طالب ذكي لكنه يواجه صعوبة في التركيز، يعاني أحمد من تشتت مستمر":

- باستخدام كود بايثون البرمجي، نقوم باستدعاء مكتبة NLTK.
- تتم عملية التقطيع وإزالة كلمات الوقف العربية.
- **النتيجة:** نحصل على قائمة تحتوي فقط على الجوهر (أحمد، طالب، ذكي، يواجه، صعوبة، التركيز، يعاني، تشتت، مستمر).

⚠ تحديات معالجة اللغة العربية في علم النفس

يجب على الباحث الوعي بخصوصية اللغة العربية التي تفرض تحديات تقنية:

- **التشكيل:** كلمة (عَلِمَ) تختلف عن (عَلَّمَ)، ويجب توحيدها تقنياً لضمان دقة التحليل.
- **اللواصق:** العربية غنية بالسوابق واللواحق (مثل: وبالوالدين)، مما يتطلب أدوات متقدمة لقص الزوائد.
- **اللهجات:** ميل الطلاب للاستجابة بالعامية يتطلب قواميس خاصة غير متوفرة دائماً بشكل قياسي.

💡 **الاستنتاج:** إن قوة بايثون لا تكمن فقط في توفير الوقت، بل في حماية الباحث من "التحيز الذاتي"؛ فبينما قد يميل الباحث للاهتمام بكلمة سمعها، يخبره الحاسوب بالكلمة التي تكررت فعلياً بدقة رياضية لا تقبل الجدل.