المحاضرة 3

الدرس الثالث: أدوات معالجة النصوص بايثون - التجهيز والتحليل النوعي

في هذا الدرس، سنركز على سلسلة المعالجة (NLP Pipeline)، وهي الخطوات التي يمر بها النص الخام ليتحول إلى بصيرة تربوية. بالنسبة لباحث علم النفس، هذه المرحلة تشبه عملية تفريغ المقابلات وتصنيفها، لكننا هنا سنعلم الحاسوب القيام بها آلياً وبسرعة فائقة.

1. مكتبة NLTK: الحقيبة الأدواتية للأخصائي الرقمي

قبل أن نبدأ، يجب أن نعرف أن بايثون لا تأتي بكل شيء مدمج؛ بل نستخدم مكتبات (Libraries) جاهزة. أهم مكتبة لنا هي .(Natural Language Toolkit) **NLTK**

تخيل NLTK كأنها حقيبة أدوات تحتوي على مقصات لتقطيع الجمل، ومساحات لمسح الكلمات الزائدة، وقواميس لفهم أصول الكلمات.

2. مرحلة التنظيف (Preprocessing): تحضير النص للتحليل

النص الذي نكتبه كبشر يحتوي على الكثير من الضجيج الذي لا يحتاجه الحاسوب لفهم المعنى النفسي. لذا نمر بالخطوات التالية:

أ. التقطيع (Tokenization)

هي عملية تقسيم النص الكبير إلى وحدات صغيرة (كلمات أو جمل).

لماذا؟ الحاسوب لا يرى الجملة كفكرة، بل يراها كسلسلة أحرف. بالتقطيع، نحولها إلى قائمة من الكلمات المستقلة التي يمكننا عدها وتحليلها.

ب. إزالة كلمات الوقف (Stop Words Removal)

في كل لغة، هناك كلمات تتكرر بكثرة ولكنها لا تحمل قيمة تشخيصية كبيرة (مثل: من، إلى، على، في).

في سياقنا: إذا حللنا مقالاً عن القلق الدراسي، سنجد كلمة في تكررت 100 مرة، وكلمة قلق 20 مرة. بدون إزالة كلمات الوقف، سيظن الحاسوب أن البحث يتحدث عن حرف الجر في!.

ج. التوحيد والنمذجة (Normalization & Stemming)

الحاسوب يرى يذاكر، تذاكر، مذاكرة ككلمات مختلفة تماماً. نحن نستخدم أدوات تعيد هذه الكلمات إلى أصلها (ذكر) ليتمكن الحاسوب من جمعها تحت مفهوم واحد.

الخطوة	الوصف التقني	الفائدة في علم النفس التربوي
التقطيع	تحويل الفقرة إلى كلمات منفصلة	عد تكرار المصطلحات السلوكية في المقابلة
إزالة كلمات الوقف	حذف حروف الجر والوصل	التركيز على الكلمات التي تحمل شحنة انفعالية
إزالة علامات الترقيم	حذف الفواصل والنقاط	منع تداخل العلامات مع الكلمات (مثلاً: قلق. تُعامل ك قلق)
التوحيد (Stemming)	رد الكلمات لأصولها	جمع كل مشتقات العنف أو التنمر في إحصائية واحدة

3. تحليل تكرار الكلمات (Word Frequency Analysis)

هذا هو قلب التحليل النوعي المحوسب. من خلال عد الكلمات الأكثر تكراراً، يمكننا اكتشاف الموضوعات المهيمنة (Dominant Themes) في استجابات الطلاب.

مثال: إذا قمت بتحليل 500 رسالة كتبها طلاب يعانون من صعوبات التعلم، ووجدت أن كلمة محبط وصعب وفشل هي الأكثر تكراراً، فهذا يعطيك مؤشراً مباشراً على الحالة الوجدانية العامة لهذه الفئة.

4. سحابة الكلمات (Word Cloud): عندما تتحدث البيانات بصرياً

تعد سحابة الكلمات من أجمل وأبسط وسائل عرض النتائج النفسية للجمهور غير المتخصص (مثل المعلمين أو أولياء الأمور).

هي صورة تظهر فيها الكلمات الأكثر تكراراً بحجم أكبر، والكلمات الأقل تكراراً بحجم أصغر.

التطبيق: بدلاً من تقديم جدول إحصائي جاف لمدير المدرسة، قدم له سحابة كلمات توضح أهم مشكلات الطلاب في ساحة المدرسة؛ ستكون الرسالة أسرع وأقوى.

5. تطبيق عملي (بايثون لتحليل ملاحظات المعلم)

لنفترض أن لدينا ملاحظة كتبها معلم عن طالب: *أحمد طالب ذكي لكنه يواجه صعوبة في التركيز، يعاني أحمد من تشتت مستمر.* سنستخدم بايثون لتنظيف هذا النص:

Python

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

# النص المراد تحليله
text = أحمد طالب ذكي لكنه يواجه صعوبة في التركيز، يعاني أحمد من تشتت مستمر.

# 1. التقطيع
tokens = word_tokenize(text)

# 2. إزالة كلمات الوقف (العربية)
stop_words = set(stopwords.words('arabic'))
filtered_text = [w for w in tokens if not w in stop_words]

print(filtered_text)
# المخرجات المتوقعة: ['أحمد', 'طالب', 'ذكي', 'يواجه', 'صعوبة', 'التركيز', 'يعاني', 'أحمد', 'تشتت', 'مستمر']
```

بإزالة الكلمات غير الضرورية، أصبح لدينا جوهر الملاحظة التربوية.

6. تحديات تحليل النصوص العربية في علم النفس

اللغة العربية غنية ومعقدة، ويجب أن يدرك طالب علم النفس بعض التحديات التقنية:

التشكيل: كلمة عِلم تختلف عن عَلَم، والحاسوب قد يخلط بينهما إذا لم نقم بعملية إزالة التشكيل لتوحيد التحليل.

اللواصق: العربية تعتمد على السوابق واللواحق (مثل: وبالوالدين أصلها والد). نحتاج لأدوات متقدمة في بايثون لقص هذه الزوائد.

اللهجات: استجابات الطلاب قد تكون بالعامية، مما يتطلب قواميس خاصة لا تتوفر دائماً في المكتبات القياسية.

الاستنتاج والتكليف الدراسي

يا أبنائي، إن أداة تحليل النصوص في بايثون هي قوتكم الخارقة القادمة. فهي لا توفر الوقت فحسب، بل تحميكم من التحيز الذاتي؛ فالباحث قد يميل للاهتمام بكلمة معينة سمعها، لكن الحاسوب سيخبره بالكلمة التي تكررت فعلياً وبدقة رياضية.

■ **تكليف الأسبوع:** حاولوا كتابة فقرة قصيرة تصف شعوركم تجاه هذا المقرر، ثم حددوا يدوياً ما هي كلمات الوقف التي يجب حذفها لتظل الكلمات المعبرة عن رأيكم فقط. في الدرس القادم، سنتعلم كيف نجعل بايثون تقوم بهذه العملية بشكل آلي وتكراري باستخدام الحلقات والشروط.[1]