

## المحور الرابع : التحليل العنقودي في البحث العلمي

في ظل تزايد حجم البيانات وتنوع مصادرها، يُعد التحليل العنقودي من الأدوات الإحصائية الفعّالة التي تُستخدم لتجميع الحالات في مجموعات متجانسة دون الحاجة إلى متغير تابع. يساعد هذا الأسلوب في اكتشاف الأنماط والفئات الطبيعية داخل البيانات، مما يجعله مناسباً للبحوث الاستكشافية في مجالات متعددة مثل الطب، والإدارة، والتعليم، والعلوم الاجتماعية. كما يُعد التحليل العنقودي خطوة أولية مهمة قبل بناء النماذج التنبؤية أو التفسيرية، نظراً لقدرته على الكشف عن البنية الخفية للبيانات وتصنيفها بطريقة علمية وموضوعية.

### أولاً : مفهوم التحليل العنقودي

التحليل العنقودي هو أسلوب إحصائي يُستخدم لتقسيم البيانات إلى مجموعات متجانسة تُسمى "عناقيد"، بحيث تتشابه العناصر داخل كل مجموعة وتختلف عن المجموعات الأخرى. يتم ذلك دون الاعتماد على متغير تابع، بل بناءً على درجة التشابه بين الحالات. يُستخدم هذا التحليل في مجالات متعددة مثل التسويق، والطب، والعلوم الاجتماعية، لاكتشاف الأنماط وتصنيف البيانات بطريقة استكشافية.

### ثانياً : أنواع التحليل العنقودي

يُقسم التحليل العنقودي إلى عدة أنواع، تختلف في الأساليب التي تستخدمها لتجميع البيانات، ومن أبرز هذه الأنواع:

#### 1- التحليل العنقودي الهرمي

هو أحد أكثر الأساليب شيوعاً، ويقوم ببناء هيكل هرمي من التجميعات يُعرف بـ "شجرة التشعب" ينقسم هذا النوع إلى طريقتين:

- **التحليل التراكمي**: يبدأ بمعاملة كل حالة على أنها عنقود مستقل، ثم يُدمج الأقرب فالأقرب حتى تتجمع كلها في عنقود واحد.

- التحليل الانقسامي: يبدأ بجميع الحالات في عنقود واحد، ثم يُقسّم تدريجيًا حتى تصل إلى العناقيد المنفصلة.  
✓ المزايا:

- لا يتطلب تحديد عدد العناقيد مسبقًا.
  - يعطي تصورًا بصريًا لتدرج التجميع.
- ✓ العيوب:

- حساس للقيم الشاذة.
- أقل كفاءة في التعامل مع البيانات الكبيرة.

## 2- التحليل العنقودي غير الهرمي

أشهر خوارزمياته هي خوارزمية K-Means ، وهي تعتمد على تقسيم البيانات إلى عدد محدد مسبقًا من العناقيد (K) ، حيث تقوم الخوارزمية بتعيين نقاط مركزية (Centroids) وتحسب المسافات لتجميع الحالات حول أقرب مركز.

✓ المزايا:

- مناسب للبيانات الكبيرة.
  - أسرع من التحليل الهرمي.
- ✓ العيوب:

- يتطلب معرفة عدد العناقيد مسبقًا.
- قد يعطي نتائج مختلفة حسب مراكز البداية.

## 3- التحليل العنقودي القائم على الكثافة

أشهر خوارزمياته هي DBSCAN. يعتمد هذا النوع على فكرة أن العناقيد هي مناطق ذات كثافة بيانات عالية، ويفصل بينها مناطق ذات كثافة منخفضة. يُستخدم هذا النوع كثيرًا في تحليل البيانات ذات الشكل غير المنتظم.

✓ مزايا:

- لا يحتاج إلى تحديد عدد العناقيد مسبقًا.
- يتعامل جيدًا مع العناقيد غير الخطية والقيم الشاذة.

## ✓ العيوب:

- . اختيار معايير الكثافة قد يكون صعبًا.
  - . لا يعمل بكفاءة في البيانات ذات الكثافة غير المتجانسة.
- ### ثالثًا : شروط استخدام التحليل العنقودي في البحث العلمي

قبل تطبيق التحليل العنقودي في البحث العلمي أو التحليل البياني، يجب التأكد من تحقق مجموعة من الشروط الأساسية التي تساهم في ضمان دقة وفعالية النتائج المستخلصة. فيما يلي الشروط المهمة لاستخدام التحليل العنقودي:

1. يجب أن تحتوي البيانات على خصائص متعددة لتقسيمها إلى مجموعات
2. يجب معالجة القيم الشاذة التي قد تؤثر على دقة النتائج.
3. يجب أن تكون البيانات غير منحرفة ومتجانسة لتسهيل التحليل.
4. لا يجب أن يكون عدد المتغيرات كبيرًا جدًا لتجنب تعقيد التحليل.
5. يجب اختيار الخوارزمية بناءً على نوع البيانات (K-Means ، DBSCAN، الخ)
6. يجب اختيار العدد الأمثل للعناقيد باستخدام تقنيات مثل Elbow Method.
7. يجب أن تكون المجموعات مختلفة بشكل واضح.
8. توحيد أو تقييس المتغيرات لتقليل تأثير اختلاف الوحدات.
9. يجب معالجة البيانات الناقصة إما باستكمالها أو حذفها.

### رابعًا : خطوات إجراء التحليل العنقودي

التحليل العنقودي هو أحد الأساليب القوية في البحث العلمي، ويعتمد على تقسيم البيانات إلى مجموعات أو عناقيد تتشابه في الخصائص الأساسية. ولكي يتم تنفيذ هذا التحليل بشكل دقيق وفعال، يجب اتباع مجموعة من الخطوات وهي كما يلي:

## 1- تحضير البيانات

وتنقسم تلك الخطوة إلى مرحلتين:

- **تنظيف البيانات:** هي التأكد من أن البيانات خالية من القيم المفقودة أو القيم الشاذة التي قد تؤثر على نتائج التحليل. قد يتطلب الأمر استبدال القيم المفقودة أو إزالتها أو حتى استخدام طرق معينة لتقديرها.
  - **توحيد المقياس:** إذا كانت المتغيرات المستخدمة في التحليل العنقودي تختلف في الوحدات أو النطاق (مثل الطول والوزن)، فمن المهم توحيد المقياس باستخدام طرق مثل التوحيد (Normalization) أو التقييس (Standardization)، حيث إن التحليل العنقودي حساس للغاية لاختلاف المقاييس بين المتغيرات.
- ### 2- اختيار نوع التحليل العنقودي

تعتمد الخوارزمية المستخدمة في التحليل العنقودي على طبيعة البيانات ونوع العناقيد المطلوبة. أبرز الخوارزميات المستخدمة:

- **K-Means** تستخدم في حال كان عدد العناقيد محددًا مسبقًا ويفترض التوزيع الكروي للبيانات.
- **Hierarchical Clustering** تستخدم عندما لا يكون العدد معروفًا مسبقًا، وتسمح بتشكيل شجرة هيكلية للبيانات.
- **DBSCAN** يعتمد على الكثافة ولا يتطلب تحديد عدد العناقيد مسبقًا، ويستخدم بشكل أساسي في البيانات غير المنتظمة.
- **Gaussian Mixture Model (GMM)** يستخدم عندما تتبع البيانات توزيعًا احتماليًا ويمكنه التعامل مع بيانات غير خطية.

### 3- اختيار عدد العناقيد

نستخدم الأساليب التالية لتحديد العدد الأمثل للعناقيد وهي كالاتي:

- **Elbow Method** يتم رسم الرسم البياني لمقدار التباين المفسر مقابل عدد العناقيد. العدد الأمثل هو النقطة التي يبدأ عندها التغير في التباين بالتناقص بشكل حاد.

- **Silhouette Score** تقيس هذه الطريقة مدى تماسك العناقيد. يُفضل أن تكون قيمة المؤشر أعلى لتحديد العدد الأمثل للعناقيد.
- **Gap Statistic** تستخدم لتحديد العدد الأمثل للعناقيد عبر مقارنة التوزيع الداخلي للمجموعات الفعلية بالمجموعات العشوائية.

#### 4- تطبيق الخوارزمية

بعد تحديد عدد العناقيد واختيار الخوارزمية المناسبة، يتم تطبيق الخوارزمية على مجموعة البيانات. في حالة K-Means ، يتم توزيع البيانات بشكل عشوائي بين العناقيد في البداية ثم يتم تحديث المواقع بشكل متكرر حتى تصل المجموعات إلى الاستقرار.

#### 5- تحليل النتائج

بمجرد تطبيق الخوارزمية، تتم دراسة العناقيد التي تم الحصول عليها من خلال النظر في خصائص البيانات داخل كل مجموعة. يجب أن تكون العناقيد متجانسة داخل كل مجموعة (الأفراد في نفس العنقود يجب أن يشبهوا بعضهم البعض) ومختلفة تمامًا بين المجموعات (الاختلاف بين العناقيد كبير). يمكن استخدام مقياس **Silhouette Coefficient** لتقييم مدى جودة العناقيد. يتراوح المقياس بين -1 و 1، وكلما اقتربت النتيجة من 1، كان العنقود أكثر تماسكًا.

#### 6 - تفسير النتائج

بمجرد أن يتم تحديد العناقيد، يتم تحليل الخصائص الفريدة لكل مجموعة، مثل المتوسطات أو الفروق بين المتغيرات المختلفة داخل كل مجموعة. يجب تفسير العناقيد بناءً على المفاهيم النظرية للبحث. على سبيل المثال، في حالة دراسة الأسواق، قد يشير العنقود إلى مجموعة معينة من العملاء الذين يتشاركون في سلوك شراء مشابه.

#### 7 - تقديم التوصيات

بناءً على خصائص العناقيد، يمكن للباحثين اتخاذ القرارات المناسبة. في الأبحاث التسويقية، يمكن استخدام العناقيد المستخلصة في النماذج المستقبلية أو حتى في الأعمال المستقبلية لتوجيه الأنشطة أو قرارات الاستثمار بناءً على فهم أعمق للبيانات.

## خامسا : مميزات التحليل العنقودي في البحث العلمي

التحليل العنقودي له فوائد كثيرة تجعله من الأساليب المفضلة لدى الباحثين في تحليل بياناتهم. ومن أجل ذلك، نستعرض أبرز المميزات بشكل نقاط واضحة:

1. يساعد في كشف مجموعات طبيعية داخل البيانات ما كانت واضحة من قبل.
2. يعكس بعض الأساليب الأخرى، يمكن استخدامه، حتى لو ما كان فيه متغير يمكن أن نعتمد عليه.
3. يقدر الباحث أن يطبقه على أنواع متعددة من البيانات، سواء كانت رقمية أو وصفية (بشروط).
4. يعطي تصنيفات قابلة للتفسير تربط بين البيانات والنظرية.
5. أداة ممتازة كبدائية لفهم شكل البيانات قبل استخدام طرق تحليل أخرى.
6. مفيد جدًا في المجالات التطبيقية مثل التسويق أو التعليم، لأنه يوضح الفروقات بين المجموعات.

## سادسا : عيوب التحليل العنقودي في البحث العلمي

رغم أن التحليل العنقودي يُعد من الأساليب القوية في تحليل البيانات، إلا أن له بعض العيوب التي ينبغي الانتباه لها عند استخدامه في البحث العلمي. وفيما يلي أبرز هذه العيوب:

1. يتأثر التحليل العنقودي بشكل كبير بطريقة تحضير البيانات، مثل القيم المفقودة أو المقاييس المختلفة بين المتغيرات.
2. لا توجد طريقة واحدة مؤكدة لاختيار العدد الأمثل للعناقيد، مما قد يؤدي إلى نتائج غير دقيقة إذا تم التقدير بشكل خاطئ.
3. وجود القيم الشاذة في البيانات قد يؤدي إلى تكوين عناقيد غير منطقية أو مشوشة.

4. أحياناً يصعب تفسير معنى كل عنقود خاصةً في الحالات التي لا تكون فيها الفروق واضحة بين المجموعات.
5. تختلف النتائج بشكل كبير حسب الخوارزمية المستخدمة، وقد لا تعطي جميع الخوارزميات نفس الجودة أو التماسك في النتائج.

### سابعاً : مجالات استخدام التحليل العنقودي

يُستخدم التحليل العنقودي في مجالات متعددة نظراً لقدرته على تصنيف البيانات واكتشاف الأنماط المخفية. وفيما يلي أبرز المجالات التي يستفيد فيها الباحثون من هذا التحليل:

1. **التسويق وتقسيم السوق**: يُستخدم لتقسيم العملاء إلى مجموعات متشابهة بناءً على السلوك الشرائي أو الاهتمامات، مما يساعد في تصميم حملات تسويقية مخصصة.
2. **الطب والصحة**: يساهم في تصنيف المرضى بناءً على الأعراض أو التحاليل، مما يُمكن الأطباء من تقديم علاجات موجهة لكل مجموعة.
3. **التعليم**: يُستخدم لتقسيم الطلاب إلى فئات حسب الأداء أو أساليب التعلم، مما يساعد المعلمين على تطوير استراتيجيات تعليمية مناسبة لكل فئة.
4. **علم الاجتماع**: يُستخدم في تحليل أنماط سلوك الأفراد أو المجموعات، مما يُسهم في فهم العلاقات الاجتماعية بشكل أعمق.
5. **تحليل البيانات الجغرافية**: يُستخدم لتجميع المواقع أو المناطق حسب خصائص معينة، مثل الكثافة السكانية أو مستوى الدخل.
6. **البحوث البيئية**: يساعد في تصنيف المناطق البيئية أو أنواع الكائنات الحية، مما يسهل دراسة التنوع الحيوي أو آثار التغير المناخي.