

المحور الثالث: التحليل العنقودي

1. مقدمة:

في عالم التسويق الحديث، لا يمكن التعامل مع جميع العملاء بنفس الطريقة. فكل مجموعة من العملاء لديها احتياجات وسلوكيات مختلفة، مما يستدعي تقسيم السوق إلى شرائح متجانسة لتصميم استراتيجيات فعالة. يُعرف هذا الأسلوب بـ **تقسيم السوق (Market Segmentation)**، ويُعدُّ **التحليل العنقودي (Clustering)** أحد أبرز الأدوات الكمية لتحقيق هذا الهدف.

على عكس الطرق التقليدية التي تعتمد على الفئات الديموغرافية (كالعمر أو الجنس)، يعتمد التحليل العنقودي على البيانات الكمية (مثل معدل الشراء أو الولاء للعلامة) لتصنيف العملاء إلى مجموعات ذات سلوكيات متشابهة. على سبيل المثال، قد يُظهر التحليل أن 30% من العملاء يشتركون بانتظام بمنتجات باهظة الثمن، بينما يفضل 50% المنتجات ذات الأسعار المعقولة. باستخدام هذه المعلومة، يمكن للمسوقين تصميم حملات إعلانية موجهة أو تحسين تجربة العملاء.

2. الإطار النظري للتحليل العنقودي

أ. تعريف التحليل العنقودي:

التحليل العنقودي هو تقنية إحصائية تُصنف البيانات إلى مجموعات (عناقيد) بحيث تكون العناصر داخل كل مجموعة متشابهة، وتختلف عن العناصر في المجموعات الأخرى. يُستخدم هذا الأسلوب في التسويق لفهم أنماط العملاء، مثل:

• تصنيف العملاء بناءً على معدل الإنفاق.

• تحديد المناطق الجغرافية ذات الكثافة السكانية العالية.

ينقسم التحليل العنقودي إلى نوعين رئيسيين:

1. **التحليل الهرمي (Hierarchical Clustering)**: يُظهر العلاقات بين العناقيد عبر مخططات شجرية (Dendrograms).

2. **التحليل التقسيمي (Partitional Clustering)**: مثل خوارزمية K-means، التي تقسم البيانات إلى عدد محدد مسبقاً من العناقيد.

ب. أهداف التحليل العنقودي في التسويق:

• فهم تفضيلات العملاء: مثل تحديد العملاء المهتمين بالجودة مقابل السعر.

• تحسين التواصل التسويقي: توجيه الحملات الإعلانية إلى الشرائح الأكثر احتمالاً للشراء.

• زيادة رضا العملاء: عبر تخصيص الخدمات بناءً على احتياجاتهم.

ج. مراحل التحليل العنقودي:

1. جمع البيانات: مثل بيانات الشراء، التقييمات، أو التفاعلات على منصات التواصل.

2. اختيار طريقة القياس: تحديد كيفية حساب التشابه بين العناصر، مثل المسافة الإقليدية.

3. تطبيق الخوارزمية: استخدام أدوات مثل K-means أو التحليل الهرمي.

4. تقييم النتائج: التأكد من أن العناقيد متميزة وذات معنى عملي.

3. العلاقات الرياضية في التحليل العنقودي

أ. قياس التشابه والاختلاف:

تعتمد معظم خوارزميات التحليل العنقودي على حساب "المسافة" بين العناصر. أشهر المقاييس هي:

• المسافة الإقليدية:

تُحسب عبر القانون:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

مثال: إذا كان العميل أ (ينفق 100 ريال شهرياً ويعطي تقييماً 5 نجوم، والعميل ب (ينفق 200 ريال ويعطي 3 نجوم، فإن المسافة بينهما تُظهر مدى اختلاف سلوكياتهما.

• **المسافة الإحصائية لماتن:**

تُحسب عبر جمع الفروق المطلقة بين المتغيرات:

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|$$

ب. **خوارزمية K-means:**

هي أشهر خوارزميات التحليل التقسيمي، وتعمل وفق الخطوات التالية:

1. تحديد عدد العناقيد (K).
2. اختيار K مراكز عشوائية.
3. تعيين كل عنصر إلى أقرب مركز.
4. إعادة حساب مراكز العناقيد بناءً على العناصر المضافة إليها.
5. التكرار حتى تثبيت المراكز.

التحدي الرئيسي: اختيار القيمة المثلى لـ K ، غالباً عبر **طريقة Elbow** رسم منحني بين K ودرجة الخطأ، واختيار النقطة التي يبدأ عندها الانخفاض في التحسن.

ج. **التحليل العنقودي الهرمي:**

- **الربط الأحادي (Single Linkage):** يحسب المسافة بين أقرب عنصرين في عنقودين.
- **الربط الكامل (Complete Linkage):** يحسب المسافة بين أبعد عنصرين.
- **المخططات الشجرية (Dendrograms):** تُظهر كيفية دمج العناقيد عبر مستويات مختلفة.
- 4. **الطرق المستخدمة لحل نماذج التحليل العنقودي**

أ. **الطرق التقسيمية:**

• **K-means:**

- **المزايا:** سرعة التنفيذ، مثالية للبيانات الكبيرة.
- **العيوب:** حساسة للقيم الشاذة، تحتاج إلى تحديد K مسبقاً.

• **K-medoids:**

- تشبه K-means ، لكنها تستخدم عناصر البيانات الفعلية كمراكز، مما يجعلها أقل تأثراً بالقيم المتطرفة.

ب. **الطرق الهرمية:**

- **التجميع التراكمي (Agglomerative):** يبدأ كل عنصر كعنقود منفصل، ثم يدمج العناقيد الأقرب تدريجياً.
- **التجميع التفكيكي (Divisive):** يبدأ بعنقود واحد، ثم يُقسم إلى عناقيد أصغر.

ج. **الطرق القائمة على الكثافة:**

مثل خوارزمية **DBSCAN** ، التي تُحدد العناقيد بناءً على مناطق الكثافة العالية في البيانات، مما يسمح باكتشاف عناقيد ذات أشكال غير منتظمة.

5. **تطبيق عملي: حالة دراسية**

لنفترض أن لدينا بيانات 1000 عميل لمتجر إلكتروني، تتضمن:

- العمر.
- الإنفاق الشهري.
- عدد الزيارات الشهرية.

الخطوات:

1. **تنظيف البيانات:** إزالة القيم المفقودة أو غير المنطقية.
2. **تطبيق K-means:** بافتراض $K=3$ ، نحصل على:
 - **العنقود 1:** عملاء ينفقون أكثر من 500 ريال، ويزورون المتجر 10 مرات شهرياً.
 - **العنقود 2:** عملاء إنفاقهم بين 200-500 ريال، وزياراتهم 5 مرات.
 - **العنقود 3:** عملاء إنفاقهم أقل من 200 ريال، وزياراتهم نادرة.
3. **الاستراتيجية التسويقية:**
 - تقديم خصومات للعنقود 3 لتشجيعهم على الشراء.
 - إرسال عروض حصرية للعنقود 1 لتعزيز ولائهم.
6. **التحديات والاعتبارات**
 - **تحديد عدد العناقيد:** قد تؤدي القيمة غير الصحيحة لـ K إلى نتائج مضللة.
 - **البيانات غير المتوازنة:** بعض العناقيد قد تكون صغيرة جداً أو كبيرة.
 - **التفسير الخاطئ:** العناقيد الرياضية قد لا تكون ذات معنى تسويقي (مثل: عنقود يضم عملاء من جميع الأعمار دون تشابه حقيقي).

7. الخاتمة

أصبح التحليل العنقودي أداة لا غنى عنها لفهم تعقيدات السوق الحديث. عبر تصنيف العملاء إلى مجموعات متجانسة، يمكن للمسوقين تحسين مواردهم وتوجيهها بشكل أكثر فعالية. مع ذلك، يجب دائماً دمج النتائج الرياضية مع الخبرة العملية لضمان نجاح الاستراتيجيات. اليوم، ومع توفر أدوات مجانية مثل **Excel** أو مكتبات **Python** (مثل **Scikit-learn**)، أصبح تطبيق هذه التقنية في متناول الجميع.

طريقة حل مسائل التحليل العنقودي (خوارزمية K-means)

سنوضح الطريقة العامة لحل مسائل التحليل العنقودي باستخدام خوارزمية **K-means** ، مع توضيح العلاقات الرياضية المستخدمة، وذلك بالاعتماد على الأمثلة السابقة.

الخطوة 1: تعريف المشكلة وتحديد المتغيرات

- **الهدف:** تقسيم البيانات إلى مجموعات (عناقيد) متجانسة داخلياً ومختلفة خارجياً.
- **المتغيرات:** تُحدد بناءً على طبيعة المشكلة، مثل: العمر، الإنفاق، الولاء، عدد الزيارات، إلخ.
- **عدد العناقيد (K):** يُحدد مسبقاً (مثال $K=2$ أو $K=3$)، أو عبر طرق مثل طريقة **Elbow**.

الخطوة 2: اختيار المراكز الأولية (Initialization)

- **طريقة العشوائية:** اختيار K نقطة من البيانات كمراكز أولية (Centroids).
 - مثال: في بيانات العملاء، قد يُختار العميل رقم 1 والعميل رقم 8 كمركزين ابتدائيين.
- **ملاحظة:** قد تؤثر المراكز الأولية على النتيجة النهائية، لذا تُكرر الخوارزمية مع مراكز مختلفة في بعض الحالات.

الخطوة 3: حساب المسافات بين البيانات والمراكز

تُستخدم المسافة الإقليدية (Euclidean Distance) لقياس التشابه بين كل عنصر والمركز، وتحسب بالقانون:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

حيث:

- نقطة البيانات (مثال: عميل مع العمر، الإنفاق، الزيارات): $x = (x_1, x_2, \dots, x_n)$
- مركز العنقود: $c = (c_1, c_2, \dots, c_n)$
- عدد المتغيرات (مثال 2: أو 3 متغيرات): n

الخطوة 4: تعيين العناصر إلى العناقيد

- تُعين كل نقطة بيانات إلى العنقود الذي يمتلك أقرب مركز لها.
- مثال: إذا كانت المسافة بين العميل (25, 100) والمركز (30, 150) هي 50.25، بينما المسافة إلى المركز (45, 300) هي 212.13، يُعين العميل إلى العنقود الأول.

الخطوة 5: تحديث المراكز

- يُحسب مركز كل عنقود جديد كـ متوسط العناصر المضافة إليه.
- القانون الرياضي:

$$c_j = \left(\frac{1}{m_j} \sum_{x \in C_j} x_1, \frac{1}{m_j} \sum_{x \in C_j} x_2, \dots, \frac{1}{m_j} \sum_{x \in C_j} x_n \right)$$

حيث:

- c_j : مركز العنقود j .
- m_j : عدد العناصر في العنقود j .

الخطوة 6: التكرار حتى التقارب

- تُعاد الخطوات 3-5 حتى:
- a. لا تتغير المراكز بين تكرارين متتاليين.
- b. أو الوصول إلى عدد محدد من التكرارات (Iterations).

الخطوة 7: تقييم النتائج

- جودة العناقيد:
 - التباين الداخلي: (Within-Cluster Sum of Squares)

$$WCSS = \sum_{j=1}^K \sum_{x \in C_j} d(x, c_j)^2$$

كلما قلّ WCSS، كانت العناقيد أكثر تجانساً.

- طريقة **Elbow**: رسم منحنى WCSS مقابل K واختيار نقطة الانكسار (Elbow Point).
- العلاقات الرياضية الرئيسية

1. المسافة الإقليدية:

$$d(x, c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2}$$

2. تحديث المراكز:

$$c_j = \frac{1}{m_j} \sum_{x \in C_j} x$$

3. حساب WCSS:

$$WCSS = \sum_{j=1}^K \sum_{x \in C_j} \|x - c_j\|^2$$