



إسم المادة: تحليل البيانات الإحصائية

إسم المحاضر: المهندس خليل محمد

الأكاديمية العربية الدولية - منصة أعد

مقدمة في تحليل البيانات الإحصائية

مفهوم تحليل البيانات الإحصائية: تحليل البيانات الإحصائية هو العملية التي نستخدم فيها الأدوات والأساليب الإحصائية لفحص وتفسير البيانات. هذه العملية تهدف إلى تحويل البيانات الخام إلى معلومات ذات قيمة يمكن استخدامها في اتخاذ القرارات أو فهم الظواهر المختلفة. يمكن لعلماء البيانات، والباحثين، والمهندسين، استخدام هذه التحليلات لتحديد الأنماط، واختبار الفرضيات، وتوقع النتائج المستقبلية.

الإحصاء الوصفي مقابل الإحصاء الاستدلالي:

. **الإحصاء الوصفي (Descriptive Statistics):** يركز الإحصاء الوصفي على تلخيص وتنظيم البيانات بطريقة مفهومة. من خلال الإحصاء الوصفي، يمكننا تقديم نظرة عامة على مجموعة البيانات باستخدام مقاييس مثل المتوسط، الوسيط، التشتت، وغيرها. على سبيل المثال، إذا كان لدينا مجموعة من درجات الطالب في اختبار معين، يمكننا استخدام الإحصاء الوصفي لحساب متوسط الدرجات (الوسط الحسابي) أو لتحديد مدى انتشار الدرجات (الانحراف المعياري).

. **الإحصاء الاستدلالي (Inferential Statistics):** بينما يركز الإحصاء الوصفي على البيانات الحالية، يهدف الإحصاء الاستدلالي إلى استنتاجات تتجاوز البيانات المتوفرة. يستخدم الإحصاء الاستدلالي تقنيات مثل اختبار الفرضيات والنمذجة التنبؤية لتوقع أو تقدير خصائص مجتمع أكبر بناءً على عينة صغيرة. على سبيل المثال، يمكننا استخدام عينة من درجات طلاب مجموعة صغيرة لتقدير أداء جميع الطلاب في المدرسة.

تطبيقات تحليل البيانات الإحصائية

1. الصحة والرعاية الطبية:

- **تحليل التجارب السريرية:** يستخدم الإحصاء لتحديد فعالية الأدوية الجديدة أو طرق العلاج من خلال تحليل البيانات من التجارب السريرية.
- **التنبؤ بتفشي الأمراض:** يمكن استخدام النماذج الإحصائية لتحليل البيانات المتعلقة بالصحة العامة والتنبؤ بتفشي الأمراض مثل الإنفلونزا أو-COVID-19.

2. الأعمال والتجارة:

- **تحليل السوق:** تستخدم الشركات الإحصاءات لفهم سلوك العملاء وفضائلهم، مما يساعد في تصميم استراتيجيات تسويقية أكثر فعالية.
- **إدارة المخاطر:** تعتمد المؤسسات المالية على الإحصاءات لتقدير المخاطر المحتملة في الاستثمارات واتخاذ قرارات مستنيرة.

3. العلوم الاجتماعية:

- **دراسات الرأي العام:** تحليل نتائج استطلاعات الرأي يساعد في فهم توجهات السكان حول موضوعات اجتماعية أو سياسية معينة.
- **البحوث الديموغرافية:** يستخدم الإحصاء لدراسة التغيرات في السكان بمرور الوقت، مما يساعد الحكومات والمنظمات في تخطيط السياسات العامة.

دور الإحصاء في دعم اتخاذ القرارات المعتمدة على البيانات

- .**1. تقليل الشكوك:** تساعد التحليلات الإحصائية في تقليل الشكوك التي تواجهها الشركات والمؤسسات عند اتخاذ القرارات. من خلال توفير تقديرات دقيقة لاتجاهات المستقبلية أو تقييم المخاطر المحتملة، يصبح اتخاذ القرارات أكثر ثقة.
- .**2. التنبؤ والتخطيط:** يمكن استخدام النماذج الإحصائية للتنبؤ بالاتجاهات المستقبلية مثل المبيعات، الطلب على الخدمات، أو التغيرات في السوق. هذه التوقعات تساعد الشركات في التخطيط للمستقبل واتخاذ قرارات استراتيجية.
- .**3. تحليل الفرضيات:** يسمح الإحصاء للباحثين والمسؤولين بتقييم الفرضيات واختبارها بناءً على البيانات المتاحة. يمكن من خلاله تحديد ما إذا كانت القرارات المتخذة مبنية على أدلة قوية أم لا.
- .**4. تقييم الأداء:** يساعد تحليل البيانات الإحصائية المؤسسات في تقييم أدائها مقارنة بالأهداف المحددة، سواء كان ذلك في سياق الأعمال، التعليم، الصحة، أو غيرها من المجالات. هذا التقييم يساعد في تحديد مجالات التحسين واتخاذ إجراءات تصحيحية عند الحاجة.

الأساليب الإحصائية الأساسية للبيانات 1 - التحليل الوصفي

1- المقاييس المركزية

الوسط الحسابي: الوسط الحسابي هو أحد أكثر المقاييس شيوعاً في تحليل البيانات الإحصائية، ويشير إلى القيمة المتوسطة لمجموعة من البيانات. يتم حسابه عن طريق جمع كل القيم في مجموعة البيانات ثم تقسيم المجموع على عدد القيم. الوسط الحسابي يعتبر مقياساً جيداً للبيانات التي تكون موزعة بشكل متباين، حيث يعكس القيمة "المتوسطة" التي يمكن توقعها.

الوسيط: الوسيط هو القيمة التي تقع في منتصف مجموعة البيانات عند ترتيبها من الأصغر إلى الأكبر. إذا كان عدد القيم فردياً، يكون الوسيط هو القيمة الموجودة في المركز. أما إذا كان عدد القيم زوجياً، فيتم حساب الوسيط كمتوسط لقيمتين الموجودتين في المركز. الوسيط مفيد بشكل خاص في البيانات التي تحتوي على قيم شاذة أو متطرفة، حيث يمكن أن يكون أكثر دقة في تمثيل "المركز" من الوسط الحسابي.

المنوال: المنوال هو القيمة أو القيم الأكثر تكراراً في مجموعة البيانات. في بعض الأحيان، قد تكون هناك أكثر من قيمة واحدة للمنوال، أو قد لا يكون هناك منوال إذا كانت جميع القيم مختلفة. المنوال يستخدم عادةً في الحالات التي يكون فيها من المفید معرفة القيمة الأكثر شيوعاً، مثل تحديد أكثر المنتجات مبيعاً أو أكثر الأوقات ازدحاماً خلال اليوم.

الأساليب الإحصائية الأساسية للبيانات 1 - التحليل الوصفي

2- مقاييس التشتت

الانحراف المعياري : الانحراف المعياري يقيس مدى تشتت البيانات حول الوسط الحسابي. يعطي فكرة عن مدى انتشار البيانات؛ فإذا كان الانحراف المعياري صغيراً، فهذا يعني أن البيانات قريبة من الوسط الحسابي، والعكس صحيح. يستخدم الانحراف المعياري في العديد من المجالات لتقدير مدى تقلب البيانات وتقييم المخاطر.

التباين : التباين هو مقياس آخر للتشتت، وهو يعكس متوسط المربعات للانحرافات عن الوسط الحسابي. التباين يشير إلى مدى تباعد البيانات عن بعضها البعض. في الواقع، التباين هو مربع الانحراف المعياري، ويُستخدم بشكل متزايد في التحليل الإحصائي، خاصةً عندما يكون هناك حاجة لفهم مدى انتشار البيانات بشكل تفصيلي.

المدى : المدى هو أبسط مقياس للتشتت، ويعبر عن الفرق بين أكبر وأصغر قيمة في مجموعة البيانات. يعطي المدى فكرة سريعة عن نطاق البيانات ومدى التفاوت بين القيم القصوى. على الرغم من بساطته، إلا أنه قد يكون مفيداً في حالات معينة حيث يكون من المهم معرفة الفجوة بين القيم العليا والدنيا.

الأساليب الإحصائية الأساسية للبيانات 2- التحليل التوزيعي

1- التوزيع الطبيعي

خصائص التوزيع الطبيعي وأهميته في الإحصاء: التوزيع الطبيعي، المعروف أيضاً بمنحنى الجرس(Bell Curve) ، هو أحد أكثر التوزيعات استخداماً في الإحصاء بسبب خصائصه الفريدة. يتميز التوزيع الطبيعي بأن معظم القيم في مجموعة البيانات تتركز حول الوسط الحسابي، مع تناقص تدريجي في التكرارات كلما ابتعدت القيم عن المركز. يتميز التوزيع الطبيعي بعدة خصائص، منها:

- التماثل: التوزيع الطبيعي متماثل حول الوسط الحسابي، مما يعني أن القيم على جانبي الوسط الحسابي تتوزع بشكل متساوٍ.
- التجمع المركزي: حوالي 68% من البيانات تقع ضمن انحراف معياري واحد من الوسط، وحوالي 95% ضمن انحرافين معياريين.
- المنحنى الجرس: شكل المنحنى الذي يتمثل فيه التوزيع الطبيعي يشبه الجرس، وهو ما يجعله سهل التعرف عليه.

كيفية استخدام التوزيع الطبيعي في تحليل البيانات: في تحليل البيانات، يمكن استخدام التوزيع الطبيعي لتقدير احتمالات وقوع الأحداث المختلفة. على سبيل المثال، إذا كانت نتائج اختبارات الطلاب تتبع توزيعاً طبيعياً، يمكن استخدام الوسط الحسابي والانحراف المعياري للتنبؤ بأداء الطلاب المستقبلي.

الأساليب الإحصائية الأساسية للبيانات 2- التحليل التوزيعي

2- التوزيع التجريبي: التوزيع التجريبي هو الطريقة التي يتم بها تقدير التوزيع الإحصائي للبيانات بناءً على المعلومات المتوفرة من مجموعة بيانات فعلية.

1. جمع البيانات: يبدأ تقدير التوزيع التجريبي بجمع مجموعة من البيانات الفعلية ذات الصلة بالمسألة قيد التحليل.
2. تحديد الفئات: (Bins) لتقدير التوزيع التجريبي، تُقسم البيانات إلى فئات أو "سلال" وفقاً لنطاق القيم. على سبيل المثال، إذا كان لديك مجموعة من القيم العددية، يمكنك تقسيمها إلى فئات مثل "من 0 إلى 10"، "من 11 إلى 20"، وهكذا.
3. حساب الترددات: يتم حساب عدد البيانات أو التكرارات في كل فئة. هذا يعني ببساطة عدّ عدد القيم التي تقع ضمن كل فئة من الفئات المحددة.
4. رسم التوزيع التجريبي: يتم رسم التوزيع التجريبي عادةً باستخدام رسم بياني يُظهر ترددات القيم أو كثافتها عبر الفئات. يمكن أن يكون الرسم البياني عبارة عن هستوغرام(Histogram)، حيث يتم تمثيل التكرارات بشكل أعمدة، أو يمكن أن يكون رسمًا بيانيًا للتوزيع التجريبي الذي يعرض كيفية توزيع البيانات عبر الفئات.
5. تحليل النتائج: من خلال النظر إلى التوزيع التجريبي، يمكن فهم كيفية توزيع البيانات الفعلية وتحليل الأنماط الموجودة. يمكن أن يساعد هذا في تحديد ما إذا كانت البيانات تتبع نمطاً معيناً أو تتشابه مع توزيع معروف، أو إذا كانت هناك أي شذوذات.

الأساليب الإحصائية المتقدمة للبيانات 1- التحليل التبايني

1- **التحليل التبايني (ANOVA)**: هو مجموعة من الأساليب الإحصائية التي تستخدم لمقارنة متوازنات عدّة مجموعات أو أكثر. الهدف من ANOVA هو تحديد ما إذا كانت هناك اختلافات ذات دلالة إحصائية بين المتوازنات لمجموعات مختلفة بناءً على متغير معين.

كيفية استخدام ANOVA لمقارنة متوازنات بين مجموعات متعددة:

1. تحديد الفرضيات:

الفرضية الصفرية (H_0): تفترض أن جميع المجموعات التي يتم مقارنتها لها متوازنات متساوية.

الفرضية البديلة (H_1): تفترض أن هناك على الأقل مجموعة واحدة تختلف فيها المتوازنات عن الأخرى.

2. جمع البيانات: تحتاج إلى جمع بيانات من عدة مجموعات مختلفة. على سبيل المثال، يمكن أن تكون هذه المجموعات مجموعات علاجية في تجربة طبية أو مجموعات من طلاب يدرسون مواضيع مختلفة.

الأساليب الإحصائية المتقدمة للبيانات 1- التحليل التبايني

3. تحليل البيانات:

- حساب المتوسطات: يتم حساب متوسطات البيانات لكل مجموعة.
- حساب التباين داخل المجموعات: قياس مدى تباين القيم داخل كل مجموعة.
- حساب التباين بين المجموعات: قياس مدى اختلاف المتوسطات بين المجموعات.

4. اختبار الفرضيات:

- يستخدم تحليل ANOVA قيمة F التي تقارن بين التباين بين المجموعات والتباين داخل المجموعات.
- إذا كانت قيمة F المحسوبة أكبر من قيمة F الحرجة من الجداول الإحصائية، فإن ذلك يشير إلى وجود اختلاف ذو دلالة إحصائية بين المتوسطات.

5. تفسير النتائج:

- إذا كانت النتائج تشير إلى وجود اختلاف، يتم إجراء اختبارات مقارنة متعددة لتحديد أي المجموعات تختلف بشكل ملحوظ.

الأساليب الإحصائية المتقدمة للبيانات 2- التحليل الانحداري

أ) مفهوم الانحدار البسيط والمتعدد وكيفية بناء نماذج الانحدار لتوقع البيانات

- الانحدار هو تقنية إحصائية تُستخدم لفهم العلاقة بين متغير تابع واحد (ما نحاول التنبؤ به) ومتغير واحد أو أكثر من المتغيرات المستقلة (التي نستخدمها للتنبؤ).
- 1- الانحدار البسيط: هو تحليل العلاقة بين متغير تابع واحد ومتغير مستقل واحد فقط.
 - بناء النموذج: يتم بناء النموذج باستخدام معادلة خطية، حيث نرسم خطًا مستقيمًا يمثل العلاقة بين المتغيرين. الهدف هو تقدير العلاقة بين المتغيرين، بحيث يمكننا استخدام المتغير المستقل لتوقع المتغير التابع.
 - 2- الانحدار المتعدد: هو تحليل العلاقة بين متغير تابع واحد وأكثر من متغير مستقل.
 - بناء النموذج: يشمل هذا النموذج إنشاء معادلة تحتوي على أكثر من متغير مستقل. يهدف النموذج إلى تقدير العلاقة بين المتغير التابع وكل من المتغيرات المستقلة.

الأساليب الإحصائية المتقدمة للبيانات 2- التحليل الانحداري

ب) تقييم نموذج الانحدار

1. **معامل التحديد:** (R^2) : معامل التحديد هو مقياس يوضح مدى قدرة النموذج على تفسير التباين في البيانات. قيمة R^2 تقترب من 1 تعني أن النموذج يفسر معظم التباين، بينما قيمة قريبة من 0 تعني أن النموذج لا يفسر التباين بشكل جيد. **التفسير:** كلما كانت قيمة R^2 أعلى، كلما كان النموذج أفضل في تفسير التباين في البيانات.
2. **اختبار الفرضيات:** **الفرضية الصفرية:** تفترض أن المتغير المستقل ليس له تأثير على المتغير التابع. - **الفرضية البديلة:** تفترض أن المتغير المستقل له تأثير على المتغير التابع.
3. **فحص الاستقلالية والتعددية:** تعني أن المتغيرات المستقلة يجب أن تكون غير مرتبطة ببعضها البعض. - **التعددية:** تشير إلى وجود ارتباطات قوية بين المتغيرات المستقلة، مما قد يؤدي إلى صعوبات في تفسير النموذج. - **التفسير:** يتم استخدام اختبارات متعددة لفحص ما إذا كان هناك علاقات قوية بين المتغيرات المستقلة، والتأكد من أن كل متغير يساهم بشكل منفصل في النموذج.
4. **تحليل المتبقيات:** المتبقيات : هي الفروقات بين القيم الفعلية والقيم المتوقعة بواسطة النموذج. - **التفسير:** تحليل المتبقيات يساعد في التتحقق مما إذا كان النموذج يناسب البيانات بشكل جيد. المتبقيات التي تظهر نمطًا معيناً قد تشير إلى أن النموذج يحتاج إلى تحسين أو تعديل.

الأساليب الإحصائية المتقدمة للبيانات 3- الاختبارات الإحصائية

أ) اختبارات الفرضيات هي أدوات إحصائية تُستخدم لاختبار صحة افتراضات حول بيانات معينة. تشمل هذه الاختبارات فحص مدى صحة الفرضيات الأساسية بناءً على البيانات المتوفرة.

1- اختبار (T-test): هو اختبار إحصائي يُستخدم لمقارنة متوسطين لتحديد ما إذا كانت هناك فروق ذات دلالة إحصائية بينهما.

• أنواع اختبار T:

◦ اختبار T لعينة واحدة: يقارن متوسط مجموعة بيانات مع قيمة معينة أو مع الافتراض.

◦ اختبار T لمجموعتين مستقلتين: يقارن بين متوسطين من مجموعتين مختلفتين، مثل مقارنة درجات امتحان بين مجموعتين من الطلاب.

◦ اختبار T لمجموعتين مرتبطتين: يستخدم لمقارنة متوسطين لمجموعة واحدة تم قياسها في نقطتين زمنيتين مختلفتين.

2- اختبار كاي تربيع (Chi-Square Test): هو اختبار إحصائي يُستخدم لتقييم ما إذا كانت هناك علاقة بين متغيرين تصنifieين أو لقياس مدى توافق البيانات مع التوزيع النظري.

• أنواع اختبار كاي تربيع:

◦ اختبار كاي تربيع للجدول المتقاطع: يقيم العلاقة بين متغيرين تصنifieين. على سبيل المثال، يمكن استخدامه لتحليل العلاقة بين الجنس والفضائل.

◦ اختبار كاي تربيع للتواافق: يقارن توزيع البيانات الفعلي بتوزيع نظري محدد.

الأساليب الإحصائية المتقدمة للبيانات 3- الاختبارات الإحصائية

ب) **التحليل متعدد المتغيرات** هو مجموعة من الأساليب الإحصائية التي تُستخدم لتحليل أكثر من متغير في وقت واحد، بهدف فهم العلاقات والأنماط بين المتغيرات.

- **تحليل المكونات الرئيسية (PCA)** : التعريف : هو تقنية إحصائية تُستخدم لتقليل أبعاد البيانات مع الحفاظ على أكبر قدر ممكن من المعلومات.
- الهدف : يساعد PCA في تبسيط البيانات وتخفيض عدد المتغيرات المعقدة، مما يسهل تحليل البيانات وفهم الأنماط الرئيسية.

التطبيقات:

- **تقليل الأبعاد**: تقليل عدد المتغيرات المستخدمة في النموذج، مما يقلل من التعقيد ويسهل الأداء.
- **تصور البيانات**: تسهيل تصور البيانات عالية الأبعاد عبر تقليلها إلى أبعاد أقل.
- **تنظيف البيانات**: إزالة الضوضاء وتقليل تأثير المتغيرات ذات الأهمية الأقل على التحليل.

مثال: إذا كان لديك مجموعة كبيرة من المتغيرات مثل قياسات مختلفة لجسم الإنسان (الطول، الوزن، محيط الخصر، إلخ)، يمكن استخدام PCA لتحديد المكونات الرئيسية التي تلخص هذه القياسات وتبسيط التحليل.

تقنيات التحليل الإحصائي في علم البيانات 1- التعلم الآلي والإحصاء

أ) التعلم الآلي الموجة: هو نوع من التعلم الآلي حيث يتم تدريب النموذج على بيانات تحتوي على تسميات أو نتائج معروفة. الهدف هو أن يتعلم النموذج كيفية التنبؤ بالقيم أو التصنيفات بناءً على البيانات المدخلة. تنقسم تقنيات التعلم الآلي الموجة إلى فئتين رئيسيتين: الانحدار والتصنيف.

1- تقنيات الانحدار:

مفهوم الانحدار: يستخدم الانحدار لتقدير العلاقة بين متغير تابع واحد أو أكثر ومتغير مستقل واحد أو أكثر. الهدف هو بناء نموذج يمكنه التنبؤ بالقيمة الكمية للمتغير التابع بناءً على القيم المدخلة.

أنواع الانحدار:

- **الانحدار الخطي:** يستخدم لتحديد العلاقة بين المتغيرات بشكل خطى. على سبيل المثال، يمكن استخدامه لتوقع سعر المنزل بناءً على مساحته وعدد الغرف.
- **الانحدار المتعدد:** هو توسيع للانحدار الخطي يتعامل مع أكثر من متغير مستقل. على سبيل المثال، يمكن أن يتضمن الانحدار المتعدد عوامل متعددة مثل المساحة، الموقع، ونوع المنزل لتوقع السعر.

تقنيات التحليل الإحصائي في علم البيانات 1- التعلم الآلي والإحصاء

2- تقنيات التصنيف: هو عملية تحديد الفئة التي ينتمي إليها كل عنصر بناءً على ميزات معينة. يستخدم في الحالات التي يكون فيها الهدف هو تصنیف البيانات إلى فئات أو مجموعات محددة.

• أنواع التصنيف:

- الآلات الداعمة للتوجّه (**SVM**) خوارزمية تستخدم لتصنيف البيانات إلى فئات عن طريق إيجاد أفضل فاصل بين الفئات. على سبيل المثال، يمكن استخدامها لتصنيف الرسائل كرسائل غير مرغوب فيها أو لا.
- الجيران الأقربون: (**k-NN**) خوارزمية تصنيف تعتمد على قياس المسافة بين النقاط لتحديد الفئة التي ينتمي إليها عنصر معين بناءً على أغلب الجيران الأقربون.

تقنيات التحليل الإحصائي في علم البيانات 1- التعلم الآلي والإحصاء

ب) التعلم غير الموجه:

مفهوم التجميع(Clustering)

التجميع هو إحدى تكنيات التعلم غير الموجه، ويشير إلى عملية تقسيم مجموعة البيانات إلى مجموعات (أو "عناقيد") بحيث تكون البيانات في نفس المجموعة مشابهة أكثر من البيانات في مجموعات أخرى. الهدف من التجميع هو اكتشاف هيكليات البيانات واكتشاف الأنماط التي قد تكون غير واضحة من خلال التحليل التقليدي.

كيف يعمل التجميع؟

- . **التشابه والاختلاف**: في التجميع، يتم استخدام مقاييس التشابه لقياس مدى قرب أو بعد العناصر من بعضها البعض. العناصر التي تكون مشابهة يتم تجميعها في نفس العنقود، بينما العناصر التي تختلف يتم وضعها في عناقيد منفصلة.
- . **الأنماط غير المعروفة**: التجميع لا يتطلب معرفة مسبقة بفئات البيانات، بل يكتشف الأنماط والهيكليات من خلال تحليل البيانات نفسها.

تقنيات التحليل الإحصائي في علم البيانات 1- التعلم الآلي والإحصاء

أنواع تقنيات التجميع:

1. التجميع بناءً على المركز:

واحدة من أكثر تقنيات التجميع شيوعاً. تقوم بتقسيم البيانات إلى عدد محدد مسبقاً من العناقيد. (K) يقوم خوارزمية K-means بتحديث مراكز العناقيد بناءً على البيانات في كل مجموعة حتى يتقارب النموذج.

2. التجميع بناءً على الكثافة:

تعتمد على كثافة البيانات لتحديد العناقيد. يمكنها اكتشاف العناقيد ذات الأشكال المختلفة والتعامل مع الضوضاء (البيانات التي لا تنتمي إلى أي عنقود).

3. التجميع بناءً على التسلسل الهرمي:

يقوم بإنشاء شجرة (دالة شجرية) تمثل تسلسل العناقيد من الأكثر تجانساً إلى الأقل تجانساً. يمكن تقسيم البيانات إلى عدة مستويات من العناقيد بناءً على الحاجة.

تقنيات التحليل الإحصائي في علم البيانات 2- التحليل البياني

التحليل البياني المتقدم: التحليل البياني المتقدم يتجاوز التصورات الأساسية للبيانات لتوفير رؤى أعمق وأدق حول المعلومات المعقدة.

- 1- **الخريطة الحرارية (Heatmaps)**: الخريطة الحرارية تستخدم لتصور البيانات التي تتغير عبر مستويين أو أكثر. يتم استخدام ألوان مختلفة لتمثيل قيم مختلفة، مما يسهل رؤية الأنماط والاتجاهات.
 - الاستخدامات: يمكن استخدامها في تحليل الشبكات الاجتماعية، حيث يتم تمثيل كثافة التفاعل بين المستخدمين.
- 2- **الرسوم البيانية الشبكية (Network Graphs)**: الرسوم البيانية الشبكية تُستخدم لتصوير العلاقات بين العناصر في شبكة.
 - الاستخدامات: تُستخدم في تحليل الشبكات الاجتماعية لتوضيح العلاقات بين الأشخاص أو الكيانات. كما تُستخدم في تحليل شبكات النقل أو أي نظام معقد يتم فيه تتبع كيفية تفاعل العناصر المختلفة.
- 3- **الرسوم البيانية الثلاثية الأبعاد (3D Plots)**: الرسوم البيانية الثلاثية الأبعاد توفر رؤية مكعبية للبيانات، مما يسمح بمزيد من التفصيل حول كيفية تغيير البيانات عبر ثلاثة أبعاد.
 - الاستخدامات: تُستخدم في النمذجة العلمية والهندسية، مثل عرض بيانات التجارب المعقدة.

تقنيات التحليل الإحصائي في علم البيانات 2- التحليل البياني

- 4- **الرسوم البيانية المتداخلة(Stacked Plots)** : الرسوم البيانية المتداخلة تعرض بيانات متعددة في نفس الرسم البياني بطريقة تظهر كيفية تداخل البيانات مع بعضها البعض. يمكن أن تكون متراكبة أو مصفوفة.
- الاستخدامات: تُستخدم لعرض كيف تتغير مقاييس متعددة بمرور الوقت، مثل إيرادات الشركة من أقسام مختلفة أو نسب المبيعات المختلفة في رسم بياني واحد.
- 5- **الرسوم البيانية المتسلسلة الزمنية المعقدة(Advanced Time Series Plots)** : تستخدم لتحليل بيانات السلسلة الزمنية بتفاصيل أكثر، مثل تحليل الاتجاهات الموسمية أو تحديد الأنماط غير العادية.
- الاستخدامات: تُستخدم في التنبؤ بالأعمال والاقتصاد، مثل التنبؤ بالمبيعات أو تحليل تغيرات السوق المالية على مدار الوقت.
- 6- **التحليل التوزيعي المتقدم(Advanced Distribution Analysis)** : يتضمن تحليل التوزيع المعقد للبيانات لتحديد الأنماط غير العادية أو التغيرات الدقيقة في التوزيع.
- الاستخدامات: يُستخدم في تحليل البيانات الكبيرة لتحديد الأنماط الخفية، مثل البحث عن قيم شاذة في مجموعة بيانات ضخمة.
- 7- **الرسوم البيانية التفاعلية المتقدمة(Advanced Interactive Charts)** : تقدم الرسوم البيانية التفاعلية المتقدمة ميزات تفاعلية متعددة.
- الاستخدامات: تُستخدم في تطبيقات اللوحات المعلوماتية لتقديم تحليلات ديناميكية في الوقت الحقيقي، حيث يمكن للمستخدمين تخصيص وتحليل البيانات بطريقة تفاعلية.

تقنيات التحليل الإحصائي في علم البيانات 3- التحليل الزمني

1- سلاسل الزمن: هي مجموعة من البيانات التي تم جمعها بانتظام على فترات زمنية متتالية. يمكن أن تشمل هذه البيانات أي شيء يتم تسجيله بمرور الوقت، مثل درجات الحرارة اليومية، المبيعات الشهرية، أو أسعار الأسهم اليومية. سلاسل الزمن مهمة لأنها تساعد في تحليل كيف تغير البيانات على مدى فترة طويلة، وتتيح التعرف على الأنماط الموسمية، الاتجاهات العامة، والدورات الاقتصادية.

أنواع الأنماط في سلاسل الزمن:

- الاتجاهات (**Trends**): تشير إلى التغيرات المستمرة في البيانات على مدى فترة زمنية طويلة. على سبيل المثال، قد تظهر زيادة تدريجية في المبيعات بسبب نمو السوق أو تطوير منتج جديد.
- التغيرات الموسمية (**Seasonal Variations**): هي الأنماط التي تتكرر على فترات زمنية ثابتة، مثل زيادة في المبيعات خلال عطلات نهاية الأسبوع أو الأعياد.
- التغيرات الدورية (**Cyclical Variations**): هي التغيرات التي تحدث على فترات غير ثابتة، مثل تقلبات السوق الاقتصادية.
- التغيرات العشوائية (**Random Variations**): هي التغيرات التي لا يمكن التنبؤ بها والتي تحدث بسبب عوامل غير منتظمة.

تقنيات التحليل الإحصائي في علم البيانات 3- التحليل الزمني

- 2- التنبؤ بالاتجاهات: هو عملية استخدام بيانات سلاسل الزمن الحالية والتاريخية لتوقع المستقبل. يشمل هذا التنبؤ:
- تحديد الأنماط والاتجاهات : من خلال تحليل البيانات التاريخية، يمكن تحديد الاتجاهات العامة، مثل الزيادة أو النقصان في البيانات بمرور الوقت.
 - نموذج التنبؤ: بناء نماذج تنبؤية يمكن أن تساعد في تحديد كيف ستتغير البيانات في المستقبل بناءً على الأنماط الحالية. هناك عدة أنواع من النماذج يمكن استخدامها:
 - النماذج الخطية: مثل الانحدار الخطي، والتي تفترض أن البيانات تتبع نمطاً خطياً.
 - نماذج السلسلة الزمنية: مثل نموذج ARIMA (Autoregressive Integrated Moving Average)، الذي يمكنه التعامل مع البيانات التي تظهر تبايناً موسمياً وتغيرات دائمة.
 - نماذج التنبؤ المتقدمة: مثل الشبكات العصبية الاصطناعية، والتي يمكنها التعلم من البيانات التاريخية والتنبؤ بالاتجاهات المستقبلية بشكل أكثر تعقيداً.
 - تقييم النماذج: من خلال اختبار النماذج على بيانات تاريخية، يمكن التحقق من دقتها وموثوقيتها في التنبؤ بالاتجاهات المستقبلية.
 - تطبيقات التنبؤ: تُستخدم تقنيات التنبؤ في مجالات متعددة، مثل:
 - التجارة: توقع المبيعات، إدارة المخزون، وتحطيط الطلب. - المالية: توقع أسعار الأسهم، التنبؤ بالأزمات الاقتصادية. - الصحة: التنبؤ بأوبئة أو الأمراض المستقبلية.

تحليل البيانات في المشاريع الحقيقة

١- **تحليل بيانات المبيعات في الشركات التجارية:** في شركات البيع بالتجزئة، يستخدم تحليل البيانات لتحسين استراتيجيات المبيعات وإدارة المخزون. من خلال جمع وتحليل بيانات المبيعات، يمكن تحديد:

- **أنماط الشراء:** تحليل البيانات لتحديد أوقات الذروة والأيام التي تشهد أعلى معدلات شراء.
- **المنتجات الرائجة:** معرفة المنتجات التي تحقق أعلى مبيعات وبالتالي التركيز على تحسين مخزونها.
- **الاستراتيجيات التسويقية:** تقييم فعالية الحملات التسويقية من خلال مقارنة البيانات قبل وبعد الحملة.

٢- **تحسين الرعاية الصحية من خلال تحليل البيانات الطبية:** في القطاع الصحي، يستخدم تحليل البيانات لتحسين الرعاية الصحية والوقاية من الأمراض. يشمل ذلك:

- **تحليل بيانات المرضى:** مراقبة وتقييم حالات المرضى لتحديد الأنماط والاتجاهات في الأمراض.
- **التنبؤ بالأوبئة:** استخدام النماذج التنبؤية لتوقع انتشار الأمراض بناءً على البيانات التاريخية والحالية.
- **تحسين العمليات:** تحليل بيانات العمليات الطبية لتحسين فعالية العلاج وتقليل الأخطاء الطبية.

تحليل البيانات في المشاريع الحقيقة

- 3- **تحليل بيانات السوق المالية:** في الأسواق المالية، يُستخدم تحليل البيانات لتوقع تحركات السوق واتخاذ قرارات استثمارية مستنيرة. يتضمن ذلك:
- **تحليل الاتجاهات السوقية:** دراسة البيانات التاريخية للأسوق لتحديد الاتجاهات والتوقعات المستقبلية.
 - **تقييم المخاطر:** تحليل بيانات الأداء المالي لتقدير المخاطر المحتملة والفرص الاستثمارية.
 - **تحليل الأسهم:** تقييم بيانات أسعار الأسهم والأرباح لتحديد الأسهم الوعادة للاستثمار.
- 4- **استخدام تحليل البيانات في صناعة التكنولوجيا:** في صناعة التكنولوجيا، يُستخدم تحليل البيانات لتحسين الأداء وتطوير المنتجات. يشمل ذلك:
- **تحليل بيانات المستخدمين:** فهم كيفية تفاعل المستخدمين مع التطبيقات والخدمات لتحسين تجربة المستخدم.
 - **تحليل أداء الأنظمة:** مراقبة وتحليل بيانات أداء الأنظمة لتحديد المشكلات وتحسين الأداء.
 - **تطوير المنتجات:** استخدام البيانات لتوجيه تطوير المنتجات الجديدة بناءً على احتياجات المستخدمين واتجاهات السوق.

التحديات في تحليل البيانات

- 1. **التعامل مع البيانات المفقودة:** البيانات المفقودة هي مشكلة شائعة في تحليل البيانات، ويمكن أن تؤثر بشكل كبير على نتائج التحليل.
 - **حذف البيانات المفقودة:** في بعض الحالات، قد يكون من الأسهل حذف السجلات التي تحتوي على بيانات مفقودة، خاصة إذا كانت نسبة البيانات المفقودة صغيرة.
 - **ملء القيم المفقودة:** يمكن استخدام تقنيات ملء القيم المفقودة، مثل ملء القيم المفقودة بمتوسط القيم المتاحة، الوسيط، أو القيم الأكثر تكراراً.
 - **التقدير بواسطة النماذج:** يمكن استخدام أساليب مثل الانحدار أو النماذج الإحصائية الأخرى لتقدير القيم المفقودة بناءً على العلاقات مع المتغيرات الأخرى في البيانات.
- 2. **التعامل مع البيانات الشاذة:** البيانات الشاذة هي القيم التي تختلف بشكل كبير عن بقية البيانات في مجموعة البيانات.
 - **التعرف على البيانات الشاذة:** تبدأ عملية التعامل مع البيانات الشاذة بتحديد القيم الشاذة، والتي يمكن تحديدها باستخدام تقنيات مثل تحليل الرسوم البيانية (مثل الرسوم البيانية الصندوقية) أو حساب المقاييس الإحصائية (مثل الانحراف المعياري).
 - **فحص الأسباب:** من الضروري فحص الأسباب التي أدت إلى ظهور القيم الشاذة. يمكن أن تكون الأسباب أخطاء في جمع البيانات، قياسات غير دقيقة، أو وجود حالات غير عادلة.

دعوة للتعلم المستمر

في عالم تتطور فيه التقنيات بشكل سريع، من الضروري أن يواصل الطالب التعلم والتدريب في مجال التحليل الإحصائي. هذا يشمل:

- **استكشاف أساليب وتقنيات جديدة:** تشجيع الطالب على البحث واستخدام أساليب وتقنيات إحصائية جديدة لمواكبة أحدث التطورات في مجال البيانات.
- **المشاركة في المشاريع والتطبيقات العملية:** التفاعل مع مشاريع حقيقة وتطبيق ما تعلموه على بيانات فعلية يعزز من مهاراتهم ويعمق فهمهم.
- **التدريب المستمر والتطوير المهني:** المشاركة في دورات تدريبية، وورش عمل، ومؤتمرات لمواصلة تحسين مهاراتهم ومواكبة أحدث الاتجاهات في علم البيانات والإحصاء.



الأكاديمية العربية الدولية
Arab International Academy

آمل أن تكونوا قد حفظتم الفائدة