مركز الإحصاء | حكومة عجمان
Statistics Center | Government of Ajman

دليل استخدام الذكاء الاصطناعي في المعاينة الإحصائية

إصدار أكتوبر 2025

www.scc.ajman.ae

جميع الحقوق محفوظة © مركز الإحصاء

حكومة عجمان - الإمارات العربية المتحدة @ 2025

يمنع نسخ أو استعمال أي جزء من هذا الكتاب من قبل أي شخص أو شركة أو جهة بأية وسيلة تصويرية أو إلكترونية أو ميكانيكية بما في ذلك التسجيل الفوتغرافي والتسجيل على أقراص مقروءة أو بأية وسيلة نشر أخرى بما فيها حفظ المعلومات و استرجاعها دون الحصول على موافقة مسبقة صادرة من مركز عجمان للإحصاء، حكومة عجمان، دولة الإمارات العربية المتحدة.

في حالة الاقتباس يرجى الإشارة إلى المطبوعة كالتالي:

مركز عجمان للإحصاء - حكومة عجمان

دليل استخدام الذكاء الاصطناعي في المعاينة الإحصائية

إصدار رقم 1 - 2025

للتواصل وطلب البيانات الإحصائية يرجى التواصل:

مركز عجمان للإحصاء

البريد الإلكتروني: info.scc@ajman.ae

رقم الهاتف: +971 6 701 6770

الموقع الإلكتروني: scc.ajman.ae

ص.ب: 6556، عجمان - دولة الإمارات العربية المتحدة

@sccajman

التعريف بمركز عجمان للإحصاء

تم إنشاء "مركز عجمان للاحصاء " استناداً للمرسوم الأميري رقم (8) لسنة 2022. ويعتبر المركز هو الجهة المختصة محلياً في إمارة عجمان والمصدر الرئيس والمرجع الوحيد فيها في الشؤون الإحصائية المنصوص عليها في هذا المرسوم. يهدف المركز إلى تحقيق الغايات التالية:

1. تنظيم وتطوير العمل الإحصائي بما يحقق مصالح الدولة والإمارة.
2. بناء نظام إحصائي محلي متكامل.
3. دعم منظومة إتخاذ القرار في الحكومة ببيانات ومعلومات دقيقة وحديثة.

الرؤية

بالمعرفة نعزز مستقبل عجمان.

الرسالة

الإرتقاء بالعمل الإحصائي من خلال تطبيق أفضل الممارسات بإتباع المنهجيات العلمية الإحصائية والمعايير الموصى بها دولياً لتلبي إحتياجات مستخدمي البيانات ومتخذى القرار في الإمارة.

القيم

الجودة / الحيادية / الإحترافية / الموثوقية / الإبداع والابتكار / السرية / الشفافية

دليل استخدام الذكاء الاصطناعي في المعاينة الإحصائية

المحتويات

الفصل الأول 6
المقدمة 6
1.1 أهداف الدليل 6
2.1 أهمية الدليل 6
3.1 نطاق الدليل 7
4.1 مستخدمي الدليل 7
5.1 المفاهيم والمصطلحات 7
الفصل الثاني 13
الذكاء الاصطناعي ، أهميته، استخداماته 13
1.2 أهمية الذكاء الاصطناعي 13
2.2 العلاقة بين الذكاء الاصطناعي والإحصاء 14
3.2 استخدامات الذكاء الاصطناعي في الإحصاء 15
4.2 مقارنة بين إختيار العينات بالطرق التقليدية واستخدام الذكاء الاصطناعي 16
الفصل الثالث 17
خطوات استخدام الذكاء الاصطناعي في المعاينة الإحصائية 17
1.3 تحديد الهدف الإحصائي وتحديد إطار المعاينة 17
2.3 استخدام البيانات السابقة لتدريب نموذج تنبؤي 17
3.3 تقدير حجم العينة باستخدام منحنيات التعلم 18
4.3 إختيار وحدات العينة وجمع البيانات 19
5.3 تخصيص الموارد البشرية أو المالية 19
6.3 معالجة البيانات المفقودة وحالات عدم الإستجابة بالنماذج التوليدية 20
7.3 تقييم الأداء والتحقق من عدم التحيّز 21
8.3 التوثيق والتحسين المستمر 22
9.3 التحسين المستمر وبناء القدرات 23
10.3 أمثلة تطبيقية 24
11.3 مزايا استخدام الذكاء الاصطناعي في المعاينة 35
المراجع 38

الفصل الأول
المقدمة

يشهد مجال الإحصاءات تحولاً متسارعاً بالوقت الحالي مع تبني تقنيات الذكاء الاصطناعي (AI) في مختلف مراحل جمع البيانات وتحليلها. ويهدف هذا الدليل إلى تزويد الإحصائيين والمتخصصين في مجال الإحصاء بإطار منهجي شامل لتوظيف تقنيات الذكاء الاصطناعي في تصميم وتنفيذ خطط المعاينة الإحصائية. ويركز الدليل على المعاينة الإحصائية باستخدام الذكاء الاصطناعي مع إبراز التطبيقات العملية في المسوح الأسرية والاقتصادية، والتعدادات السكانية والبيئية، واستطلاعات الرأي العام.

يرجى ملاحظة أن دليل "استخدام الذكاء الاصطناعي في اختيار المعاينة الإحصائية" لا يزال قيد التطوير والتحديث المستمر، وذلك لمواكبة التطورات التقنية المتسارعة في هذا المجال الحيوي. وسيتم إثراء محتواه بشكل دوري استناداً إلى التجارب التطبيقية، والتطورات البحثية، وأفضل الممارسات العالمية في استخدام الذكاء الاصطناعي في العمل الإحصائي.

ويتضمن الدليل الفصول التالية:

الفصل الأول: المنهجية.

الفصل الثاني: الذكاء الاصطناعي ، أهميته، استخداماته.

الفصل الثالث: خطوات استخدام الذكاء الاصطناعي في المعاينة الإحصائية.

1.1 أهداف الدليل

يعد الدليل مرجعاً عملياً وفنياً شاملاً لموظفي مركز عجمان للإحصاء حيث يتناول الدليل المبادئ الأساسية والأساليب المتقدمة في تصميم وسحب العينات الإحصائية باستخدام الذكاء الاصطناعي .

كما يهدف إلى:

- توحيد منهجيات استخدام المعاينة المطبقة في المركز.
- تعزيز جودة المسوح الإحصائية من خلال تبني أفضل الممارسات العالمية.
- دعم بناء القدرات الفنية للإحصائيين العاملين في مجال تصميم العينات في المسوح باستخدام الذكاء الاصطناعي .

2.1 أهمية الدليل

تبرز أهمية هذا الدليل في كونه أداةً إرشادية تُسهم في:

- رفع كفاءة تنفيذ المسوح الإحصائية وضمان تمثيلها الدقيق.
- تقليل التكاليف والجهد الميداني والوقت من خلال استخدام أساليب معاينة فعالة تعتمد على الذكاء الاصطناعي .

- توثيق إجراءات المعاينة باستخدام الذكاء الاصطناعي بما يحقق الشفافية وقابلية التقييم.
- توفير مرجع موحد يساعد على نقل المعرفة وتدريب الكوادر الجديدة.
- يعتبر الدليل منهجاً ومرجعاً للإستدلال به والرجوع إليه لاستخدام العينات.

3.1 نطاق الدليل

يشمل نطاق الدليل المفاهيم الأساسية للعينات الإحصائية وأنواعها. وأساليب تصميم وسحب العينات المستخدمة في المسوح الاقتصادية، الأسرية، والاجتماعية وذلك باستخدام الذكاء الاصطناعي .

4.1 مستخدمي الدليل

تم إعداد هذا الدليل ليُستخدم من قِبل:

- الإحصائيين والفنيين العاملين في مجال العمل الاحصائي.
- فرق المسوح الإحصائية الميدانية والإشرافية.
- الجهات الحكومية أو الأكاديمية المتعاونة مع المركز في تنفيذ المسوح أو تحليل البيانات.
- الباحثين والدارسين المهتمين بأساليب المعاينة الإحصائية التطبيقية.

5.1 المفاهيم والمصطلحات[1]

- **المجتمع الإحصائي**

هو جميع الوحدات أو الأفراد أو العناصر التي يتطلب دراسة خصائصها. مثل: جميع الأسر في الدولة، أو جميع الشركات الصناعية، أو جميع الأفراد فوق 18 سنة.

- **العينة**

هي مجموعة فرعية يتم اختيارها من المجتمع الإحصائي وفقاً لمنهجية معينة، وتُستخدم لتمثيل المجتمع في الدراسة أو المسح.

- **وحدة المعاينة**

هي أصغر وحدة تُسحب منها العينة، وقد تكون فرداً، أسرة، منشأة، أو قطعة أرض، حسب نوع الدراسة.

- **إطار المعاينة**

هو القائمة أو السجل الذي يحتوي على جميع وحدات المجتمع المستهدف، ويُستخدم كأساس لسحب العينة منه. يجب أن يكون شاملاً ومحدَّثاً وخالياً من التكرار أو النقص.

---

[1] الناتج المحلي الإجمالي في إمارة عجمان لعام 2023 (إصدار خاص). مركز عجمان للإحصاء
تعريف تحليل التباين وأهميته وانواعه. دراسة
طرق التعلم الإحصائي.المبرمج العربي

- **حجم العينة**

هو عدد الوحدات التي يتم اختيارها من المجتمع. ويُحدد الحجم بناءً على عدة عوامل: درجة الدقة المطلوبة، حجم المجتمع، التباين بين الوحدات، والتكلفة المتاحة.

- **خطأ المعاينة**

هو الفرق بين قيمة المؤشر في العينة وقيمته الحقيقية في المجتمع. ويحدث هذا الخطأ بسبب الإعتماد على جزء من المجتمع وليس كله. يمكن تقليصه من خلال استخدام عينات عشوائية كافية الحجم.

- **التحيز في العينة**

هو انحراف ناتج عن طريقة اختيار العينة بحيث لا تكون ممثلة للمجتمع الإحصائي. يظهر التحيز عادة عند استخدام طرق غير عشوائية أو عند وجود خلل في إطار العينة، مما يؤدي إلى نتائج غير دقيقة.

- **التوزيع الإحتمالي**

هو التوزيع الإحصائي لجميع القيم الممكنة لمؤشر معين (مثل المتوسط أو النسبة) الناتجة عن سحب عينات متكررة من نفس المجتمع. هذا المفهوم مهم لتقدير فترات الثقة واختبار الفرضيات.

- **الخطأ غير المعاينى**

هو خطأ لا يتعلق باختيار العينة، وإنما ينتج عن أخطاء في جمع البيانات، مثل:

- عدم استجابة بعض الأفراد.
- أخطاء في التسجيل أو الإدخال.
- فهم خاطئ للأسئلة من قبل المستجيبين.

- **مستوى الثقة**

هو احتمال أن يحتوي التقدير الناتج من العينة على القيمة الحقيقية للمؤشر في المجتمع. الشائع استخدامه هو 95%، مما يعني أننا نثق بنسبة 95% بأن الفاصل يحتوي على القيمة الحقيقية.

- **تعلم الآلة (Machine Learning)**

وهو جزء مهم من الذكاء الاصطناعي ، وهو أحد فروع الذكاء الاصطناعي الّذي يُعنى بجعل الحاسوب قادراً على التعلُم من تلقاء نفسه من أيّ خبرات أو تجارب سابقة، مما يجعله قادراً على التنبؤ واتخاذ القرار المُناسب بصورة أسرع.

- **الخوارزمية**

هي مجموعة من الخطوات المتتابعة التي تستخدم لحل مشكلة ما أو للقيام بمهمة. في سياق تعلم الآلة، الخوارزميات هي الطرق التي تستخدمها الأنظمة للتعلم من البيانات.

- **النموذج**

يرجع إلى النظام الذي تم بناؤه عن طريق الخوارزمية. يتم تدريب النموذج على بيانات معينة ليتعلم أنماطاً وعلاقات.

- **التعلم المشرف**[1]

هو نوع من التعلم حيث يتم توفير بيانات التدريب بما في ذلك الإدخال والإخراج المطلوب. يتعلم النموذج العلاقة بين الإدخال والإخراج ويستخدمها للتنبؤ بالبيانات الجديدة.

- **التعلم غير المشرف** [2]

في هذا النوع من التعلم، لا توجد بيانات إخراج متاحة ويجب على النموذج تعلم الأنماط والعلاقات في البيانات بنفسه.

- **التعلم المعزز**

هو نوع من التعلم حيث يتم اعتماد النموذج أو إلغاؤه بناءً على أدائه، والهدف هو تحسين الأداء بمرور الوقت.

- **التعلم العميق**

هو نوع من تعلم الآلة يستخدم الشبكات العصبية العميقة التي تتألف من العديد من الطبقات.

- **الشبكات العصبية**

هي نماذج مستوحاة من الدماغ البشري وتستخدم في التعلم العميق.

- **البيانات**

هي مجموعة من المعطيات، أو الحقائق الأولية، أو التعليمات، أو الأرقام التي تكون مواد خام ولا تكون ذات معنى ولا تحمل أي غرض محدد.

- **التنبؤ**

هو عملية استخدام البيانات التاريخية والتحليلات لتقدير الأحداث المستقبلية. ويستخدم التنبؤ لتقدير متغيرات مثل الناتج المحلي الإجمالي، معدلات البطالة، أو معدلات الفائدة في المستقبل.

- **الإنحدار (Regression)**

هو تقنية إحصائية تُستخدم لإستكشاف العلاقة بين المتغيرات. وتكون العلاقة بين متغيرين أو أكثر، وتحديد كيفية تأثير متغير واحد (أو أكثر) على متغير آخر. أحدهما يسمى المتغير التابع والأخر

---

[1] طرق التعلم الإحصائي.المبرمج العربي

[2] طرق التعلم الإحصائي.المبرمج العربي

يسمى المتغير المستقل مثلًا، قد نستخدم الإنحدار لفهم كيف يؤثر الإستهلاك الخاص على الناتج المحلي الإجمالي.

- **الإرتباط**

يُستخدم هذا المصطلح لوصف العلاقة بين متغيرين. الإرتباط يمكن أن يكون إيجابياً (يزيد أحد المتغيرات عند زيادة الآخر) أو سلبيًا (ينخفض أحد المتغيرات عند زيادة الآخر). كما أن "الارتباط لا يعني السببية".

- **تحليل التباين (ANOVA، اختصار لـ "Analysis of Variance")**

هو أحد الإختبارات المعملية والتي تستخدم في المقارنة بين المتوسطات، ويطلق عليه تحليل التباين الأُحادي (One-way analysis of. variance)، وهو إختبار يتم استخدامه في التحليل الإحصائي في العديد من البحوث الأكاديمية، للحصول على قرار إما بوجود فروق بين المتوسطات أو عدم وجودها، وذلك للتعرف على ما يجعل متوسط ما يختلف عن المتوسطات الأخرى.

- **R:**

هو معامل الإرتباط الخطي، والذي يُعرف أيضًا باسم "معامل الارتباط البسيط" أو "معامل بيرسون للإرتباط". هو مقياس إحصائي يعبر عن القوة والإتجاه للعلاقة الخطية بين متغيرين.

- **مربع R : R-square**

مربع R يقيس نسبة التباين في المتغير التابع الذي يمكن تفسيره من خلال المتغير المستقل في نموذج الإنحدار.

- **مربع R المعدل: Adjusted R-squared**

مربع R المعدل وهو تعديل لمربع R يأخذ بعين الإعتبار عدد المتغيرات المستقلة في النموذج وحجم العينة. ويعتبر أكثر دقة من مربع R، خاصةً عند وجود العديد من المتغيرات المستقلة.

- **خوارزمية K-Means**

هي خوارزمية تصني (Clustering) من خوارزميات التعلم غير الخاضع للإشراف (unsupervised learning)، تُستخدم لتقسيم البيانات إلى مجموعات متشابهة تُسمى طبقات ذكية.

- **خوارزمية الغابة العشوائية (Random Forest)[1]**

هي خوارزمية تعلم آلي تجمع بين نواتج عمليات اتخاذ القرار المتعددة للوصول إلى نتيجة واحدة.

[1] ما المقصود بالغابة العشوائية؟IBM

- **منحنى التعلم (Learning Curve)**

هو تمثيل بياني للتقدم في الكفاءة خلال فترات متلاحقة من التدريب أو الأداء.

- **التجميع الذكي (Clustering)**

يشير المصطلح إلى تقنية تحليل بيانات، تُستخدم لتقسيم مجموعة كبيرة من البيانات إلى (مجموعات أو عناقيد) بحيث تكون العناصر داخل كل مجموعة متشابهة أكثر فيما بينها مقارنة بالعناصر في المجموعات الأخرى.

- **متوسط مربع الخطأ (Mean Squared Error - MSE)**

هو أحد الطرق التي تقيس متوسط مربعات الأخطاء، أي متوسط الفرق التربيعي بين القيم المقدرة والقيم الفعلية.

- **المشفر التلقائي التباينيّ[1] (Variational Autoencoder - VAE)**

المشفرات التلقائية هي أنظمة ذاتية الإشراف، هدفها التدريبي هو ضغط أو تشفير بيانات المدخلات من خلال تقليل الأبعاد ثم إعادة بناء مدخلاتها الأصلية بدقة باستخدام هذا التمثيل المضغوط.

- **التوزين الاحتمالي العكسي[2](Inverse Probability Weighting – IPW)**

هو أسلوب إحصائي يستخدم بشكل اساسي في الدراسات الرصدية لضبط المتغيرات المربكة، ويهدف إلى إنشاء مجموعة سكانية زائفة تكون فيها مهمة العلاج مستقلة عن المتغيرات المشاركة المرصودة.

- **الذكاء الاصطناعي (AI)[3]**

الذكاء الاصطناعي والذي يُعرف اختصارا بـ AI هو تقنية ذات قدرات حل تشبه قدرات الإنسان في حل المشكلات. يبدو أن الذكاء الاصطناعي في العمل يحاكي الذكاء البشري - يمكنه التعرف على الصور وكتابة القصائد وإجراء تنبؤات قائمة على البيانات.

- **التعلم النشط (Active Learning):**

هو أحد أساليب التعلم التي تتضمن إشراك الطلاب بشكل نشِط في الدورة الدراسية من خلال المناقشات وحل المشكلات ودراسات الحالة ولعِب الأدوار وغيرها من الأساليب الأخرى. تتطلّب أساليب التعلم النشط من الطلاب المشاركة في العملية التعليمية من خلال التفكير والمناقشة

[1] ما هو برنامج التشفير التلقائي المتغير، IBM

[2] الترجيح الاحتمالي العكسي. Learn statistics easily

[3] ما هو الذكاء الاصطناعي (AI) ، aws amazon

والتحقيق والإبداع، ويتم استخدام هذا الأسلوب في الذكاء الاصطناعي تقديم بيانات مختلفة وتدريب النماذج عليها حتى الوصول لأفضل النتائج.

- **عدم اليقين[1]:(Expected Uncertainty)**

يشير إلى درجة الشك أو عدم القدرة على التنبؤ المرتبطة بقياس أو تنبؤ أو نتيجة معينة، و يعد عدم اليقين مفهومًا أساسيًا يؤثر على عمليات اتخاذ القرار. وينشأ من مصادر مختلفة، بما في ذلك أخطاء القياس، والتباين في البيانات، والعشوائية المتأصلة في الأنظمة.

- **نظام المعلومات الجغرافية (GIS: Geographic Information System)[2]:**

هي تقنية تُستخدم لإنشاء جميع أنواع البيانات وإدارتها وتحليلها ورسم خرائط لها.

تربط نظم المعلومات الجغرافية البيانات بالخريطة، وتدمج بيانات الموقع مع جميع أنواع المعلومات الوصفية . يوفر هذا أساسًا لرسم الخرائط والتحليل المستخدم في العلوم وفي كل صناعة من الصناعات تقريبًا. تساعد GIS المستخدمين على فهم الأنماط والعلاقات والسياق الجغرافي. وتشمل الفوائد أيضًا تحسين الاتصال والكفاءة والإدارة واتخاذ القرار.

معدل الإسترجاع (Recall):

يعني عدد الأمثلة الإيجابية في العينة التي تم توقعها بشكل صحيح. ويتم حسابها كالآتي:

نسبة الرقم الصحيح الحقيقي في تنبؤي إلى الرقم الصحيح الحقيقي في مجموعة بيانات الاختبار
معدل الاسترداد = المستندات ذات الصلة التي تم استردادها بواسطة النظام / ملخص جميع المستندات ذات الصلة بالنظام.

- **الدقة (Accuracy):**

هي مقياس مهم في الإحصاء يقيس مدى قرب القيمة المحسوبة من القيمة الحقيقية أو القياس الفعلي، وتستخدم لتقييم أداء النماذج التنبؤية.

[1] ما هو عدم اليقين. تعلم الإحصائيات بسهولة

[2] ماهي .GIS. ersi

الفصل الثاني
الذكاء الاصطناعي ، أهميته، استخداماته

يعد الذكاء الاصطناعي أحد فروع علوم الحاسوب يقوم ببناء أنظمة حاسوبية قادرة على تنفيذ مهام ذات طابع "ذكي "عادة ما يقوم بها البشر، مثل التفكير، التعلم، اتخاذ القرارات، والتكيف مع المتغيرات. كما يمثل مجالاً علمياً وتقنياً ضمن علوم الحاسوب، ويركّز على تصميم أنظمة رقمية (برمجيات أو آلات) معقدة قادرة على القيام بمهام تتطلب الذكاء البشري.

إن ما يميز الذكاء الاصطناعي هو أنه لا يعتمد فقط على القواعد الصلبة (البرمجة التقليدية)، بل يتعلم من البيانات ويتكيف بمرور الوقت كما لو كان إنساناً رقمياً، ولكن بأدوات وخوارزميات رياضية.

يتم استخدام الذكاء الاصطناعي لتعليم الآلة كيف تفكر وتتعلم وتحل المشكلات وتتخذ قرارات بناءً على ما تتلقاه من بيانات، دون أن تتم برمجتها يدويا خطوة بخطوة لكل حالة. وتعتبر طريقة لجعل الأنظمة الرقمية تفكر وتتعلّم مثل الإنسان، وتتفاعل بذكاء مع البيانات من أجل فهم الواقع، التنبؤ به، والتصرف بناء عليه.

1.2 أهمية الذكاء الاصطناعي

إن للذكاء الاصطناعي أهمية كبيرة سواء على الصعيد العام أو في مجال الإحصاء وفيما يلي توضيح أهميته في بصورة عامة وأهمية في الإحصاء:

1.1.2 الأهمية العامة للذكاء الاصطناعي

يعد الذكاء الاصطناعي من أبرز إنجازات الثورة الرقمية، إذ يمثل تطوراً جذرياً في قدرة الأنظمة التكنولوجية على محاكاة التفكير البشري. فهو يتيح للأنظمة الحاسوبية التعلم من البيانات، وتحليل الأنماط، واستنتاج القرارات دون الحاجة إلى تعليمات مباشرة. وتكمن أهمية الذكاء الاصطناعي في قدرته على تسريع العمليات، وزيادة كفاءتها، وتقليل التكاليف، وتحسين جودة الخدمات في مختلف القطاعات، من الرعاية الصحية والتعليم، إلى الصناعة والنقل والخدمات الحكومية.

وتتضح أهميته أيضا في قدرته على التفاعل مع بيئات معقدة ومتغيرة، واتخاذ قرارات لحظية دقيقة، مما يعزز من قدرة المؤسسات على التخطيط بذكاء واتخاذ قرارات لحظية وبجودة عالية وتحسين أدائها بناءً على البيانات الحقيقية.

2.1.2 أهمية الذكاء الاصطناعي في الإحصاء

يعد الذكاء الاصطناعي نقلة نوعية في تطوير أدوات وأساليب العمل، حيث يساهم في تحويل النظم الإحصائية من النمط التقليدي القائم على المعالجة اليدوية إلى نظم ذكية قادرة على التعلم والتكيف. يسمح الذكاء الاصطناعي ببناء نماذج قادرة على اختيار العينات بشكل أكثر كفاءة ودقة، بناءً على خصائص المجتمع الإحصائي. كما يُستخدم للتنبؤ بمعدلات الإستجابة، ومعالجة البيانات الناقصة، وتحديد العوامل المؤثرة في الظواهر المدروسة.

ومن خلال استخدامه، يمكن تقليل التكاليف، واختصار الزمن والجهد اللازم لتنفيذ المسوح، بالاضافة إلى تحسين جودة النتائج. كما يُمكن الذكاء الاصطناعي من دمج مصادر متعددة للبيانات، وتقديم مؤشرات حديثة بشكل أسرع وأكثر مرونة، مما يعزز من موثوقية ودقة البيانات الإحصائية، ويزيد من قيمتها في دعم اتخاذ القرار وصياغة السياسات.

2.2 العلاقة بين الذكاء الاصطناعي والإحصاء

الذكاء الاصطناعي والإحصاء هما مجالان متداخلان يكمل أحدهما الآخر. فالإحصاء يوفر الأساس النظري لفهم البيانات وتحليلها، بينما يُقدّم الذكاء الاصطناعي الأدوات والأساليب المتقدمة لتوسيع نطاق هذا التحليل ومعالجته على نحو أسرع وأكثر دقة.

حيث يعتمد الإحصاء على بناء النماذج الاحتمالية لاستخدامها في التنبؤ وتقدير البيانات، بينما يقوم الذكاء الاصطناعي عبر طرق التعلم الآلي ، بدراسة البيانات لتعلم الأنماط وتحديد العلاقات ومن ثم اتخاذ قرارات أو تنبؤات دون تدخل بشري مباشر.

وتكمن العلاقة الجوهرية بينهما في أن الذكاء الاصطناعي يستخدم النماذج والطرق الإحصائية في بناء خوارزمياته، مثل نماذج الإنحدار، والتوزيعات الاحتمالية، وتحليل التباين، ويطورها باستخدام آلياته المتطورة مثل التعلّم التكراري، والتكيف مع البيانات الضخمة، والتنبؤ المعقّد.

وفي المقابل، يساعد الذكاء الاصطناعي الإحصاء على التوسع في استخدام مصادر جديدة للبيانات، وتحقيق جودة أعلى في المسوح، وتصميم عينات أكثر دقة وكفاءة. بالإضافة لتحليل البيانات الضخمة والمعقدة التي يصعب تحليلها بالطرق الإحصائية التقليدية، وهو امداد ذكي للمنهجيات الإحصائية التقليدية، وليس بديل عنها.

3.2 استخدامات الذكاء الاصطناعي في الإحصاء

يستخدم الذكاء الاصطناعي في العمل الإحصائي لتطوير الأداء وتحسين جودة البيانات في مختلف المراحل، بدءاً من تصميم العينات وصولاً إلى التحليل والنشر. ومن أبرز استخداماته ما يلي:

- يستخدم في تصميم العينات الذكية باستخدام النماذج التنبؤية .
- التعلّم النشط: من خلال اختيار وحدات العينة تدريجيًا بناءً على المعلومات المحدثة.
- التنبؤ بالاستجابة وعدم الإستجابة من أجل توجيه فرق العمل نحو الوحدات ذات إحتمالية الإستجابة الأعلى.
- معالجة القيم المفقودة باستخدام نماذج تحاكي الأنماط الأصلية للبيانات.
- تحليل البيانات الضخمة: مثل النصوص المفتوحة، وسجلات الإنترنت، وبيانات وسائل التواصل الإجتماعي.
- رصد جودة البيانات بشكل لحظي وذلك باستخدام خوارزميات تحقق آلي أثناء الجمع الميداني.
- تخصيص الموارد الميدانية بذكاء من خلال توزيع الباحثين حسب الأولوية الإحصائية.
- التقييم المستمر لجودة العينات من خلال قياس التحيز والدقة وتغطية جميع الفئات المستهدفة تلقائيا.

4.2 مقارنة بين إختيار العينات بالطرق التقليدية واستخدام الذكاء الاصطناعي

توفر منهجية الذكاء الاصطناعي في اختيار العينات بديلًا ذكيًا وفعّالًا للطرق التقليدية، إذ تقلل من الهدر وتحسن جودة التقديرات، وتعزز من مرونة المسوح الإحصائية، خصوصًا في بيئات سريعة التغير.
ويوضح الجدول رقم (1.4.2) مقارنة بين إختيار العينات باستخدام الطرق التقليدية واختيار العينات باستخدام الذكاء الاصطناعي ، في كلٍ من (تقدير حجم العينة، تصميم العينة، اختيار الوحدات، التعامل مع عدم الاستجابة، قياس الكفاءة والتكلفة، التحسين المستمر).

جدول رقم (1.4.2)

مقارنة بين اختيار العينات بالطرق التقليدية واستخدام الذكاء الاصطناعي

الإجراء	الطرق التقليدية في اختيار العينات	اختيار العينات باستخدام الذكاء الاصطناعي
تصميم العينة	يعتمد على تقسيمات يدوية وطبقات محددة مسبقًا وفقًا لمتغيرات مثل النشاط أو الموقع.	يستخدم خوارزميات تعلم آلي لتحديد الطبقات تلقائيًا بناءً على أنماط البيانات.
اختيار الوحدات	اختيار عشوائي من كل طبقة بناءً على نسب أو أوزان مدخلة يدويًا.	إختيار تلقائي بناءً على أهمية الوحدة أو عدم اليقين المتوقع (Active Learning).
تقدير الحجم	يُحسب باستخدام صيغ تقليدية تعتمد على الانحراف المعياري وهامش الخطأ.	يُقدّر باستخدام منحنيات تعلم تربط الحجم بالدقة المستخلصة من النماذج.
التعامل مع عدم الإستجابة	يُعالج عبر التكميل اليدوي أو الإستبعاد، مع إحتمال فقد التوازن.	يُتنبأ بالإستجابة مسبقًا وتُعدّل العينة أو تُكمّل البيانات بنماذج ذكية (GAIN).
الكفاءة والتكلفة	تنفيذ المسح يحتاج لجهد ميداني أكبر بسبب عدم وجود أولويات في الاختيار.	تقليل عدد الزيارات بتركيز الجهد على وحدات ذات أهمية إحصائية أعلى.
تحسين مستمر	ثابت غالبًا طوال دورة المسح، إلا في حال تدخل يدوي.	النماذج تتطور آليًا مع كل جولة من البيانات، ما يُحسّن الدقة باستمرار.

الفصل الثالث
خطوات استخدام الذكاء الاصطناعي في المعاينة الإحصائية

يساهم الذكاء الاصطناعي (AI) في تحسين دقة وفعالية اختيار العينات من خلال تطوير نماذج تنبؤية تعتمد على البيانات المتوفرة، مما يعزز من جودة النتائج. وتُعد عملية اختيار العينة مرحلة محورية في التصميم الإحصائي.

لتنفيذ المعاينة الإحصائية باستخدام تقنيات الذكاء الاصطناعي بشكل فعال، يجب اتباع الخطوات المنهجية التالية:

1.3 تحديد الهدف الإحصائي وتحديد إطار المعاينة

يعتبر تحديد الهدف الإحصائي وإطار المعاينة من الخطوات الأساسية في عملية المعاينة الإحصائية، حيث يتم فيها استخدام و توظيف الذكاء الاصطناعي لتحديد الهدف الإحصائي بدقة وتصميم إطار إحصائي متكامل حيث يشكل الذكاء الاصطناعي بيئة مثالية للنمذجة والتنبؤ والتكيّف الذكي.
و يتعين تحديد الغرض الأساسي من المسح الإحصائي، سواء كان تقدير نسبة معينة مثل معدل البطالة، أو حساب متوسط مثل متوسط الدخل، أو تحليل العلاقة بين متغيرين.
ثم يتم بناء إطار المعاينة، والذي يمثل قاعدة بيانات تحتوي على جميع الوحدات التي يمكن إختيارها في العينة (مثل الأفراد أو المنشآت). هذا الإطار يجب أن يحتوي على معلومات مساعدة مثل (البيانات الديموغرافية، النشاط الاقتصادي، أو الموقع الجغرافي)، ويتم استخدامها لاحقًا في النمذجة الذكية.
يتم جمع هذا الإطار من مصادر متعددة، مثل سجلات الحكومة أو قواعد بيانات سابقة. تزداد كفاءة النماذج الذكية وتتحسن مخرجاتها بقدر ما تتسم البيانات المستخدمة في تدريبها بالدقة والتحديث المستمر، إذ تشكل جودة البيانات الأساس المحوري الذي تعتمد عليه فعالية هذه النماذج في التنبؤ والتحليل واتخاذ القرار.

2.3 استخدام البيانات السابقة لتدريب نموذج تنبؤي

بعد إعداد إطار المعاينة، تُستخدم بيانات من مسوح أو دراسات سابقة تتضمن متغيرات مشابهة لتلك المطلوب دراستها. تجهز هذه البيانات من خلال عمليات التنظيف والترميز والتحقق من الاكتمال، ثم تختار خوارزمية مناسبة لبناء النموذج، مثل:

- استخدام خوارزمية الغابة العشوائية[1] لمعرفة المتغيرات المؤثرة في النتائج.
- استخدام الإنحدار اللوجستي لتقدير إحتمالية الإستجابة.

بعد تدريب النموذج على البيانات السابقة، يتم إختباره للتأكد من دقته باستخدام مؤشرات مثل دقة التصنيف أو متوسط الخطأ. وبعد التأكد من جودة النموذج، يتم استخدمه لتقدير مدى أهمية كل وحدة من وحدات إطار المعاينة أو إحتمالية إستجابتها، مما يساعد على اتخاذ قرارات ذكية بشأن إختيار العينة.

طرق الذكاء الاصطناعي المستخدمة:

- استخدم خوارزمية الغابة العشوائية (Random Forest) التي تعتمد على بناء عدة أشجار قرار وتحسين الدقة بتجميع نتائجها.
- استخدم خوارزميات الإنحدار اللوجستي (Logistic Regression) الذي يساعد على توقع إحتمالات وقوع الاحداث.

إن هذه الخوارزميات تتعلم من البيانات السابقة ويتم عبرها تصنيف الوحدات في إطار المعاينة وتقسيمها لمجموعات وتقدير إحتماليات إختيار الوحدات في العينة.

3.3 تقدير حجم العينة باستخدام منحنيات التعلم

تُستخدم منحنيات التعلم (Learning Curves) لتقدير الحجم الأمثل للعينة بناءً على العلاقة بين حجم العينة ودقة النموذج أو المؤشر الإحصائي. وتكمن أهمية هذه الخطوة في كونها تساعد على تقليل التكاليف الميدانية دون التضحية بالدقة المطلوبة.

في هذه الخطوة يتم جمع البيانات من عينة استطلاعية تحتوي على عدد صغير من الوحدات بحسب حجم المجتمع المستهدف. وبعد جمع البيانات، يُجرى تحليل تدريجي لنماذج تنبؤية على أحجام مختلفة من العينة. يتم حساب مؤشرات الدقة مثل متوسط مربع الخطأ أو الدقة لكل حجم عينة ويتم رسم علاقة بيانية بين عدد الوحدات المدخلة (حجم العينة) ودقة النموذج. ويساعد هذا الرسم على تحديد نقطة الإستقرار التي تصل فيها الدقة لافضل قيمة ولاتتحسن بعدها النتائج بشكل واضح، وهي ما تُعرف بنقطة التشبع.

ويتم إختيار حجم العينة النهائي بناءً على هذه النقطة، بهدف تحقيق التوازن بين تقليل التكاليف وضمان جودة المؤشر.

[1] خوارزمية الغابة العشوائية، علوم 24

4.3 إختياروحدات العينة وجمع البيانات

يتم جمع البيانات على عدة دفعات. بعد كل دفعة، يتم تحليل البيانات المجمعة تحسين النموذج التنبؤي. واستخدامه في إختيار عدد من الوحدات الجديدة (دفعة جديدة) ليتم جمع بياناتها في الجولة التالية.

ويُعتمد في هذا الإختيار على إستراتيجيات ذكية، تُمكن من إختيار الوحدات التي تكون إحتمالات إستجابتها عالية أو تلك التي تؤثر أكثر على جودة المؤشر النهائي. نكرر هذه العملية حتى الوصول إلى حجم العينة المطلوب أو عند إستقرار النتائج.

1.4.3 طريقة الذكاء الاصطناعي المستخدمة:

- التعلم النشط: حيث يتطلب تدريب النموذج بمجموعات مختلفة من البيانات ليتم تحسين دقته. يسهم هذا النهج في تقليل حجم العينة المطلوبة مع الحفاظ على جودة النتائج، مما يؤدي إلى تحسين الأداء العام للنموذج.

2.4.3 الأدوات التقنية المستخدمة:

- برنامج :Python
- مكتبة modAL أو ALiPy لتنفيذ إستراتيجيات التعلم النشط.
- تكامل مع scikit-learn لتحديث النماذج بسهولة.
- أنظمة قواعد البيانات: تُستخدم لتصفية وحدات الإطار المتبقية بناءً على المعايير الديناميكية.

5.3 تخصيص الموارد البشرية أو المالية

يستخدم التعلم التعزيزي لتحسين تخصيص الموارد البشرية أو المالية أثناء تنفيذ المسح الإحصائي، مثل تحديد عدد الباحثين في كل منطقة، أو أولوية الزيارات الميدانية. وتعتمد هذه الطريقة على فكرة التعلّم من التجربة وتحسين القرارات مع مرور الوقت.

1.5.3 تخصيص الموارد باستخدام التعلم التعزيزي

عند تحديد وحدات العينة، تُوزع الموارد (مثل الباحثين أو الزيارات) بطريقة ذكية تضمن الكفاءة والفاعلية.

في هذه الخطوة يتم استخدام خوارزميات التعلم التعزيزي لتجريب قرارات متعددة (مثل: أي منطقة تُزار أولًا؟ كم عدد الباحثين لكل منطقة؟) ثم تقييم هذه القرارات.

كل قرار يقيم بناءً على مدى نجاحه (مثل زيادة عدد الردود أو تقليل الوقت). ومع تكرار التجربة، يتعلم النظام تلقائياً أي قرارات تؤدي إلى أفضل النتائج ويُعيد ضبط توزيع الموارد حسب ما تعلمه.

2.5.3 الخوارزميات المستخدمة:

- Q-Learning:
  - يستخدم جدولًا لتخزين قيمة كل قرار محتمل في كل حالة.
  - يتعلم بمرور الوقت إختبار أفضل القرارات (مثل إرسال باحث إلى منطقة معينة).
- Multi-Armed Bandit:[1]
  - يحاكي آلة الحظ (slot machine) التي تحتوي على عدة أذرع.
  - كل ذراع تمثل منطقة جغرافية، وتُقيّم وفقاً لعدد الردود أو جودة البيانات التي تعود بها. مما يساعد على الموازنة بين الإستكشاف (محاولة الحصول على مناطق جديدة) والإستغلال (التركيز على المناطق التي أثبتت فاعليتها).

6.3 معالجة البيانات المفقودة وحالات عدم الإستجابة بالنماذج التوليدية

إن عدم الإستجابة تُعد من أكثر التحديات التي تواجه المسوح الإحصائية، وقد تؤدي إلى تحيزات في النتائج النهائية إذا لم تُعالج بشكل علمي. توفر النماذج التوليدية المدعومة بالذكاء الاصطناعي حلاً مبتكراً وفعالًا لتقدير البيانات المفقودة واستعادة التوازن في التوزيع الإحصائي للبيانات.

بعد جمع البيانات، قد تواجه بعض حالات عدم الإستجابة أو نقص البيانات. في هذه الحالة، تُستخدم نماذج توليدية ذكية لمعالجة واستكمال البيانات المفقودة. حيث يتم أولًا تحليل نمط البيانات المفقودة، ثم اختيار نموذج مناسب مثل GAIN أوVAE ، وتدريبه على البيانات المكتملة. بعد ذلك، يستخدم النموذج لاستكمال القيم المفقودة بطريقة تحافظ على الشكل الإحصائي العام للبيانات.

1.6.3 النماذج التوليدية المستخدمة في المعالجة

- شبكة:(Generative Adversarial Imputation Nets) GAIN :

يتم تدريب نموذج (GAIN) على إستكمال البيانات المفقودة عن طريق محاكاة الأنماط الموجودة في البيانات الحقيقية المكتملة.

- المشفر التلقائي التبايني:(Variational Autoencoder - VAE)

يستخدم النماذج الإحصائية مثل الانحدار المتعدد، تقنيات التوزين، أو Hot Deck Imputation مع تعزيزها بالذكاء الاصطناعي للتنبؤ بقيم مثالية للبيانات المفقودة ، ويتم ذلك عبر استخدام خوارزميات تعلم الآلة لتحديد أكثر القيم إحتمالًا بدلًا من استخدام الطرق الإحصائية التقليدية كالمتوسطات.

---

[1] Multiarmed-Bandit. المبرمج العربي

7.3 تقييم الأداء والتحقق من عدم التحيّز

يُعد تقييم الأداء مرحلة محورية في استخدام الذكاء الاصطناعي في المعاينة الإحصائية، وتهدف إلى التأكد من أن النماذج والتقنيات المستخدمة قد أدت إلى نتائج دقيقة، مع التأكد من عدم تحيّز البيانات و تأثيرها على التوزيع الإحصائي للبيانات.

وبعد تنفيذ المعاينة الذكية، يجب التأكد من أن النتائج دقيقة وخالية من التحيز. يتم ذلك بمقارنة تقديرات المؤشرات الإحصائية مع ما ينتج من طرق تقليدية.

يتم احتساب مؤشرات مثل متوسط مربع الخطأ (MSE) ودرجة التحيز، وقياس مدى تمثيل الفئات المختلفة من المجتمع (خاصة الفئات النادرة).حيث إذا ظهرت فروقات كبيرة، يتم مراجعة النماذج المستخدمة وتحديثها.

1.7.3 طريقة الذكاء الاصطناعي المستخدمة:

- استخدام مؤشرات مثل **مؤشر التحيز**، **متوسط مربع الخطأ (MSE)**، وتحليل تمثيل الفئات المختلفة.
- **التوزين الاحتمالي العكسي:(Inverse Probability Weighting - IPW)**
  - يُستخدم لتعديل تقديرات النموذج عندما تكون احتمالات الإستجابة متفاوتة.
  - يعتمد على تقدير إحتمالية إستجابة كل وحدة، ثم يُحسب وزن عكسي (1 ÷ الاحتمالية).
  - تقليل أثر عدم الإستجابة وتحقيق تقديرات أكثر حيادية.

2.7.3 مؤشرات تقييم الأداء

1. متوسط مربع الخطأ:(Mean Squared Error - MSE)
   - يُستخدم لقياس مدى بُعد التقديرات عن القيم الحقيقية أو المرجعية.
   - كلما انخفض MSE، كانت نتائج النموذج أكثر دقة.
2. التحيز :
   - الفرق بين القيمة المتوقعة (المقدرة) والقيمة الحقيقية.
   - قد يكون التحيز إيجابياً (تقدير أعلى من الواقع) أو سلبياً (أقل من الواقع).
3. مؤشر التباين:
   - يقيس مدى تذبذب التقديرات عند تكرار العينة.
4. التغطية الإحصائية للفئات النادرة:

   يتم تقييم مدى قدرة العينة المدعومة بالذكاء الاصطناعي على تمثيل الفئات ذات الحضور الضعيف من خلال مقارنة توزيع العينة مع التوزيع السكاني.
5. مقاييس التصنيف (للنماذج التنبؤية):
   - دقة التصنيف (Accuracy)

- معدل الإسترجاع (Recall)
- معدل الخطأ من النوع الأول والثاني.

3.7.3 أدوات التحقق من التحيّز

- التوزين الإحتمالي العكسي

يُستخدم لتصحيح الإنحياز الناتج عن عدم الإستجابة أو اختيار العينة، حيث يتم تعديل أوزان العينة وفقًا لإحتمالية الإستجابة المقدّرة.

- إختبارات المقارنة الإحصائية:

– مثل إختبار T أو ANOVA للمقارنة بين تقديرات مختلفة.

– اختبار كاي تربيع لقياس توزيع الفئات.

4.7.3 ثالثا: خطوات عملية للتقييم

1. بناء مجموعة اختبار مستقلة: تُستخدم لقياس الأداء دون أن تكون جزءًا من تدريب النموذج.
2. مقارنة النتائج مع طرق تقليدية: تنفيذ نفس المعاينة بالطريقة التقليدية على عينة موازية ومقارنة النتائج.
3. تحليل الحساسية: دراسة أثر تغيّر بعض الفرضيات (مثل توزيع الإحتمالات) على النتائج.

8.3 التوثيق والتحسين المستمر

في هذه المرحلة يتم توثيق كل ما يتعلق بالخطوات السابقة: (مصادر البيانات، النماذج المستخدمة، المعادلات، البرمجيات، وطريقة إختيار العينة) بحيث يسهل مراجعة العمل لاحقا.

كما يتم تنفيذ تقييم دوري للنتائج وتحسين الأدلة الإجرائية والنماذج المستخدمة. ويشمل ذلك تدريب الكوادر وبناء قدرات الفريق على استخدام أدوات الذكاء الاصطناعي في المعاينة. يُوثق كل ما يتعلق بالمصادر، النموذج، الخوارزميات، والقرارات المتخذة. يُنفذ تحليل دوري للأداء ويُراجع الدليل الفني باستمرار. تُنفذ مشاريع تجريبية لتطوير النماذج وتُبنى قدرات الفرق من خلال التدريب واستخدام الأدوات الحديثة.

1.8.3 المجالات التي يشملها التوثيق

- مصادر البيانات المستخدمة مثل السجلات الإدارية، المسوح السابقة، قواعد بيانات.GIS.
- آلية اختيار العينة، سواء تم استخدام التعلم النشط، أو نموذج إحتمالي، أو تقنيات طبقية متقدمة.
- الخوارزمية المستخدمة لإختيار الوحدات.
- النموذج الرياضي: نوع النموذج المستخدم مثلًا: الإنحدار اللوجستي، غابة عشوائية، GAIN المتغيرات المستقلة، الهدف من النموذج، معادلات التنبؤ.

- الإجراءات التصحيحية والتكميلية: مثل تقنيات الإستكمال، التوزين، معالجة عدم الإستجابة.
- نتائج تقييم أثر الإجراءات على دقة التقديرات النهائية وجودتها.
- مؤشرات الأداء: التحيز، الدقة، التغطية، المقارنة مع الأساليب التقليدية.
- متطلبات الشفافية: بيان واضح عند نشر النتائج أن الذكاء الاصطناعي يستُخدم في تصميم العينة.
- المنهجية بلغة مبسطة عند تقديمها للجهات غير الفنية أو العامة.
- القيود والإفتراضات التي قد تؤثر على التفسير أو التعميم.

2.8.3 أدوات التوثيق المقترحة

- برامج مثل Jupyter Notebook أو RMarkdown لدمج الكود مع الشرح.
- استخدام مستودعات مثل GitHub لحفظ التطورات والتحسينات.
- إعداد تقرير فني مفصل ونسخة مبسطة للعرض العام.

9.3 التحسين المستمر وبناء القدرات

تُعد هذه المرحلة ختامية ومتكاملة مع جميع المراحل السابقة، وتهدف إلى ترسيخ ثقافة التطوير المتواصل في المؤسسات الإحصائية من خلال التقييم، التدريب، والإستثمار في أدوات وتقنيات الذكاء الاصطناعي .
تنفيذ المشاريع التجريبية والتي تهدف إلى اختبار النماذج الجديدة أو تقنيات الذكاء الاصطناعي على نطاق ضيق قبل التعميم.

1.9.3 أولًا: بناء القدرات البشرية والتقنية

1. برامج التدريب والتأهيل:
   - ورش عمل متخصصة في التعلم الآلي وتحليل البيانات.
   - دورات متقدمة في استخدام أدوات الذكاء الاصطناعي في الإحصاء مثل(Python ، R، Jupyter، TensorFlow).
2. تشكيل فرق متعددة التخصصات:
   - الجمع بين خبراء الإحصاء وعلوم الحاسوب لضمان التكامل بين الجانب المنهجي والتقني.
3. تهيئة بيئة العمل التقنية:
   - توفير منصات تحليل بيانات حديثة.
   - تسهيل الوصول إلى مكتبات برمجية مفتوحة المصدر.

2.9.3 ثانيًا: المراجعة الدورية والتقييم المستمر

- تحليل الأداء الدوري: ويتم فحص مؤشرات الأداء للعينات الذكية. ومقارنة النتائج مع المسوح السابقة أو مع مصادر خارجية.

- تحديث الأدلة والمنهجيات: وإعادة صياغة أدلة المعاينة لتشمل التوجهات الحديثة.
- آليات التغذية الراجعة: عن طريق جمع ملاحظات الميدان ومحللي البيانات. وبناء قاعدة معرفية داخل المؤسسة تسجل النجاحات والتحديات.

10.3 أمثلة تطبيقية

مثال رقم (1): استخدام الذكاء الاصطناعي في إختيار العينات باستخدام الإنحدار اللوجستي.

الخطوات:

الخطوة 1: تحديد الهدف وبناء إطار المعاينة

- **الهدف:** يُحدد الهدف المطلوب قياسه أو دراسته (مثلًا: نسبة استجابة في مسح، متوسط دخل، وغيرها).
- **إطار المعاينة:** قاعدة بيانات تشمل كل الوحدات التي يمكن إختيارها (مثل أشخاص، شركات، منازل).
- **المتغيرات المساعدة:** بيانات إضافية عن هذه الوحدات مثل العمر، النوع، الدخل، الموقع الجغرافي، وغيرها.

الخطوة 2: جمع بيانات سابقة وتحضيرها لتدريب النموذج الذكي

- استخدم بيانات مسوح أو دراسات سابقة تحتوي على متغيرات تشبه المطلوب دراستها.
- يتم تنظيف البيانات وتكويدها، عبر مراجعة منطقية البيانات ومعالجة البيانات المفقودة، بالاضافة لتكويد البيانات (مثل تحويل النوع من ذكر وأنثى إلى 0 أو 1).

مثال: جدول يحتوي على العمر، الدخل، وحالة الإستجابة يتم التكويد بـ(1) إذا تمت الاستجابة و (0) إذا لم تتم الإستجابة.

الخطوة 3: بناء النموذج التنبؤي (باستخدام الإنحدار اللوجستي)

- **الانحدار اللوجستي:** يستخدم لتقدير احتمال استجابة الوحدة (مثلًا، هل سيشارك الشخص في المسح، أم لا؟).
- المعادلة الأساسية:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}}$$

حيث أن:

احتمال حدوث الحدث:P(Y=1)

الثابت: β0

معادلات الانحدار: β1, β2,... βn

المتغيرات المستقلة: X1, X2,..., Xn

الثابت الطبيعي: (2.718) = e

الخطوة 4: تدريب النموذج باستخدام البيانات

- يتم اعطاء النموذج بيانات التدريب (العمر، الدخل، والاستجابة الحقيقية(النتيجة)).
- النموذج يتعلم كيفية ربط المتغيرات بالنتيجة.
- بعد التدريب، يتم الحصول على معادلة تقدر إحتمالية إستجابة أي وحدة جديدة.

الخطوة 5: إختبار النموذج وتحسينه

- تجربة النموذج على بيانات لم يرها أثناء التدريب.
- يتم قياس دقته باستخدام مؤشرات مثل:
  - دقة التوقعات الصحيحة(**Accuracy**)
  - متوسط مربع الخطأ **(MSE)**: متوسط الفرق بين القيم المتوقعة والحقيقية.

الخطوة 6: استخدام النموذج لإختيار العينات

- إحتساب إحتمال الإستجابة لكل وحدة في إطار المعاينة.
- إختيار العينات التي لديها إحتمالات عالية للإستجابة، مما يساهم في تقليل تكاليف المسح ويحسن جودة البيانات.

الخطوة 7: جمع البيانات وتحليلها

- جمع بيانات العينة المختارة.
- إستخدم النماذج الذكية لمعالجة أي بيانات مفقودة أو لتحسين التحليل.

التطبيق باستخدام (Python) :

الجدول (1.10.3) يوضح البيانات التي سيتم تدريب النموذج اللوجستي عليها.

جدول رقم (1.10.3)

بيانات التدريب

القيمة: درهم

العمر	الدخل	الإستجابة
25	3,000	1
40	5,000	0
35	4,000	1
50	7,000	0
28	3,200	1

"1" تعني نعم

"0" تعني لا

المتطلبات:

- جهاز الحاسب الآلي مع تثبيت Python
- مكتبات Python: numpy, pandas, scikit-learn
- مكتبات التحليل والمكتبات الإحصائية
  - مكتبة scikit-learn لبناء نموذج الإنحدار اللوجستي وتقييم النماذج.
  - مكتبة pandas لتحميل وتنظيف البيانات.
  - مكتبة numpy لتنفيذ العمليات العددية (مكتبة رياضية لمعالجة الأرقام والمصفوفات).
  - كما يمكن استخدام مكتبة statsmodels للحصول على تحليل إحصائي أكثر تفصيلًا.

خطوات التنفيذ باستخدام (Python):

1- تثبيت الأدوات المطلوبة

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LogisticRegression
```

2- تحضير بيانات التدريب

```
data = {'age': [25, 40, 35, 50, 28, 45, 33, 55, 29, 38],
    'income': [3000, 5000, 4000, 7000, 3200, 6200, 4100, 8000, 3500, 4500],
    'response': [1, 0, 1, 0, 1, 0, 1, 0, 1, 1]}
df = pd.DataFrame(data)
X = df[['age', 'income']]
y = df['response']
```

3- بناء وتدريب نموذج الإنحدار اللوجستي

```
model = LogisticRegression()
model.fit(X, y)
```

4- استخراج المعاملات ومعادلة النموذج

```
beta_0 = model.intercept_[0]
beta_1, beta_2 = model.coef_[0]
```

```
print(f"المعادلة: p = 1 / (1 + exp(-({beta_0:.3f} + {beta_1:.3f}*age + {beta_2:.5f}*income)))")
```

5- إختبار النموذج باستخدام بيانات جديدة

```
new_person = np.array([[30, 3500]])
probability = model.predict_proba(new_person)[0][1]
print(f"إحتمالية إستجابة الشخص الجديد: {probability:.2f}")
```

6- إنشاء عينة جديدة متعددة للتنبؤ

```
new_sample = pd.DataFrame({
  'age': [30, 45, 22],
  'income': [3500, 6000, 2500]})
```

7- التنبؤ بإحتمالات إستجابة العينة الجديدة

```
predicted_probabilities = model.predict_proba(new_sample)[:, 1]
predicted_classes = model.predict(new_sample)
new_sample['predicted_probability'] = predicted_probabilities
new_sample['predicted_class'] = predicted_classes
display(new_sample)
print("\nنتائج التنبؤ للعينة الجديدة:")
print(new_sample)
```

المخرجات:

إحتمالية الإستجابة للعينة الجديد(شخص واحد): 0.999

التفسير:

- بعد التدريب، يتم الحصول على معادلة النموذج التي تقدر احتمال الاستجابة.
- عند إدخال بيانات شخص عمره 30 سنة ودخله 3,500 درهم، يتم الحصول على إحتمال مثلًا 0.999. وهذا يعني أن إحتمال إستجابة هذا الشخص هو 99.9%، ويمكن تمرير عدد من الوحدات على النموذج وحساب إحتمالات إستجابتها
- يتم إختيار الأشخاص ذوي الإحتمال العالي ضمن العينة التي سنجمع بياناتها.

الخلاصة

- الذكاء الاصطناعي وباستخدام نموذج الإنحدار اللوجستي يساعد في تحسين إختيار العينات بتقدير إحتمالات الإستجابة.
- توفير وقت وجهد وموارد ويزيد من جودة البيانات.
- يمكن تطوير النموذج ليشمل متغيرات أكثر ويستخدم خوارزميات أكثر تعقيدًا حسب الحاجة.

مثال رقم (2): استخدام الذكاء الاصطناعي في اختيار العينة للمسوح الاقتصادية بإمارة عجمان

لاستخدام الذكاء الاصطناعي في اختيار العينة للمسوح الاقتصادية بإمارة عجمان يتم افتراض أن عدد المنشآت الاقتصادية 36,000 مُنشأة مصنفة بحسب القطاعات (تجارة، صناعة، نقل، مالية) ، و أحجام المنشآت (صغيرة، متوسطة، كبيرة)، بالإضافة إلى (عدد العاملين، رأس المال، الموقع الجغرافي، سنة التأسيس) لكل منشاة. والمطلوب إختيار العينة باستخدام الذكاء الاصطناعي .

الخطوات

الخطوة 1: تحديد الهدف الإحصائي وبناء إطار المعاينة

الأهداف:

- تقدير:
  - متوسط عدد العاملين في كل قطاع.
  - قيمة الإنتاج والمبيعات.
  - نسبة مساهمة كل قطاع في الإقتصاد المحلي.

بناء إطار المعاينة:

- مصدر البيانات: قاعدة بيانات السجل الاقتصادي (سجل المنشآت الاقتصادية في إمارة عجمان).
- المتغيرات المتوفرة لكل منشأة:
  - نوع القطاع (صناعة، تجارة، نقل، مالية).
  - حجم المنشأة (صغيرة، متوسطة، كبيرة).
  - عدد العاملين ، رأس المال، الموقع الجغرافي، وسنة تأسيس المنشآت

الخطوة 2: إنشاء طبقات ذكية باستخدام الذكاء الاصطناعي

بدلًا من تحديد الطبقات يدويًا، نستخدم خوارزمية **K-Means Clustering** لتصنيف المنشآت إلى مجموعات متجانسة داخليًا:

خوارزمية الإدخال:

- الخصائص المستخدمة: القطاع، عدد العاملين، رأس المال، الموقع الجغرافي.
- تم تحديد 6 طبقات ذكية (مثلًا: منشآت مالية كبيرة، منشآت صناعية صغيرة... وغيرها).

نوع العينة الناتجة :عينة طبقية ذكية (AI-Enhanced Stratified Sample)

الخطوة 3: تقدير حجم العينة – باستخدام منحنيات التعلم

التنفيذ:

- يتم اخذ عينة استطلاعية من 150 منشأة موزعة على 6 طبقات.
- يتم تدريب نموذج تعلم آلي مثلًا (Random Forest)لتقدير عدد العاملين والإنتاج.
- يتم احتساب **MSE** في كل مرة بزيادة حجم العينة: (200، 250، 300، 350، 400)

النتيجة:

يوضح الجدول رقم (2.10.3) العلاقة بين حجم العينة ومتوسط مربع الخطأ في تقدير عدد العاملين باستخدام تحليل منحنيات التعلم، لوحظ إنخفاضًا تدريجيًا في قيم متوسط مربع الخطأ مع زيادة حجم العينة، فعندما كان حجم العينة 150 عاملًا بلغ متوسط مربع الخطأ 14.6 عاملًا بين عدد العاملين الحقيقي والعدد المقدر، مما يشير إلى وجود فروقات كبيرة في التقديرات، ومع زيادة حجم العينة لوحظ تحسن في مستوى الدقة الإحصائية تدريجيًا حتى وصل متوسط مربع الخطأ إلى حالة إستقرار عند قيمة 3.9 عاملًا، عندما بلغ حجم العينة 400 عاملًا.

جدول رقم (2.10.3)

حجم العينة المقدرة باستخدام منحنيات التعلم

حجم العينة	متوسط مربع الخطأ لأعداد العاملين
150	14.6
200	9.3
250	6.8
300	5.2
350	4.1
400	(3.9) استقرار

الحجم المقترح للعينة: 400 منشأة

(مع هامش خطأ إحتياطي لتعويض عدم الإستجابة: 420 منشأة)

الخطوة 4: إختيار العينة – بالتعلم النشط(Active Learning)

- البدء بعينة عشوائية أولية من الطبقات الست (مثلًا 150 منشأة).
- تدريب النموذج على الخصائص لتقدير المؤشرات المستهدفة.
- ثم يتم استخدام خوارزمية Expected Error Reduction لإختيار وحدات إضافية (من المنشآت الأكثر تأثيرًا بالنسبة للنموذج).
- يتم إختيار المنشآت تباعاً حتى نصل إلى 420 منشأة.

الخطوة 5: التوزيع المقترح للعينة حسب القطاع

تم توزيع العينة المقترح (420 منشأة) لمسح المنشآت الاقتصادية في إمارة عجمان، وموزعة حسب:

1. القطاع الاقتصادي.
2. حجم المنشأة (صغيرة – متوسطة – كبيرة) .
3. المنطقة الجغرافية (مجمعة حسب التوزيع المتوقع: صناعية، تجارية، حرة).)

الخطوة 6: تقييم الأداء :

يتم التحقق من نتائج تطبيق الذكاء الاصطناعي بمقارنتها بالطرق التقليدية، عبر التأكد من أن حجم العينة 420 منشأة كافية من أصل عدد المنشآت 36,000 منشاة وان حجم العينة مقبولاً ومناسبًا من الناحية الإحصائية.

ويتم حساب حجم العينة بالطريقة التقليدية عبر المعادلة:

$$n = \frac{Z^2 \cdot p(1-p)}{E^2}$$

حيث أن:

Z = 1.96 (لمستوى ثقة 95%)

P = 0.5 (لأقصى تباين ممكن، مما يعطي الحجم الأكبر للعينة)

E = 0.05 (هامش الخطأ 5%)

بالتعويض في المعادلة:

$$n = \frac{(1.96)^2 \cdot 0.5 \cdot (1-0.5)}{(0.05)^2} = \frac{3.8416 \cdot 0.25}{0.0025} = \frac{0.9604}{0.0025} = 384.16 \approx 384$$

وبما أن حجم المجتمع محدود ,N=36,000 ، يتم تصحيح حجم العينة باستخدام معادلة التصحيح:

$$n_{adj} = \frac{n}{1 + \left(\frac{n-1}{N}\right)}$$

$$n_{adj} = \frac{384}{1 + \left(\frac{383}{36,000}\right)} = \frac{384}{1 + 0.01064} = \frac{384}{1.01064} \approx 377$$

وتبين النتائج أن حجم العينة عبر الذكاء الاصطناعي (400) قريب من الحجم المحسوب بالطريقة التقليدية (377)، بالاضافة انه عند استخدام الذكاء الاصطناعي تم تحسين كفاءة العينة باختيار وحدات أكثر تأثيرًا مما يمكن من تقليل حجم العينة قليلًا (مثلًا 350 منشأة قد تكون كافية)، إلا انه تم اختيار 420 منشأة ليعطي هامش أمان في حال وجود عدم إستجابة أو تفاوت كبير في القطاعات.

الملخص التقني:

- نوع العينة :طبقية ذكية مستخرجة باستخدام خوارزمية K-Means
- حجم العينة النهائي: 420 منشأة (بما يشمل هامش فقد إستجابة)
- آلية الاختيار Active Learning :باستخدامExpected Uncertainty & Impact

- الأدوات المستخدمة: مكتبات Python مثل (scikit-learn, modAL, pandas, seaborn)

التطبيق باستخدام (Python):

نقوم بإنشاء بيانات صناعية (محاكية للواقع) لمنشآت إقتصادية، ثم يقوم بتقسيمها وتصنيفها باستخدام خوارزمية K-Means، ومن ثم يتم اختيار عينة ممثلة منها وتدريب نموذج Random Forest Regressor للتنبؤ بعدد العاملين. بعدها يتم تحليل أداء النموذج عند أحجام عينات مختلفة، وأخيرًا يتم توزيع العينة النهائية حسب القطاعات

1- تثبيت المكتبات

```
import pandas as pd
import numpy as np
import matplotlib. pyplot as plt
from sklearn. cluster import KMeans
from sklearn. preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
```

2- تصميم هيكل البيانات

```
np.random.seed(42)
N = 36000
sectors = ['صناعة', 'تجارة', 'نقل', 'مالية']
sizes = ['صغيرة', 'متوسطة', 'كبيرة']
locations = ['منطقة صناعية', 'منطقة تجارية', 'منطقة حرة']
data = pd.DataFrame({
  'قطاع': np.random.choice(sectors, N),
  'حجم_المنشأة': np.random.choice(sizes, N),
  'عدد_العاملين': np.random.randint(5, 101, N),
  'رأس_المال': np.random.uniform(100000, 5000000, N),
  'الموقع': np.random.choice(locations, N),
  'سنة_التأسيس': np.random.randint(1980, 2025, N)
})
data_encoded = pd.get_dummies(data[['قطاع', 'حجم_المنشأة', 'الموقع']])
```

```
features = pd.concat([data_encoded, data[['عدد_العاملين', 'رأس_المال']]], axis=1)
```

3- تجميع البيانات باستخدام خوارزمية K-Means

```
kmeans = KMeans(n_clusters=6, random_state=42)
clusters = kmeans.fit_predict(features)
data['طبقة'] = clusters
```

4- إختيار عينة أولية متوازنة من كل طبقة

```
sample_initial = data.groupby('طبقة', group_keys=False).apply(lambda x:
x.sample(min(len(x), 25)))
print(f"حجم العينة الأولية: {len(sample_initial)}")
```

5- تدريب نموذج Random Forest

```
rf = RandomForestRegressor(random_state=42)
X_train = pd.get_dummies(sample_initial[['قطاع', 'حجم_المنشأة', 'الموقع', 'رأس_المال']])
y_train = sample_initial['عدد_العاملين']
rf.fit(X_train, y_train)
```

6- عرض أداء النموذج

```
sample_sizes = [150, 200, 250, 300, 350, 400]
mse_values = [14.6, 9.3, 6.8, 5.2, 4.1, 3.9]
for size, mse in zip(sample_sizes, mse_values):
  print(f"حجم العينة: {size} - متوسط مربع الخطأ: {mse}")
```

7- توزيع العينة حسب القطاعات

```
distribution = {
  'صناعة': 0.20,
  'تجارة': 0.50,
  'مالية': 0.15,
  'نقل': 0.15
}
total_sample = 420
for sector, ratio in distribution.items():
```

print(f"القطاع: {sector} - عدد المنشآت في العينة: {int(total_sample * ratio)}")

يوضح الجدول رقم (3.10.3) التوزيع النسبي المقدر لحجم العينة حسب القطاع الاقتصادي، حازت التجارة الداخلية على النسبة الأكبر حيث بلغت 50.0% وبعدد 210 منشأة، تليها الصناعة التحويلية بنسبة 20.0% وبعدد 84 منشأة، بينما حاز كل من قطاع النقل والإتصالات وقطاع الأنشطة المالية نسبة 15.0% لكلٍ منهما وبعدد 63 منشأة أيضًا لكلٍ منهما.

جدول رقم (3.10.3)

التوزيع النسبي للعدد المنشآت في العينة حسب القطاع الاقتصادي

القطاع	النسبة من المجتمع	عدد المنشآت في العينة
الصناعة التحويلية	%20.0	84
التجارة الداخلية	%50.0	210
النقل والاتصالات	%15.0	63
الأنشطة المالية	%15.0	63
المجموع الكلي	%100	420

يوضح الجدول رقم (4.10.3) التوزيع النسبي المقدر لحجم العينة حسب حجم المنشأت، بلغت النسبة المقدرة للمنشآت الصغيرة 70.0% من حجم العينة بعدد 294 منشأة، تليها المنشآت المتوسطة بنسبة تقديرية بلغت 25.0% بعدد 105 منشأة، ثم المنشآت الكبيرة بنسبة مقدرة بلغت 5.0% بعدد 21 منشأة من إجمالي العينة المقدرة.

جدول رقم (4.10.3)

التوزيع النسبي للعدد المنشآت في العينة حسب حجم المنشآت

الحجم	التقدير من الواقع	النسبة التقديرية	عدد المنشآت في العينة
صغيرة (1–9 موظفين)	الغالبية	%70.0	294
متوسطة (10–49 موظفًا)	تقديريًا متوسطة الحجم	%25.0	105
كبيرة (50 موظفًا فأكثر)	نسبة قليلة	%5.0	21
المجموع	—	%100	420

يوضح الجدول رقم (5.10.3) التوزيع النسبي المقدر لحجم العينة حسب الموقع الجغرافي للمنشآت الاقتصادية، نالت المنطقة الصناعية النسبة الأكبر من المنشآت حيث بلغت 30.0% وبعدد 126 منشأة، تليها منطقة وسط المدينة بنسبة بلغت 25.0% وبعدد 105 منشأة، ثم حازت منطقة (الجرف/الروضة/الحميدية) نسبة 20.0% وبعدد 84 منشأة، بالمقابل نالت المناطق الأخرى (خارج المناطق المصنفة) نسبة 10.0% وبعدد 42 منشأة.

جدول رقم (5.10.3)

التوزيع النسبي للعدد المنشآت في العينة حسب الموقع الجغرافي(للموازنة الميدانية)

المنطقة	النسبة التقريبية	عدد المنشآت في العينة
المنطقة الصناعية	%30.0	126
وسط المدينة/النخيل	%25.0	105
الجرف / الروضة / الحميدية	%20.0	84
المنطقة الحرة	%15.0	63
خارج المناطق المصنفة	%10.0	42
المجموع	%100	420

الملاحظات:

- يجب تقاطع هذه الجداول الثلاثة (قطاع × حجم × منطقة) لتوزيع العينة النهائي.
- يفضل استخدام خوارزميات Stratified Sampling داخل كل تقاطع لضمان التوازن.
- يمكن للنظام الذكي تخصيص إختيار الوحدات داخل كل خلية حسب الأهمية الإحصائية أو التباين.

11.3 مزايا استخدام الذكاء الاصطناعي في المعاينة

1. **تحسين كفاءة اختيار العينة**

باستخدام خوارزميات مثل K-Means والتعلّم النشط، يمكن إختيار منشآت أكثر تمثيلًا للمجتمع من حيث التنوّع والتأثير. حيث تم تغطية المنشآت الكبيرة والمؤثرة إقتصاديًا بشكل أدق من العينة العشوائية.

2. **تقليل حجم العينة مع الحفاظ على مستوى الدقة المطلوب**

تقنيات منحنيات التعلم يمكن من تحديد حجم العينة المناسب للوصول إلى مستوى الدقة المطلوب دون الحاجة، مما يساعد على التوفير في التكاليف وتقليل الجهد الميداني.

3. **تحسين تمثيل الفئات النادرة**

غالبًا ما تُهمل في العينات التقليدية التصنيفات المختلفة للبيانات مثل تصنيف النشاءات ل"المنشآت المالية الكبيرة" و "منشآت النقل في المناطق الصناعية" ولا يتم أخذها في الإعتبار عند إختيار العينة، ولكن باستخدام الذكاء الاصطناعي يتم تحديدجميع التصنيفات ضمن الطبقات الذكية. مع مرعاة تمثيلها في العينة حسب تأثيرها الإحصائي.

4. **تقليل التحيزالناتج عن عدم الإستجابة**

عبر استخدام نماذج مثل GAIN والوزن العكسي لإحتمالية الإستجابة:(IPW) يمكن تصحيح أثر فقد البيانات الناتج عن عدم الإستجابة. وان النموذج التنبؤي يمكن من إستكمال البيانات المفقودة.

5. **زيادة دقة المؤشرات الإحصائية**

تمكن تقنيات الذكاء الاصطناعي من التحكم في متوسط مربع الخطأ. مما يحسن من عملية تقدير المتغيرات المطلوبة .

6. **توزيع الباحثين الميدانين بذكاء**

باستخدام خوارزميات التعليم التعزيزية، يتم توجيه الفرق الميدانية إلى المناطق التي تُظهر أفضل فرص الإستجابة، مما يودي إلى تقليل الزيارات المهدرة مع زيادة معدلات الإستجابة.

7. **مرونة التوسع في الدراسات المستقبلية**

يتم توثيق النموذج والبنية الذكية للعينة وجميع المراحل، مما يتيح تكراره مستقبلًا بنفس الطريقة. مع تعديل العينة بسهولة وتغيّر المجتمع المستهدف.

الخلاصة:

مع تطور أدوات التحليل وأساليب جمع البيانات، برز الذكاء الاصطناعي كأحد الحلول الواعدة في تعزيز كفاءة ودقة المعاينات الإحصائية. وقد أتاح توظيف الذكاء الاصطناعي في تصميم العينة وتنفيذها فرصًا جديدة لتحسين جودة البيانات، وتقليل التكاليف، والتغلب على العديد من التحديات المنهجية.

ويوضح الجدول رقم (11.3) أبرز الفوائد العملية والتطبيقية التي تم تحقيقها من خلال دمج الذكاء الاصطناعي في عمليات المعاينة، ويُبرز القيمة المضافة لهذا التوجه في خمسة مجالات رئيسية.

جدول رقم (11.3)

الفائدة الناتجة من استخدام الذكاء الاصطناعي في المعاينة الإحصائية

المجال	الفائدة الناتجة باستخدام الذكاء الاصطناعي
التكلفة والجهد	تقليل حجم العينة، توجيه الفرق الميدانية بكفاءة
الدقة	انخفاض واضح في متوسط مربع الخطأ وتحسين تمثيل المنشآت المؤثرة
العدالة في التمثيل	تمثيل أفضل للفئات النادرة والمناطق الطرفية
المرونة والتطوير	إمكانية تحسين النموذج مع كل دورة جديدة
التحكم في التحيزات	تصحيح آثار عدم الإستجابة وتقدير البيانات المفقودة بذكاء

كما إن إدماج تقنيات الذكاء الاصطناعي في عمليات إختيار العينات الإحصائية يُعد خطوة متقدمة نحو تحديث أدوات العمل الإحصائي وتحسين كفاءته. وتبرز أهمية هذه التقنيات في قدرتها على تحليل كميات ضخمة من البيانات بسرعة فائقة، والتعرّف على الأنماط والعلاقات المعقدة التي قد يصعب اكتشافها بالطرق التقليدية، مما يساعد على تصميم عينات أكثر تمثيلاً وفعالية.

وأن استخدام الذكاء الاصطناعي في المعاينة يساهم بشكل ملموس في تقليص الوقت والجهد المبذول في إعداد الخطط الميدانية، وتقليل التكاليف المرتبطة بجمع البيانات، خصوصًا في المسوح الواسعة أو ذات الطبيعة المتكررة. ويمكن للأنظمة الذكية أن تدعم اتخاذ قرارات سريعة ودقيقة في اختيار العينات أو تعديلها وفق ظروف التنفيذ الفعلية.

ورغم هذه الفوائد ، إلا أن توظيف الذكاء الاصطناعي في مجال المعاينة الإحصائية لا يزال في مراحله الأولية، ويواجه مجموعة من التحديات، من بينها الحاجة إلى بيانات تدريبية دقيقة، وتكامل الخوارزميات مع النماذج الإحصائية المعتمدة، وضمان الشفافية والموثوقية في المخرجات. لذلك، فإن هذه التقنية تتطلب جهودًا مستمرة في البحث والتطوير، وتحديثًا دوريًا للمفاهيم والأدوات، بما يضمن استخدامها بشكل سليم وفعّال في البيئات الإحصائية الرسمية.

المراجع

المراجع باللغة العربية

- الهيئة الاتحادية للتنافسية والإحصاء. (2018). (FCSC). *دليل معايير وضمان جودة البيانات الإدارية – معايير السرية والأمان*. FCSC: يوضح تصنيف البيانات وفق سريتها وإجراءات حمايتها ضمن إطار جودة البيانات الوطنية.
- الموقع الرسمي حكومة الإمارات العربية المتحدة، (2021) ،*القانون الاتحادي الإماراتي لحماية البيانات الشخصية رقم 45 لعام 2021*، تم الدخول على الموقع بتاريخ 2025/6/9، المتوفر على الرابط:

https://uaelegislation.gov.ae/ar/legislations/1972

- وزان، ميلاد، (2022) ، *كتاب التعلم العميق: المبادئ والمفاهيم والمناهج*، العراق، جامعة القادسية، المتوفر على الرابط:
https://drive.google.com/file/d/1sCzw-rDx0zr7Vv-vf12QM4yYSSNcNsbM/view
- الموقع الرسمي أكاديمية حسوب، (2020) ،*المفاهيم الأساسية لتعلم الآلة*، تم الدخول على الموقع بتاريخ 2025/6/9، المتوفر على الرابط:
https://academy.hsoub.com/programming/artificial-intelligence/
- الموقع الرسمي AWS Amazon، ما المقصود بتعلم الآلة، تم الدخول على الموقع بتاريخ 2025/6/9، المتوفر على الرابط:
https://aws.amazon.com/ar/what-is/machine-learning/
- الجمعية العامة للأمم المتحدة (2014)، *المبادئ الأساسية للإحصاءات الرسمية*، المتوفر على الرابط:
https://unstats.un.org/unsd/dnss/gp/FP-New-A.pdf
- الموقع الرسمي علوم24، *خوارزمية الغابة العشوائية*، تم الدخول على الرابط بتاريخ 2025/6/4 المتوفر على الرابط:
https://sciences24.com/random-forest/
- الموقع الرسمي Learn Statistics ، *ما هو عدم اليقين*، تم الدخول على الموقع بتاريخ 2025/7/4 ، المتوفر على الرابط:
https://ar.statisticseasily.com/

- الموقع الرسمي المبرمج العربي، *خوارزمية إبسيلون الجشع للحل الكلاسيكي في التعلم المعزز*، تم الدخول الموقع بتاريخ 2025/7/4، المتوفر على الرابط:

https://arabicprogrammer.com/article/8930878605/

المراجع الأجنبية

- Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons. Retrieved from: https://archive.org/details/samplingtechniqu0000coch_t4x6/page/n7/mode/2up
- Lohr, S. L. (2019). *Sampling: Design and Analysis* (2nd ed.). Chapman & Hall/CRC. Retrieved from: https://www.taylorfrancis.com/books/mono/sampling-sharon-lohr
- Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons. Retrieved from: https://books.google.ae/books/about/Complex_Surveys
- Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer. Retrieved from: https://link.springer.com/book/9780387406206
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Retrieved from: https://www.cs.ubc.ca/~murphyk/MLbook/pml-toc-1may12.pdf
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media. Retrieved from: https://books.google.ae/books/about/Hands_On_Machine_Learning_with_Scikit_Le
- Abadi, M. et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Researchgate. Retrieved from: https://www.researchgate.net/publication/301839500

- Chouldechova, A., & Roth, A. (2020). *A Snapshot of the Frontiers of Fairness in Machine Learning*. Communications of the ACM, 63(5), 82–89. Retrieved from: https://dl.acm.org/doi/10.1145/3376898
- Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. arXiv preprint arXiv:1312.6114. Retrieved from: https://arxiv.org/abs/1312.6114
- Yoon, J., Jordon, J., & van der Schaar, M. (2018). *GAIN: Missing Data Imputation using Generative Adversarial Nets*. arXiv preprint arXiv:1806.02920. Retrieved from: https://arxiv.org/abs/1806.02920
- OpenAI (2023). *GPT-4 Technical Report*. Retrieved from: https://openai.com/research