الوحدة الثانية: معالجة البيانات وتحليلهاData Processing and Analysis

الدرس الثاني: تقنيات تحليل البيانات(Data Analysis Techniques)

القسم الأول: تقنيات تحليل البيانات (الجزء الأول(

ما هي التقنية التي تُستخدم لتقسيم مجموعة كبيرة من البيانات إلى مجموعات (Clusters)بناءً على أوجه التشابه بين العناصر داخل كل مجموعة؟

أ. تحليل الانحدار.(Regression Analysis)

ب. تحليل المجموعات.(Clustering Analysis)

ج. تحليل التردد.(Frequency Analysis)

د. تحليل السلاسل الزمنية.(Time Series Analysis)

الإجابة الصحيحة: ب. تحليل المجموعات.(Clustering Analysis)

عندما يُستخدم تحليل المجموعات(Clustering Analysis) ، فإن الهدف الرئيسي هو:

أ. التنبؤ بقيم مستقبلية.

ب. تحديد العلاقات السببية بين المتغيرات.

ج. اكتشاف الأنماط والتصنيفات الطبيعية داخل البيانات غير المصنفة.

د. تلخيص البيانات الإحصائية فقط.

الإجابة الصحيحة: ج. اكتشاف الأنماط والتصنيفات الطبيعية داخل البيانات غير المصنفة.

أي تقنية تحليل بيانات تُستخدم لبناء نموذج يتنبأ بقيمة متغير تابع بناءً على قيمة متغير مستقل واحد أو أكثر؟

أ. تحليل الارتباط.(Correlation Analysis)

ب. تحليل الانحدار.(Regression Analysis)

ج. تحليل المجموعات.(Clustering Analysis)

د. تحليل التردد.(Frequency Analysis)

الإجابة الصحيحة: ب. تحليل الانحدار.(Regression Analysis)

إذا كنت تحاول التنبؤ بسعر منزل بناءً على مساحته وعدد غرفه، فما هي تقنية التحليل التي ستستخدمها على الأرجح؟

أ. تحليل المجموعات.

ب. تحليل الانحدار.

ج. تحليل التردد.

د. تحليل المشاعر.

الإجابة الصحيحة: ب. تحليل الانحدار.

أي من تقنيات تحليل البيانات تُستخدم لتحديد العلاقات والارتباطات بين متغيرين أو أكثر دون بالضرورة تحديد علاقة السبب والنتيجة؟

أ. تحليل المجموعات.(Clustering Analysis)

ب. تحليل الارتباط.(Correlation Analysis)

ج. تحليل الانحدار.(Regression Analysis)

د. تحليل التردد.(Frequency Analysis)

الإجابة الصحيحة: ب. تحليل الارتباط.(Correlation Analysis)

إذا وجدت أن هناك "ارتباطاً إيجابياً قوياً" بين متغيرين، فهذا يعني عادة أن:

أ. أحدهما يزداد بينما الآخر يقل.

ب. كلاهما يزداد أو يقل معاً.

ج. لا توجد علاقة بينهما.

د. أحدهما يسبب الآخر بالضرورة.

الإجابة الصحيحة: ب. كلاهما يزداد أو يقل معاً.

ما هي التقنية التي تُستخدم لحساب عدد مرات تكرار كل قيمة فريدة في مجموعة بيانات معينة؟

أ. تحليل الانحدار.(Regression Analysis)

ب. تحليل المجموعات.(Clustering Analysis)

ج. تحليل التردد.(Frequency Analysis)

د. تحليل السلاسل الزمنية.(Time Series Analysis)

الإجابة الصحيحة: ج. تحليل التردد.(Frequency Analysis)

عند دراسة اتجاهات أسعار الأسهم أو درجات الحرارة على مدى فترة زمنية، ما هي تقنية التحليل الأنسب للاستخدام؟

أ. تحليل المجموعات.(Clustering Analysis)

ب. تحليل التردد.(Frequency Analysis)

ج. تحليل السلاسل الزمنية.(Time Series Analysis)

د. تحليل المشاعر.(Sentiment Analysis)

الإجابة الصحيحة: ج. تحليل السلاسل الزمنية.(Time Series Analysis)

ما هي التقنية التي تُستخدم لاستخراج الأنماط أو القواعد من مجموعات البيانات الكبيرة، مثل "إذا اشترى العميل المنتجA ، فمن المرجح أن يشتري المنتج"B ؟

أ. تحليل المشاعر.(Sentiment Analysis)

ب. تحليل سلة التسوق.(Market Basket Analysis)

ج. تحليل الانحدار.(Regression Analysis)

د. تحليل الارتباط.(Correlation Analysis)

الإجابة الصحيحة: ب. تحليل سلة التسوق.(Market Basket Analysis)

تُستخدم تقنية "تحليل المشاعر (Sentiment Analysis) "بشكل أساسي لفهم:

أ. العلاقة بين متغيرين رقميين.

ب. عدد مرات تكرار كلمة معينة في نص.

ج. النبرة العاطفية (إيجابية، سلبية، محايدة) في البيانات النصية.

د. مجموعات البيانات المتشابهة.

الإجابة الصحيحة: ج. النبرة العاطفية (إيجابية، سلبية، محايدة) في البيانات النصية.

الوحدة الثانية: معالجة البيانات وتحليلهاData Processing and Analysis

الدرس الثاني: تقنيات تحليل البيانات(Data Analysis Techniques)

القسم الثاني: أدوات ومصطلحات تحليل البيانات (Data Analysis Tools & Terminology)

أي من البرامج التالية يُعد من أدوات تحليل البيانات الشائعة والممتازة للتعامل مع الجداول الكبيرة وإجراء التحليلات الإحصائية البسيطة؟

أ. برنامج الرسام.(Paint)

ب. مايكروسوفت وورد.(Microsoft Word)

ج. مايكروسوفت إكسل.(Microsoft Excel)

د. متصفح الويب.(Web Browser)

الإجابة الصحيحة: ج. مايكروسوفت إكسل.(Microsoft Excel)

ما هي إحدى الميزات الرئيسية لأدوات مثل Tableau و Power BI في تحليل البيانات؟

أ. القدرة على كتابة أكواد برمجية معقدة فقط.

ب. توفير واجهات رسومية قوية لإنشاء لوحات معلومات (Dashboards) وتصور البيانات.

ج. معالجة النصوص فقط.

د. استخدامها حصرياً للمستخدمين ذوي الخبرة في البرمجة.

الإجابة الصحيحة: ب. توفير واجهات رسومية قوية لإنشاء لوحات معلومات (Dashboards)وتصور البيانات.

ما هي اللغة البرمجية التي تُعد من أقوى اللغات وأكثرها شيوعاً في تحليل البيانات والتعلم الآلي بسبب مكتباتها الغنية؟

أ.HTML .

ب.CSS .

ج. بايثون.(Python)

د.Java .

الإجابة الصحيحة: ج. بايثون.(Python)

ما هي المكتبة في بايثون التي تُستخدم بشكل واسع للعمليات العددية ومعالجة المصفوفات الكبيرة، وهي أساس للعديد من مكتبات تحليل البيانات الأخرى؟

أ.Matplotlib .

ب.Pandas .

ج.NumPy .

د.Seaborn .

الإجابة الصحيحة: ج.NumPy .

أي مكتبة في بايثون تُعتبر حجر الزاوية في تحليل البيانات وتوفر هياكل بيانات مرنة مثل DataFramesللتعامل مع البيانات الجدولية؟

أ.SciPy .

ب.TensorFlow .

ج.Keras .

د.Pandas .

الإجابة الصحيحة: د.Pandas .

في سياق تحليل البيانات، يشير مصطلح "المتغير (Variable) "إلى:

أ. قيمة ثابتة لا تتغير.

ب. خاصية أو سمة يتم قياسها أو ملاحظتها في مجموعة بيانات.

ج. أداة برمجية.

د. نوع من الرسوم البيانية.

الإجابة الصحيحة: ب. خاصية أو سمة يتم قياسها أو ملاحظتها في مجموعة بيانات.

ما هو المصطلح الذي يصف مجموعة البيانات التي تم جمعها من أجل تحليل معين؟

أ. المصدر.(Source)

ب. العينة.(Sample)

ج. مجموعة البيانات.(Dataset)

د. النتيجة.(Outcome)

الإجابة الصحيحة: ج. مجموعة البيانات.(Dataset)

يشير مصطلح "القيمة الشاذة" أو "القيمة المتطرفة (Outlier) "في تحليل البيانات إلى:

أ. قيمة تحدث بشكل متكرر جداً.

ب. قيمة بعيدة بشكل غير عادي عن معظم البيانات الأخرى في مجموعة البيانات.

ج. قيمة مفقودة في البيانات.

د. قيمة متوسطة.

الإجابة الصحيحة: ب. قيمة بعيدة بشكل غير عادي عن معظم البيانات الأخرى في مجموعة البيانات.

ما هو "المتوسط (Mean) "في الإحصاء وتحليل البيانات؟

أ. القيمة الأكثر تكراراً في مجموعة البيانات.

ب. القيمة التي تقسم البيانات إلى نصفين متساويين.

ج. مجموع كل القيم مقسوماً على عددها الإجمالي.

د. الفرق بين أعلى وأدنى قيمة.

الإجابة الصحيحة: ج. مجموع كل القيم مقسوماً على عددها الإجمالي.

تُعرف "الموثوقية (Reliability) "في تحليل البيانات بأنها:

أ. مدى قدرة الأداة على قياس ما يفترض أن تقيسه.

ب. مدى اتساق النتائج عند إعادة قياس نفس الشيء عدة مرات.

ج. سرعة جمع البيانات.

د. سهولة استخدام الأداة.

الإجابة الصحيحة: ب. مدى اتساق النتائج عند إعادة قياس نفس الشيء عدة مرات.

Prepared by:موقع المناهج الدراسية الأردن

الوحدة الثانية: معالجة البيانات وتحليلهاData Processing and Analysis

الدرس الثاني: تقنيات تحليل البيانات(Data Analysis Techniques)

القسم الثالث: عرض البيانات باستخدام الرسوم البيانية(Data Visualization)

أي من الرسوم البيانية التالية يُعد الأنسب لعرض التغيرات والاتجاهات في البيانات على مدى فترة زمنية متصلة؟

أ. الرسم البياني بالأعمدة.(Bar Chart)

ب. الشكل الدائري.(Pie Chart)

ج. الرسم البياني الخطي.(Line Chart)

د. الرسم التوضيحي.(Scatter Plot)

الإجابة الصحيحة: ج. الرسم البياني الخطي.(Line Chart)

يُستخدم الرسم البياني الخطي (Line Chart) بشكل فعال لعرض:

أ. مقارنة بين فئات مختلفة في نقطة زمنية واحدة.

ب. نسبة جزء من الكل.

ج. تطور قيمة معينة (مثل درجة الحرارة أو المبيعات) عبر الزمن.

د. العلاقة بين متغيرين رقميين.

الإجابة الصحيحة: ج. تطور قيمة معينة (مثل درجة الحرارة أو المبيعات) عبر الزمن.

أي من الرسوم البيانية التالية يُفضل استخدامه للمقارنة بين قيم فئات مختلفة أو مجموعات متباينة؟

أ. الرسم البياني الخطي.(Line Chart)

ب. الشكل الدائري.(Pie Chart)

ج. الرسم البياني بالأعمدة.(Bar Chart)

د. الرسم التوضيحي.(Scatter Plot)

الإجابة الصحيحة: ج. الرسم البياني بالأعمدة.(Bar Chart)

إذا أردت عرض عدد الطلاب في كل تخصص جامعي (مثال: علوم الحاسوب، الهندسة، الطب)، فما هو الرسم البياني الأكثر ملاءمة؟

أ. الرسم البياني الخطي.(Line Chart)

ب. الرسم البياني بالأعمدة.(Bar Chart)

ج. الشكل الدائري.(Pie Chart)

د. الرسم التوضيحي.(Scatter Plot)

الإجابة الصحيحة: ب. الرسم البياني بالأعمدة.(Bar Chart)

ما هو نوع الرسم البياني الذي يُستخدم لعرض نسبة كل جزء إلى الكل (مجموع الأجزاء 100%)؟

أ. الرسم البياني بالأعمدة.(Bar Chart)

ب. الرسم البياني الخطي.(Line Chart)

ج. الشكل الدائري.(Pie Chart)

د. الرسم التوضيحي.(Scatter Plot)

الإجابة الصحيحة: ج. الشكل الدائري.(Pie Chart)

يُعد الشكل الدائري (Pie Chart) الأنسب عندما يكون عدد الفئات التي يتم تمثيلها:

أ. كبيراً جداً (أكثر من 10 فئات.(

ب. قليلاً (3-5 فئات) لسهولة المقارنة البصرية.

ج. غير محدد.

د. متغير باستمرار.

الإجابة الصحيحة: ب. قليلاً (3-5 فئات) لسهولة المقارنة البصرية.

أي رسم بياني يُستخدم لتحديد العلاقة أو الارتباط بين متغيرين رقميين، حيث يُمثل كل نقطة زوجاً من القيم؟

أ. الرسم البياني الخطي.(Line Chart)

ب. الشكل الدائري.(Pie Chart)

ج. الرسم البياني بالأعمدة.(Bar Chart)

د. الرسم التوضيحي أو مخطط التشتت.(Scatter Plot)

الإجابة الصحيحة: د. الرسم التوضيحي أو مخطط التشتت.(Scatter Plot)

إذا لاحظت في الرسم التوضيحي (Scatter Plot) أن النقاط تتجمع حول خط مستقيم يميل للأعلى من اليسار إلى اليمين، فهذا يشير إلى:

أ. عدم وجود علاقة بين المتغيرين.

ب. علاقة سلبية قوية (أحدهما يزداد والآخر يقل.(

ج. علاقة إيجابية قوية (كلاهما يزداد أو يقل معاً.(

د. علاقة غير خطية.

الإجابة الصحيحة: ج. علاقة إيجابية قوية (كلاهما يزداد أو يقل معاً.(

ما هو العنصر الذي يجب أن يكون موجوداً في أي رسم بياني فعال لتوضيح ما يمثله المحور الأفقي والمحور الرأسي؟

أ. الألوان الزاهية فقط.

ب. تسميات المحاور.(Axis Labels)

ج. حجم الخط الكبير.

د. الخلفية الشفافة.

الإجابة الصحيحة: ب. تسميات المحاور.(Axis Labels)

أي من الرسوم البيانية يسمح لك برؤية إذا كانت هناك أي "قيم شاذة (Outliers) "بشكل واضح بين متغيرين رقميين؟

أ. الشكل الدائري.(Pie Chart)

ب. الرسم البياني بالأعمدة.(Bar Chart)

ج. الرسم التوضيحي.(Scatter Plot)

د. الرسم البياني الخطي.(Line Chart)

الإجابة الصحيحة: ج. الرسم التوضيحي.(Scatter Plot)

الوحدة الثانية: معالجة البيانات وتحليلهاData Processing and Analysis

الدرس الثاني: تقنيات تحليل البيانات(Data Analysis Techniques)

القسم الخامس: أسئلة إضافية حول تقنيات تحليل البيانات

عند استخدام تحليل الانحدار(Regression Analysis) ، ما هو الهدف الأساسي من تحديد معادلة الانحدار؟

أ. تصنيف البيانات إلى مجموعات.

ب. قياس تكرار ظهور قيمة معينة.

ج. التنبؤ بقيمة متغير تابع بناءً على متغير مستقل أو أكثر.

د. اكتشاف الأنماط الموسمية.

الإجابة الصحيحة: ج. التنبؤ بقيمة متغير تابع بناءً على متغير مستقل أو أكثر.

أي من تقنيات التحليل التالية هي الأنسب لتحديد شرائح العملاء المختلفة بناءً على عاداتهم الشرائية وسلوكهم؟

أ. تحليل السلاسل الزمنية.(Time Series Analysis)

ب. تحليل الانحدار.(Regression Analysis)

ج. تحليل المجموعات.(Clustering Analysis)

د. تحليل الارتباط.(Correlation Analysis)

الإجابة الصحيحة: ج. تحليل المجموعات.(Clustering Analysis)

يُعرف "الارتباط (Correlation) "بأنه:

أ. أن متغير واحد يسبب الآخر بالضرورة.

ب. مقياس للعلاقة الإحصائية بين متغيرين، والتي قد تكون إيجابية أو سلبية أو لا توجد علاقة.

ج. توقع لقيمة مستقبلية.

د. تقسيم البيانات إلى فئات محددة مسبقاً.

الإجابة الصحيحة: ب. مقياس للعلاقة الإحصائية بين متغيرين، والتي قد تكون إيجابية أو سلبية أو لا توجد علاقة.

إذا كان معامل الارتباط (Correlation Coefficient) بين متغيرين قريباً من 0، فهذا يشير إلى:

أ. علاقة إيجابية قوية.

ب. علاقة سلبية قوية.

ج. عدم وجود علاقة خطية واضحة بين المتغيرين.

د. علاقة سببية مباشرة.

الإجابة الصحيحة: ج. عدم وجود علاقة خطية واضحة بين المتغيرين.

أي من التقنيات التالية تُستخدم بشكل شائع في أنظمة التوصية (Recommendation Systems)مثل تلك الموجودة في مواقع التسوق أو بث الأفلام؟

أ. تحليل التردد.

ب. تحليل المشاعر.

ج. تحليل المجموعات (Clustering) أو تحليل سلة التسوق (Market Basket Analysis).

د. تحليل السلاسل الزمنية.

الإجابة الصحيحة: ج. تحليل المجموعات (Clustering) أو تحليل سلة التسوق (Market Basket Analysis).

ما هي التقنية الأنسب لتحليل بيانات المبيعات الشهرية لمنتج معين على مدار خمس سنوات؟

أ. تحليل الارتباط.

ب. تحليل السلاسل الزمنية.

ج. تحليل المجموعات.

د. تحليل المشاعر.

الإجابة الصحيحة: ب. تحليل السلاسل الزمنية.

عند استخدام تحليل التردد(Frequency Analysis) ، ما هو المقياس الذي يتم التركيز عليه بشكل أساسي؟

أ. متوسط القيم.

ب. تشتت القيم.

ج. عدد مرات ظهور كل قيمة أو فئة.

د. العلاقة السببية.

الإجابة الصحيحة: ج. عدد مرات ظهور كل قيمة أو فئة.

إذا كنت ترغب في معرفة ما إذا كانت حملة تسويقية جديدة قد أدت إلى تغيير في "مشاعر" العملاء تجاه منتج ما، فما هي التقنية الأكثر ملاءمة؟

أ. تحليل الانحدار.

ب. تحليل سلة التسوق.

ج. تحليل المشاعر.

د. تحليل التردد.

الإجابة الصحيحة: ج. تحليل المشاعر.

يُعتبر تحليل المجموعات (Clustering) من طرق "التعلم غير الخاضع للإشراف " (Unsupervised Learning)لأنه:

أ. يعتمد على بيانات مصنفة مسبقاً.

ب. لا يتطلب تسميات مسبقة للبيانات، بل يكتشف الأنماط بنفسه.

ج. يحتاج إلى إشراف بشري مستمر.

د. يقوم بالتنبؤ بقيم محددة.

الإجابة الصحيحة: ب. لا يتطلب تسميات مسبقة للبيانات، بل يكتشف الأنماط بنفسه.

أي من التقنيات التالية قد تكون مفيدة في الكشف عن الاحتيال المالي من خلال تحديد المعاملات غير العادية التي تختلف بشكل كبير عن الأنماط العادية؟

أ. تحليل السلاسل الزمنية.

ب. تحليل المجموعات (كشف القيم الشاذة.(Outlier Detection/

ج. تحليل التردد.

د. تحليل المشاعر.

الإجابة الصحيحة: ب. تحليل المجموعات (كشف القيم الشاذة.(Outlier Detection/