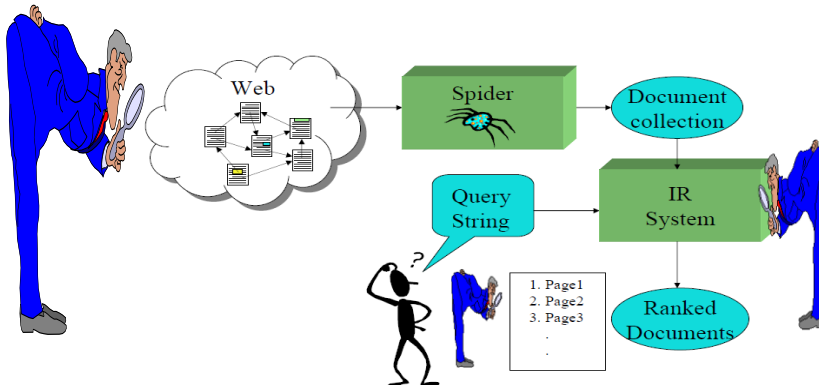


مقرر تنظيم المعلومات الإدارية Management Information Systems MIS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
مَنْ قَرَأَ هَذَا كِتَابًا
مِنْ كِتَابِي فِي يَوْمٍ
مِنْ أَيَّامِي حَسِبْتَنِي
مِنْ مَنزِلِي

الفصل الثامن

التقيب في البيانات واكتشاف المعرفة



مقدمة

• لم تقتصر حاجة المديرين في الحصول على معلومات ملخصة داخلية وخارجية من بيانات قواعد أو مستودع البيانات (DW) أو منفذ البيانات (DM) Data Mart، بل امتدت حاجتهم للحصول على معرفة مكتشفة من ذلك الكم الكبير من البيانات، حيث يتم الحصول على تلك المعرفة من خلال رفع مستوى معالجة بيانات مستودع البيانات (أو منفذ البيانات) إلى المعالجة الفائقة التي تستخدم نماذج وخوارزميات التنقيب في البيانات Data Mining (DM).

• ولقد أصبح تطبيق أساليب التنقيب في البيانات ممكناً ومجدياً مع التطور الكبير في قدرة الحاسوب على معالجة كمية كبيرة من البيانات وبسرعة مذهلة، وبدأ ذلك في منتصف التسعينات لتبدأ تطبيقات التنقيب في البيانات واستمر ذلك التطور ليستمر التوسع في تطبيقات التنقيب في البيانات ولاسيما تطبيقات علوم الإحصاء متعددة المتغيرات وتطبيقات تعلم الآلة التي تحتاج إلى معالجات هائلة،

• **فما هو مفهوم التنقيب في البيانات؟**

• **وما هي علاقته بمستودع البيانات ؟**

• **وما هي أهم أساليبه؟.**

3

مقدمة

• مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات ومخازن البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، من هنا ظهر ما يسمى بالتنقيب في البيانات كتقنية تهدف إلى استخراج المعرفة من كميات هائلة من البيانات. وهي تقنية حديثة فرضت نفسها بقوة في عصر المعلوماتية، واستخدامها يوفر للشركات والمنظمات في جميع المجالات القدرة على استكشاف والتركيز على أهم المعلومات في قواعد البيانات، كما تركز تقنيات التنقيب على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات مما يسمح باتخاذ القرارات الصحيحة في الوقت المناسب. والتي تعتبر بدورها مرحلة من مراحل عملية أكثر تعقيداً هي استكشاف المعرفة في قواعد البيانات، والمرتبطة إلى حد بعيد بعملية تطوير أخرى مهمة جداً هي مستودعات البيانات. حيث أن الكثير من الشركات والمنظمات الرائدة اليوم تستخدم عملية استكشاف المعرفة في قواعد البيانات بشكل منهجي ومنظم بوصفها تشكل جوهر العمل الذي يعتمد عليه في تفعيل النشاط وتحقيق الميزة التنافسية.

4

ثانيا: العوامل المحركة لثورة التنقيب في البيانات

- يمكن تقسيم العوامل المحركة لثورة التنقيب في البيانات إلى قسمين رئيسيين هما:
- **أ. العوامل المتعلقة بجانب العرض:**
 - تأثيرات تطورات تكنولوجيا المعلومات خاصة تكنولوجيا تخزين البيانات ومعالجتها.
 - انخفاض تكاليف الاتصالات الإلكترونية مما يسر الوصول إلى قواعد البيانات.
 - ظهور أساليب تحليل جديدة من أبرزها الشبكات العصبية، نظم الخوارزميات الوراثية، شجرة القرار وقواعد الاستقراء.
 - ثورة إعمار الحاسوب والتي تمخضت عن ولادة أجيال جديدة من البرمجيات الصديقة للمستخدم، مثل: مايكروسوفت ويندوز، وبرمجيات العميل-الخادم (Client-Server) وغيرها.
 - التطور الهائل في أساليب التخزين ومن أبرز هذه الأساليب تلك المتعلقة بمستودعات البيانات (Data Warehouse) وأسواقها.

5

ثانيا: العوامل المحركة لثورة التنقيب في البيانات

- **ب. العوامل المتعلقة بجانب الطلب:**
 - تنامي الحاجة إلى نتائج تحليلية سريعة من طرف المؤسسات المنافسة الشديدة والتغيرات السريعة في بيئة العمل.
 - انحسار التنظيم الهرمي حيث أصبح لزاما على المديرين الاعتماد على أنفسهم في الحصول على احتياجاتهم من المعلومات التحليلية، بعدما انحسر دور محلي الدعم الذين كانوا يعتمدون عليهم.
 - ازدهار حركة التأليف والنشر خاصة في مجال تكنولوجيا المعلومات والإنترنت وكل ما يرتبط بعصر الاقتصاد الرقمي.

6

مفهوم التنقيب في البيانات (Data Mining)

- ظهر مصطلح التنقيب في البيانات في منتصف التسعينات في الولايات المتحدة الأمريكية، وهو يجمع ما بين الإحصاء وتكنولوجيات الإعلام (قواعد البيانات، الذكاء الاصطناعي، التعلم الآلي «machine Learning»
- و توجد عدة تعريفات لهذا المفهوم منها، حيث يمكن تعريفها بأنها: " الاستكشاف الآلي أو المؤتمت لأنماط شائقة وغير جلية مخفية في قاعدة بيانات معينة"، أو أنها: "سيرورة تحليل دقيقة وذكية، تفاعلية وتسلسلية، تسمح لمسيرتي النشاطات عند استخدام هذه السيرورة باتخاذ قرارات والقيام بأعمال ملائمة في صالح النشاط للمسؤولين عنه والمؤسسة التي يعملون بها"، أو أنها: " عبارة عن تحليلات لكمية كبيرة من البيانات بغرض إيجاد قواعد وأمثلة ونماذج التي يمكن أن تستخدم وتفقد وتدلل أصحاب القرار، وتتنبأ بالسلوك المستقبلي"، كما يمكن تعريفها كذلك بأنها: " تحليل لمجموعات كبيرة الحجم من البيانات المشاهدة للبحث عن علاقات محتملة وتلخيص للبيانات في أشكال جديدة لتكون مفهومة ومفيدة لمستخدمها".

7

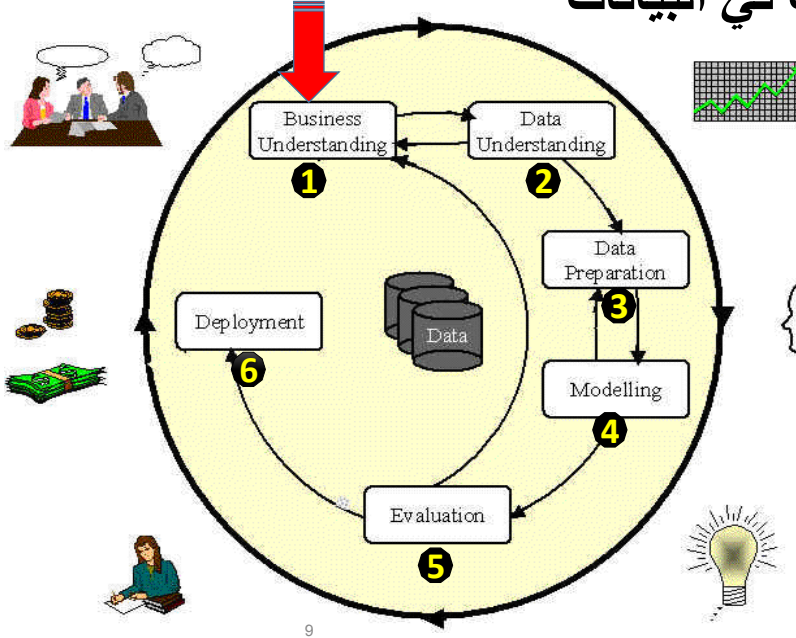
مفهوم التنقيب في البيانات

- يمكن تعريف التنقيب في البيانات DM بأنه: استطلاع وتحليل كميات كبيرة من البيانات لاكتشاف علاقات وقواعد ذات معنى بين المتغيرات، وهو بذلك يساعد في الإجابة على أسئلة من الشكل:
- لماذا حدث؟
- وكيف حدث؟
- وماذا سيحدث أو سيكون؟

أي أنه يساعد في اكتشاف معارف جديدة.

8

مراحل عملية التنقيب في البيانات



يمكن تلخيص مراحل خطوات عملية التنقيب في البيانات كما يلي:

1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

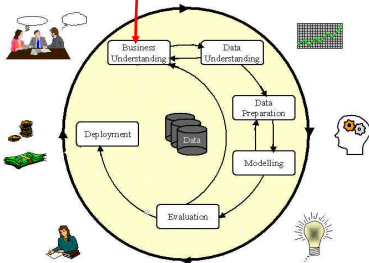
1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• يمكن تلخيص مراحل خطوات عملية التنقيب في البيانات كما يلي:

• فهم طبيعة الأعمال (Business Understanding):

- يعتبر المطلب الأول لاكتشاف المعرفة هو فهم المشاكل والمسائل التي تواجهها الأعمال. وبمعنى آخر، كيف يمكن تحقيق المنفعة الأعظم من التنقيب في البيانات، مما يتطلب وجود صيغة واضحة ومحددة لأهداف الأعمال.



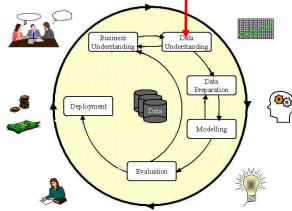
1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• فهم البيانات (Data Understanding):

• تعتبر مسألة معرفة ماهية وطبيعة البيانات عامل مهم في نجاح عملية التنقيب في البيانات واكتشاف المعرفة. حيث أن معرفة البيانات بصورة جيدة تعني مساعدة المصممين على استخدام الخوارزميات أو الأدوات المستخدمة للمسائل المحددة بدقة عالية. وهذا يقود إلى تعظيم فرص النجاح بالإضافة إلى رفع الفعالية والكفاءة لنظام اكتشاف المعرفة.

• ولا تحتاج عملية التنقيب في البيانات إلى تجميع البيانات في مستودع البيانات، أما إذا كان مستودع البيانات موجود في المنظمة، فمن الأفضل عدم احتكار المستودع بشكل مباشر لغرض التنقيب في البيانات.



11

1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• ويمكن تلخيص الخطوات الضرورية لعملية فهم البيانات كالآتي:

1. **تجميع البيانات (Data Collection):** وهي الخطوة الموجهة نحو تحديد مصدر البيانات في الدراسة بما في ذلك استخدام البيانات العامة الخارجية مثل الضرائب وغيرها.
2. **توصيف البيانات (Data Description):** وهي الخطوة التي تركز على توصيف محتويات الملف الواحد من الملفات أو الجداول.
3. **جودة البيانات وتحققها (Data Quality and Verification):** هذه الخطوة تحدد ما إذا كان تقليل أو إهمال بعض البيانات غير الضرورية أو كونها رديئة الجودة وقد لا تنفع في الدراسة. لأن النموذج الجيد يحتاج إلى بيانات جيدة مما يتوجب أن تكون البيانات صحيحة وذات مضمون دقيق.
4. **التحليل الاسترشادي للبيانات (Exploratory Analysis of Data):** تستخدم الأساليب مثل الإظهار المرئي أو التصور أو عملية التحليل المباشر (OLAP) التي تؤدي إلى إجراء التحليل الأولي للبيانات وتعتبر هذه الخطوة مهمة وضرورية لأنها تركز على تطوير الفرضيات المتعلقة بالمشكلة قيد الدراسة.



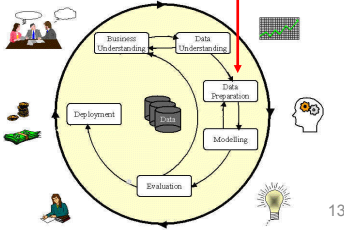
12

1. فهم طبيعة الأعمال
2. فهم البيانات
3. **تهيئة البيانات**
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• **تهيئة البيانات (Data Preparation):** وتشمل الخطوات التالية:

1. الاختيار (Selection) وتعني اختيار المتغيرات المتوقعة وحجم العينة.
2. صياغة المتغيرات وتحويلها (Construction and Transformation Variables) حيث يجب دائماً أن تصاغ المتغيرات الجديدة لبناء النماذج الفعالة.
3. تكامل البيانات (Data Integration): حيث أن مجاميع البيانات في دراسة التنقيب عن البيانات من الممكن تخزينها في قواعد بيانات متعددة الأغراض التي تكون بحاجة إلى توحيدها في قاعدة بياناتية واحدة.
4. تصميم وتنسيق البيانات (Data Formatting) حيث تتعلق هذه الخطوة في إعادة ترتيب حقول البيانات كما يتطلب في نموذج التنقيب في البيانات.



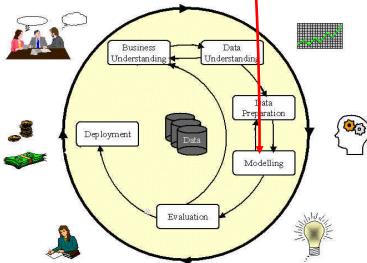
13

1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. **صياغة نماذج الحل وثبوتها**
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• **صياغة نماذج الحل وثبوتها (Model Building and Validation):** إن بناء وصياغة

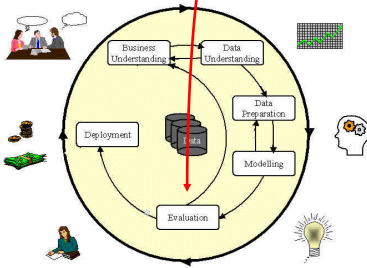
نموذج الحل السليم والدقيق يتم من خلال عملية الخطأ والصواب، حيث كثيراً ما تحتاج مثل هذه العملية إلى مساعدة المختصين في التنقيب عن البيانات بهدف اختبار وفحص مختلف البدائل للحصول على أفضل نموذج لحل المشكلة قيد الدراسة.



1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

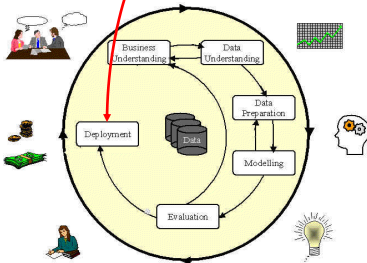
• التقييم وتعليل نتائج النموذج (Evaluation and Interpretation): حالما يتم صياغة النموذج والتحقق من ثباته وصدقه، تجري مباشرة عملية التحقق من ثبات حزمة البيانات التي يتم تغذيتها بواسطة النموذج. وبما أن نتائج هذه البيانات معروفة، لذا فإن النتائج المتوقعة تقارن مع النتائج الفعلية في ثبات حزمة البيانات قيد التشغيل. وتؤدي هذه المقارنة أو المفاضلة إلى التحقق من دقة النموذج.



1. فهم طبيعة الأعمال
2. فهم البيانات
3. تهيئة البيانات
4. صياغة نماذج الحل وثبوتها
5. التقييم وتعليل نتائج النموذج
6. نشر وتوزيع النموذج

مراحل عملية التنقيب في البيانات

• نشر وتوزيع النموذج (Model Deployment): حيث تشتمل هذه الخطوة على نشر وتوزيع النموذج داخل المنظمة لمساعدة عملية صنع القرار. وأن النموذج الصالح يجب أيضا أن يحقق الرضا لدى المستخدمين طالما أن اختيار النموذج لا بد أن يتم من خلال الدراسة الاسترشادية أو نموذج مصغر من الدراسة الشاملة.



أساليب التنقيب في البيانات

• يمكن التنقيب في البيانات من اكتشاف معارف جديدة بواسطة أساليب علم الإحصاء، وأساليب تعلم الآلة (مثل أسلوب الشبكات العصبية)، وفيما يلي موجز عن أهم أساليب التنقيب في البيانات:

1. الجار الأقرب
2. التجزئة العنقودية
3. شجرة القرار
4. الشبكات العصبية (تعلم الآلة) هنا تعطي أوزان ترجيحية
5. استقراء القاعدة
6. تحليل الانحدار والسلاسل الزمنية
7. اختبار الفروض: أسلوب إحصائي يقوم على اختبارات معلمية ولا معلمية

17

أساليب التنقيب في البيانات

• أسلوب الجار الأقرب

- تعتبر من تقنيات التنقيب في البيانات التي تهدف للتنبؤ عن طريق مقارنة السجلات الشبيهة بالسجل المراد التنبؤ له وتقدير القيمة المجهولة لهذا السجل بناء على معلومات تلك السجلات المتوفرة.
- كثيراً ما تستخدم خوارزمية الجار الأقرب في مجال الأعمال، ومن الأمثلة الشائعة الاستخدام تلك التي تساعد المستخدمين في الشراء عن طريق اختيار السلع الأقرب لاحتياجاتهم مقارنة بسلع قد تم شراؤها بالفعل.

• أسلوب التحليل بالتجزئة العنقودية

- هي عملية تجميع السجلات المتشابهة في مجموعات، ويتم ذلك بهدف الاستكشاف عالي المستوى لما يجري داخل قاعدة البيانات.

18

أساليب التنقيب في البيانات

- **أساليب تحليل المجموعة المتجانسة:**
- **تهدف لتصنيف كمية كبيرة من بيانات الأشياء أو الأشخاص أو الحالات في مجموعات معينة أقل (يحدد المستخدم عددها)، بحيث يتم تعظيم التجانس داخل كل مجموعة، وتعظيم التباين بين المجموعات المختلفة.**
- **فمثلاً يمكن: تصنيف الموردين و/أو العملاء و/أو المنتجات و/أو مناطق البيع إلى مجموعات متشابهة حسب خاصية معينة، كأن يتم:**
 - تصنيف العملاء إلى مجموعات حسب نسبة ربحيتهم أو حسب نسبة مشترياتهم،
 - تصنيف الموردين حسب نسبة التسليم في الموعد المحدد،
 - تصنيف المنتجات إلى عائلات حسب ربحيتها و/أو التشابه في خصائص الطلب عليها.
- **ولا شك أن معرفة مواصفات المجموعات المختلفة تجعل الإدارة تختار أسلوب التعامل المناسب لكل مجموعة، كأن تقوم بتحسين علاقتها بالعملاء الأكثر ربحية والموردين الأكثر التزاماً بمواعيد التسليم، وتحدد مزيجاً ترويجياً مناسباً لكل عائلة منتج ولكل شريحة من العملاء في كل منطقة جغرافية معينة.**

19

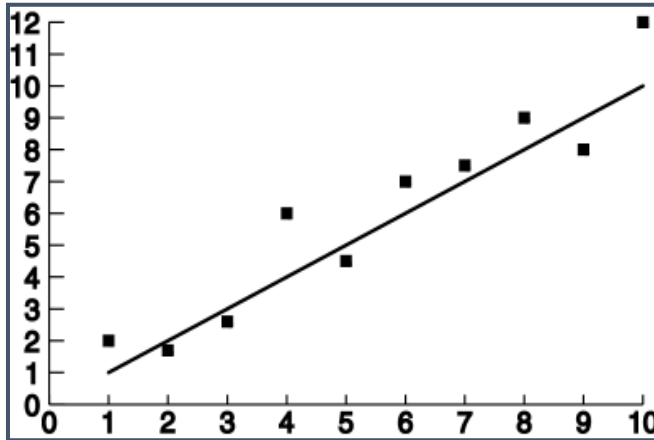
أساليب التنقيب في البيانات

- **أسلوب شجرة القرار:**
- يقوم أسلوب شجرة القرار Decision Tree بتجزئة كمية كبيرة من بيانات مجموعة من الأشياء أو الأشخاص أو الحالات إلى مجموعات رئيسية ومجموعات متفرعة منها، وتجزئة المجموعات الفرعية إلى مجموعات أصغر وهكذا لتأخذ شكل الشجرة وأغصانها، **ويستخدم هذا الأسلوب بهدف التصنيف أو التنبؤ.**
- وشجرة القرار هي نموذج استكشافي يظهر على شكل شجرة، كما يعبر اسمها، وبشكل دقيق يمثل كل فرع من فروعها سؤالاً تصنيفياً وتمثل أوراقها أجزاء من قاعدة البيانات تنتمي للتصنيفات التي تم بنائها.
- **أسلوب الشبكات العصبية:**
- خوارزمية الشبكة العصبية تشبه في تركيبها تركيبية مخ الإنسان، يقوم أسلوب الشبكة العصبية على أساس التعلم من بيانات المتغيرات الفعلية السابقة.

20

أساليب التنقيب في البيانات

Linear Regression



21

• أساليب تحليل الانحدار والسلاسل الزمنية:

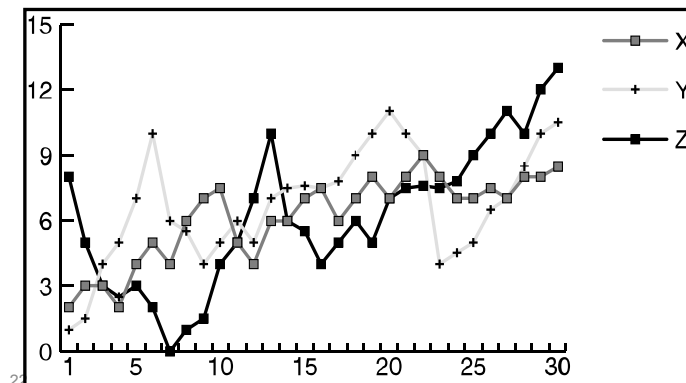
• تساعد أساليب تحليل الانحدار Regression Analysis متخذ القرار في التنبؤ بقيمة المتغير التابع من خلال تقدير بيانات المتغيرات المستقلة المؤثرة به بعد أن يتم تحديد نموذج الانحدار من واقع بيانات تاريخية.

أساليب التنقيب في البيانات

• أساليب تحليل الانحدار والسلاسل الزمنية:

• يقوم أسلوب السلاسل الزمنية بالتنبؤ بقيمة متغير تابع في أي فترة مستقبلية بعد أن يتم تصميم نموذج السلسلة الزمنية من بيانات المتغير التابع (مثل المبيعات) والمتغير المستقل (الزمن بالشهور أو السنين، الخ).

- [Example:](#)
- Stock Market
- Predict future values
- Determine similar patterns over time
- Classify behavior



22

أساليب التنقيب في البيانات

- أساليب اختبار الفروض:
- تقوم أساليب اختبار الفروض باختبار صحة فروض معينة إحصائياً، فمثلاً يمكن **اختبار الفرض القائل بأن: لا توجد علاقة بين مكان السكن واستخدام الموبايل**، وتكون نتيجة الاختبار إما الرفض أو القبول (إحصائياً) عند درجة ثقة معينة (95% عادة) ، وعند قبول الفرض يتم إثبات معرفة جديدة، وتنقسم أساليب اختبار الفروض إلى قسمين:
- الاختبارات المعلمية **Parametric Tests**
 - من أمثلة الاختبارات المعلمية: اختبار T (لمقارنة متوسطين)، واختبار F (لمقارنة أكثر من متوسط، كما في فرض استهلاك السمن السابق)، الخ،
- والاختبارات البديلة غير المعلمية **Nonparametric Tests** (تستخدم عند عدم توافر شروط الاختبارات المعلمية).
 - ومن أمثلة الاختبارات غير المعلمية: اختبار Mann-Whitney (وهو بديل لاختبار T في مجموعتين مستقلتين)، واختبار الإشارة **Sign Test** لعينة واحدة (البديل لاختبار T في عينة واحدة)، واختبار **ANOVA** **Kruskal-Wallis** (البديل لتحليل التباين).

23

أساليب التنقيب في البيانات

- التمثيل البياني لنتائج التنقيب في البيانات:
- تساعد أساليب تمثيل البيانات بيانياً **Data visualization** في العرض المرئي لمخرجات التنقيب في البيانات، ويمكن عرض البيانات بشكل مرئي على بعدين:
 - رسم شكل الانتشار،
 - والمُدرج التكراري، الخ
- أو ثلاثة أبعاد ويمكن أن يكون العرض ساكناً أو متحركاً، وتجدر الإشارة إلى أهمية الجمع بين أساليب التنقيب في البيانات (DM) آنفة الذكر وأساليب تمثيل البيانات بيانياً، وهذا ما يسمى التنقيب في البيانات المرئية **Visual Data Mining** الذي يساعد في الكشف عن علاقات ومجموعات متجانسة وقيم متطرفة.

24

أساليب التنقيب في البيانات

- **التنقيب في بيانات المجالات الوظيفية المختلفة:**
- يوفر مستودع البيانات للمديرين كماً ضخماً من البيانات التاريخية الداخلية والخارجية المتعلقة بكم كبير من المتغيرات المختلفة، كما يمكن أن تنبثق من مستودع البيانات عدة منافذ بيانات لكل مجال وظيفي (تسويق ومبيعات، إنتاج، مشتريات ومخزون) وهنا يأتي دور أدوات التنقيب في تلك البيانات لمساعدة المديرين في الكشف عن معارف جديدة في تلك المجالات الوظيفية المختلفة، وفيما يلي توضيح للتنقيب في بيانات أهم المجالات الوظيفية:

25

التنقيب في بيانات وظائف المنظمة

- **التنقيب في بيانات التسويق والمبيعات:**
- يحتوي مستودع أو منفذ بيانات التسويق والمبيعات على بيانات عن ذلك المجال، ومن أمثلة تلك البيانات:
- **المنتجات، وكمية المبيعات، وقنوات التوزيع، والعملاء، ومناطقهم وأسلوب النقل لهم، والسعر والقيمة، والحسم الممنوح، وتفصيل تكاليف المنتجات، الخ،**
- ويمكن أن يتم التنقيب في بيانات ذلك المجال للتعرف على علاقات وقواعد جديدة (معارف) تفسر سلوك تلك البيانات أو المتغيرات، ويمكن الاستفادة من تلك المعارف المكتشفة عند اتخاذ قرارات تصميم أو إعادة تصميم شبكة التوزيع، و/أو قرارات متعلقة بتخصيص كمية المنتجات المختلفة في الفترات الزمنية القادمة على قنوات التوزيع والعملاء المختلفين في المناطق المختلفة.

26

التنقيب في بيانات وظائف المنظمة

• التنقيب في بيانات التسويق والمبيعات-تابع:

ويفضل أن يتم التنقيب في بيانات المبيعات للمستهلك النهائي، التي يمكن الحصول عليها بموجب ترتيبات التعاون في مجال تبادل البيانات إلكترونياً مع العملاء الرئيسيين،

• فمثلاً يمكن أن يقوم تاجر التجزئة بتزويد الشركة الصناعية ببيانات بيع المنتجات المختلفة الخاصة بالمنتج وبمنافسيه، مقابل حصوله على بيانات مبيعات ذلك المنتج للمنافسين الرئيسيين لتاجر التجزئة،

• وإذا حصلت الشركة الصناعية على تلك البيانات من بعض العملاء الرئيسيين فقط، فإن ذلك يكفي للتنقيب فيها، وكشف علاقات وقواعد جديدة أو تأكيد علاقات وقواعد متوقعة بين كثير من المتغيرات ذات العلاقة بالمبيعات، وتقدير قوة العلاقة أو التأثير بشكل كمي محدد.

27

التنقيب في بيانات وظائف المنظمة

• التنقيب في بيانات التسويق والمبيعات-تابع:

• وتجدر الإشارة إلى أنه بدأت كثير من سلاسل السوبر ماركت العالمية تبني معلومات و/أو بيانات نقاط البيع للشركات الصناعية، ولا تحتوي تلك البيانات على البيانات التاريخية لمبيعات المنتجات المختلفة فحسب، بل وعلى البيانات الديموغرافية لمستهلكي تلك المنتجات، ويتم الحصول على البيانات الديموغرافية للمستهلك من خلال بطاقة ولاء المستهلك التي يتم توزيعها على العملاء (المستهلكين) الراغبين، حيث تمكنهم تلك البطاقة من الحصول على حسميات حسب قيمة مشترياتهم المتراكمة، وقد كانت سلسلة Safeway من أوائل سلاسل السوبر ماركت التي عملت ببيع البيانات والمعلومات، حيث تقوم مثلاً ببيع قائمة تكرارات مشتري منتج ببسي لشركة كوكا كولا، وتبيع لشركة ببسي البيان نفسه عن شركة كوكا كولا.

28

التنقيب في بيانات وظائف المنظمة

• التنقيب في بيانات الإنتاج:

• يمكن أن يتم التنقيب في البيانات التاريخية المتعلقة بمتغيرات الإنتاج مثل متغير كل من:

• **المنتجات، وأوامر الإنتاج، والمصانع المنتجة، وكميات كل أمر، وتفاصيل تكاليف المنتج في كل أمر، الخ،**

• ومن خلال ذلك التنقيب في البيانات، يمكن التعرف على علاقات وقواعد جديدة تفسر سلوك تلك المتغيرات، للاستفادة منها عند اتخاذ قرارات تصميم أو إعادة تصميم شبكة التصنيع (فتح أو إغلاق مصنع أو خط إنتاج)، و/أو قرارات متعلقة بكمية وتوقيتات وأماكن إنتاج المنتجات المختلفة.

29

التنقيب في بيانات وظائف المنظمة

• التنقيب في بيانات المشتريات ومخزون المواد الأولية:

• يمكن أن يتم التنقيب في البيانات التاريخية المتعلقة بذلك المجال:

• **(المواد الخام، ومورديها، وأسلوب شحنها، ومدة تخزينها، وتفاصيل تكاليفها، الخ)**

• بالإضافة لمتغيرات الظروف الاقتصادية والسياسية لدول الموردين، ومن خلال ذلك التنقيب في البيانات يمكن التعرف على علاقات وقواعد جديدة تفسر سلوك تلك المتغيرات، للاستفادة منها عند اتخاذ قرارات متعلقة بإقامة علاقات سلسلة توريد (قائمة على التحالف والتعاون) مع بعض الموردين، و/أو قرارات متعلقة بكميات وتوقيتات المشتريات من المواد المختلفة وأساليب شحنها وتخزينها.

30

برمجيات التنقيب في البيانات

- من أمثلة برمجيات التنقيب في البيانات برامج الإحصاء مثل:
 - برنامج STATISTICA (فيه غالبية أساليب التنقيب في البيانات)،
 - وبرنامج MINITAB
 - وبرنامج SPSS ، الخ،
- كما تنتشر برمجيات التنقيب في البيانات كمكونات إضافية في تطبيقات قواعد البيانات وتطبيقات جداول الانتشار مثل:
 - Excel