

المحاضرة الخامسة: التنقيب في البيانات Data mining

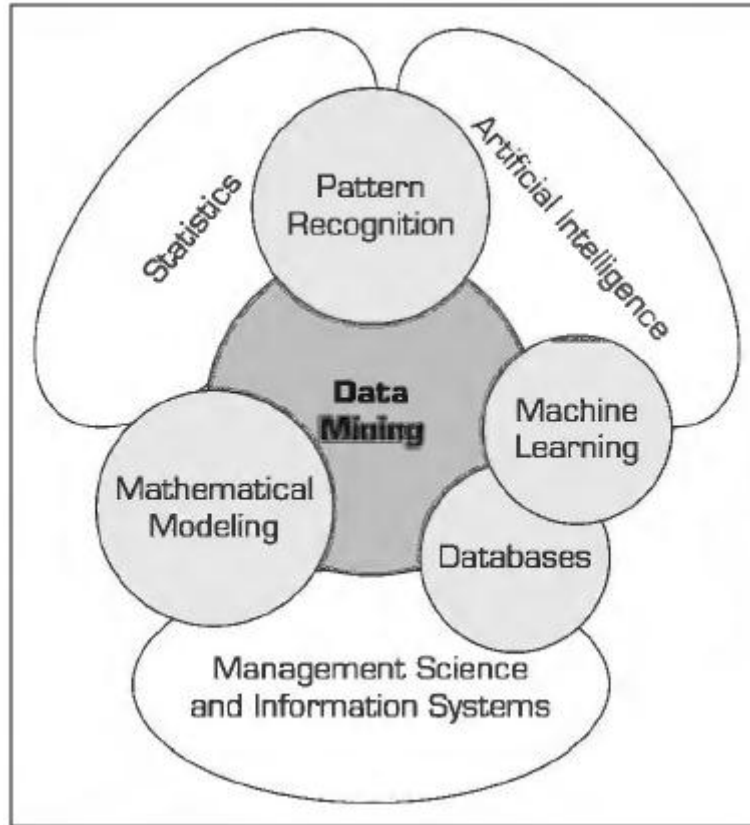
1- مفهوم التنقيب في البيانات

إن استخدام تقنيات التنقيب في البيانات يوفر للمؤسسات في جميع المجالات القدرة على استكشاف والتركيز على أهم المعلومات في قواعد البيانات، كما تركز تقنيات التنقيب في البيانات كذلك على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات مما يسمح باتخاذ القرارات الصحيحة واتخاذها في الوقت المناسب.

ظهر ما يسمى بالتنقيب في البيانات Data Mining في أواخر الثمانيات في الولايات المتحدة الأمريكية كتقنية تهدف فيالبحثن العلاقات والأنماط الموجودة في قواعد البيانات الكبيرة ولكنها مخفية بين الكم الهائل من البيانات¹، تعتمد على الخوارزميات الرياضية والتي تعتبر أساس التنقيب عن البيانات وهي مستمدة من العديد من العلوم مثل علم الإحصاء والرياضيات والمنطق، والذكاء الاصطناعي والنظم الخبيرة، وعلم الآلة، وغيرها من العلوم التي من العلوم الذكية وغير التقليدية.

فالتنقيب في البيانات هو "البحث عن العلاقات والأنماط الموجودة في قواعد البيانات الكبيرة ولكنها مخفية بين الكم الهائل من البيانات"²

الشكل رقم (01): التنقيب في البيانات كمزيج من التخصصات المتعددة



¹Rafi Ahmad Khan, S.M. K. Quadri, **Business Intelligence: An Integrated Approach**, Business Intelligence Journal - January, Vol.5 No.1, 2012, p :68.

²Usama Fayyad et autres, **From Data Mining to Knowledge Discovery in Databases**, American Association for Artificial Intelligence, 1996, pp: 37-54

2- أدوات التنقيب عن البيانات

نماذج التنقيب في البيانات نوعان: النماذج التنبؤية (Predictives) والنماذج الوصفية (Descriptives). النماذج التنبؤية تهدف إلى التنبؤ بقيمة بعض الخصائص، مثل التنبؤ باحتمال الشراء للزبون، أما النماذج الوصفية فتتقسّم إلى صنفين: نماذج العقدة التي تسمح بتجميع الأفراد، والأحداث، أو المنتجات في عناقيد، ونماذج الارتباط التي تسمح بتحديد العلاقات بينهم.

وهناك عدة أدوات للتنقيب في البيانات أهمها:

2-1- التلخيص (Summarization): يشير التلخيص إلى أساليب تفتيت كتل البيانات الكبيرة إلى مقاييس موجزة، توفر وصفا عاما للمتغيرات وعلاقاتها، ومن الأمثلة على أساليب التلخيص نذكر: المتوسطات، والجاميع، والإحصائيات الوصفية التي تتضمن مقاييس النزعة المركزية مثل المتوسط الحسابي والوسيط والمنوال، ومقاييس التشتت مثل الانحراف المعياري، وعلى الرغم من أن مقاييس التلخيص تعطي صورة كبيرة عن بعض التفاصيل ذات العلاقة فإنها غالبا ما تحمل تفاصيل أخرى ذات أهمية كبيرة تتعلق بسلوك المستهلك خصوصا¹.

2-2- التصنيف (Classification): يتمثل التصنيف في تفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى، هذه الخاصية هي عموما كيفية، ويمكن انجاز التصنيف بالاعتماد على الأساليب الإحصائية القديمة مثل الانحدار والتحليل التمييزي، أو بالاعتماد على أساليب حديثة نسبيا مثل قوى الارتباط والاستنتاج المستند إلى الحالة والشبكات العصبية².

2-3- التنبؤ (Prediction): يشبه التنبؤ التصنيف أو التقدير، ما عدا أن البيانات تصنف على أساس التنبؤ بسلوكها المستقبلي أو تقدير قيمتها المستقبلية. حيث أن المتغير التابع المتنبأ به هو متغير كمي. ومن الأدوات التقليدية المستخدمة في التنبؤ نذكر على سبيل المثال: الانحدارات بأنواعها والتحليل التمييزي. أما الأساليب الجديدة فتشتمل على قواعد الارتباط وشجرة القرار والشبكات العصبية³.

2-4- العقدة أو التجزئة (Clustering): يتمثل التجميع العنقودي أو التجزئة إلى قطاعات في البحث عن مجموعات متجانسة في مجتمع من الأفراد، ويشير التجميع العنقودي أو التجزئة إلى قطاعات إلى عملية تشكيل مجموعات أو قطاعات مؤلفة من أفراد أو أصحاب أسر، وذلك بالاستناد إلى معلومات متضمنة في مجاميع من المتغيرات التي تصفهم. والغرض من التجميع العنقودي المساعدة على تطوير برامج تسويقية مصممة على مقاسات الزبائن أنفسهم، والتي بالإمكان استخدامها لاستهداف أعضاء لكل قطاع من هذه القطاعات على أمل ترغيبهم في تكرار الشراء أو التحول إلى زبائن موالين.

2-5- تحليل الارتباط (Rule Analysis):

يتمثل الارتباط في البحث عن علاقات أو ارتباطات موجودة بين عدة خصائص، ويشير تحليل الارتباط إلى مجموعة من الأساليب التي تستخدم لربط أنماط الشراء عبر القطاعات المتقاطعة أو عبر الوقت. فمثلا يقوم أسلوب تحليل سلة السوق (نوع من أنواع الارتباط) باستخدام المعلومات الكامنة في السلع التي اشتراها المستهلكون فعليا للتنبؤ بالسلع المحتمل شراؤها إياها إذا ما تم تقديم عروض خاصة لهم أو إذا تم تعريفهم بهذه السلع.

¹ بشير عباس العلاق، الإدارة الرقمية: المجالات والتطبيقات، مركز الإمارات للدراسات والبحوث الاستراتيجية، ابوظبي، 2005، ص 92.

² Bazsalica M., Naim P., Data mining pour le Web, éd. Eyrolles, Paris, 2001, pp.68-69.

³ Berry J. A. M., Linoff G. S., Data Mining Techniques For Marketing, Sales, and Customer Relationship Management, 2° ed., Wiley Publishing, INC, Indianapolis, 2004, p. 10.

2-6- الكشف عن التغيرات أو الانحرافات (Change and deviation detection): يركز على استكشاف

التغيرات المهمة جداً في البيانات من خلال قياسات سابقة أو قيم معيارية.

3- خصائص التنقيب في البيانات

التنقيب في البيانات هو عملية تحليل كميات كبيرة من البيانات لاستخراج أنماط أو معلومات مفيدة، تتضمن هذه العملية عدة

خصائص رئيسية:

- **تحليل كميات كبيرة من البيانات:** يعتمد التنقيب في البيانات على تحليل مجموعات ضخمة من البيانات، مما يتطلب تقنيات

متقدمة للتعامل مع التعقيدات الحسابية.

- **استخراج الأنماط:** يهدف التنقيب إلى اكتشاف الأنماط والعلاقات المخفية في البيانات، مما يساعد في فهم سلوك العملاء أو

تحسين العمليات التجارية

- **تحديد الأهداف:** من المهم تحديد أهداف واضحة قبل البدء في عملية التنقيب، مثل معرفة سلوك العملاء أو تقليل التكاليف أو

اكتشاف الاحتيال.

- **استخدام تقنيات متعددة:** يتضمن التنقيب في البيانات استخدام مجموعة متنوعة من التقنيات والخوارزميات لاستخراج المعلومات، مما

يجعله أداة قوية في مجالات متعددة مثل التسويق والمالية.

- **تقديم رؤى مفيدة:** يمكن أن يوفر التنقيب في البيانات رؤى قيمة تساعد الشركات في اتخاذ قرارات استراتيجية أفضل، مما يمنحها ميزة

تنافسية.

4- تطبيقات تنقيب البيانات

أصبح تنقيب البيانات أداة شائعة في معالجة العديد من المشكلات والفرص التجارية المعقدة. وقد ثبت أنه ناجح للغاية ومفيد في

العديد من المجالات، وبعضها موضح من خلال الأمثلة التالية. الهدف من العديد من تطبيقات تنقيب البيانات التجارية هو حل مشكلة

ملحة أو استكشاف فرصة تجارية ناشئة من أجل خلق ميزة تنافسية مستدامة.

• **إدارة علاقات العملاء (CRM):** هي امتداد للتسويق التقليدي. الهدف من إدارة علاقات العملاء هو إنشاء

علاقات فردية مع العملاء من خلال تطوير فهم عميق لاحتياجاتهم ورغباتهم. مع مرور الوقت، تبني الشركات علاقات مع عملائها من

خلال مجموعة متنوعة من التفاعلات (مثل استفسارات المنتجات، المبيعات، طلبات الخدمة، تقييمات المنتجات، الاتصالات عبر وسائل

التواصل الاجتماعي)، مما يؤدي إلى تراكم كميات هائلة من البيانات. عند دمج هذه المعلومات الغنية بالبيانات مع السمات الديموغرافية

والاجتماعية والاقتصادية، يمكن استخدام هذه البيانات في: (1) تحديد الأكثر احتمالاً للاستجابة/ الشراء من المنتجات/ الخدمات

الجديدة (أي، تصنيف العملاء)؛ (2) فهم الأسباب الجذرية لفقدان العملاء من أجل تحسين احتفاظ العملاء (أي، تحليل التناقص)؛

(3) اكتشاف الارتباطات المتغيرة مع الزمن بين المنتجات والخدمات لتعظيم المبيعات وقيمة العملاء؛ و (4) تحديد العملاء الأكثر ربحية

واحتياجاتهم التفضيلية لتعزيز العلاقات وتعظيم المبيعات.

• **البنوك:** يمكن أن تساعد تقنيات التنقيب في البيانات البنوك في ما يلي: (1) أتمتة عملية تقديم طلبات القروض من خلال التنبؤ بدقة

بأكثر الأشخاص عرضة للتعثر؛ (2) اكتشاف المعاملات الاحتيالية في بطاقات الائتمان والمعاملات المصرفية عبر الإنترنت؛ (3) تحديد

طرق لتعظيم قيمة العملاء من خلال بيعهم المنتجات والخدمات ؛ (4) تحسين العائد النقدي من خلال التنبؤ بدقة بتدفق النقد في الكيانات المصرفية (مثل آلات الصراف الآلي، والفروع المصرفية).

• **البيع بالتجزئة واللوجستيات:** في صناعة البيع بالتجزئة، يمكن استخدام تقنيات التنقيب في البيانات ل: (1) التنبؤ بحجم المبيعات بدقة في مواقع البيع بالتجزئة المحددة لتحديد مستويات المخزون الصحيحة؛ (2) تحديد العلاقات البيعية بين المنتجات المختلفة لتحسين تصميم المتجر وتحسين العروض الترويجية؛ (3) التنبؤ بمستويات الاستهلاك لمختلف أنواع المنتجات (استنادًا إلى الظروف الموسمية والبيئية) لتحسين اللوجستيات وبالتالي تعظيم المبيعات؛ (4) اكتشاف أنماط مثيرة للاهتمام في حركة المنتجات (خصوصًا للمنتجات التي لها فترة صلاحية محدودة لأنها عرضة للتلف والفساد والتلوث) في سلسلة التوريد من خلال تحليل البيانات .

• **التصنيع والإنتاج.** يمكن للمصنعين استخدام التنقيب في البيانات في: (1) التنبؤ بفشل الآلات قبل حدوثها من خلال استخدام البيانات الحسية (مما يمكن ما يُسمى بالصيانة القائمة على الحالة)؛ (2) تحديد الاختلالات والتشوهات في أنظمة الإنتاج لتحسين القدرة التصنيعية؛ و(3) اكتشاف أنماط جديدة لتحديد وتحسين جودة المنتج.

• **الوساطة وتداول الأوراق المالية:** يستخدم الوسطاء والمتداولون التنقيب في البيانات في: (1) التنبؤ بموعد وكم ستتغير أسعار بعض السندات؛ (2) توقع نطاق واتجاه تقلبات الأسهم؛ (3) تقييم تأثير قضايا وأحداث معينة على حركة السوق بشكل عام؛ و(4) تحديد ومنع الأنشطة الاحتيالية في تداول الأوراق المالية.

• **التأمين.** تستخدم صناعة التأمين تقنيات التنقيب في البيانات في: (1) توقع مبالغ المطالبات لتكاليف التغطية العقارية والطبية من أجل تخطيط الأعمال بشكل أفضل؛ (2) تحديد خطط الأسعار المثلى بناءً على تحليل المطالبات وبيانات العملاء؛ (3) التنبؤ بالعملاء الذين من المرجح أن يشتروا سياسات جديدة بميزات خاصة؛ و(4) تحديد ومنع المدفوعات غير الصحيحة للمطالبات والأنشطة الاحتيالية.

• **أجهزة الكمبيوتر والبرمجيات:** يمكن استخدام تنقيب البيانات ل: (1) التنبؤ بفشل وحدات التخزين قبل حدوثها بفترة طويلة؛ (2) تحديد وتصنيف المحتوى غير المرغوب فيه على الويب ورسائل البريد الإلكتروني؛ (3) اكتشاف ومنع خروقات أمان الشبكات الحاسوبية؛ (4) تحديد المنتجات الربحية التي قد تكون غير آمنة.

• **السياحة (شركات الطيران، الفنادق/المنتجعات، شركات تأجير السيارات):** يتم استخدام تنقيب البيانات في صناعة السفر بعدة طرق. يتم استخدامه بنجاح ل (1) التنبؤ بمبيعات الخدمات المختلفة (أنواع المقاعد في الطائرات، أنواع الغرف في الفنادق/المنتجعات، أنواع السيارات في شركات تأجير السيارات) من أجل تسعير الخدمات بشكل مثالي لزيادة الإيرادات كدالة للمعاملات المتغيرة مع الزمن (المعروفة عادة بإدارة العائدات)؛ (2) التنبؤ بالطلب في مواقع مختلفة لتخصيص الموارد التنظيمية المحدودة بشكل أفضل؛ (3) تحديد العملاء الأكثر ربحية وتقديم خدمات مخصصة لهم للحفاظ على تكرار أعمالهم؛ (4) الاحتفاظ بالموظفين القيمين من خلال تحديد ومعالجة الأسباب الجذرية للتسرب.

• **الرعاية الصحية:** للتنقيب في البيانات عدد من التطبيقات في مجال الرعاية الصحية. يمكن استخدامه ل: (1) تحديد الأشخاص الذين ليس لديهم تأمين صحي والعوامل الكامنة وراء هذه الظاهرة غير المرغوب فيها؛ (2) تحديد علاقات التكلفة والفائدة الجديدة بين العلاجات المختلفة لتطوير استراتيجيات أكثر فعالية؛ (3) التنبؤ بمستوى ووقت الطلب في مواقع الخدمة المختلفة لتخصيص الموارد التنظيمية بشكل مثالي؛ (4) فهم الأسباب الكامنة وراء تسرب العملاء والموظفين.

- صناعة الترفيه. يتم استخدام تنقيب البيانات بنجاح من قبل صناعة الترفيه لـ (1) تحليل بيانات المشاهدين لتحديد البرامج التي يجب عرضها خلال وقت الذروة وكيفية زيادة العوائد من خلال معرفة أين يتم إدراج الإعلانات؛ (2) توقع النجاح المالي للأفلام قبل إنتاجها لاتخاذ قرارات الاستثمار وتحسين العوائد؛ (3) توقع الطلب في مواقع وأوقات مختلفة لجدولة الفعاليات الترفيهية بشكل أفضل وتخصيص الموارد بشكل مثالي؛ و (4) تطوير سياسات تسعير مثلى لزيادة الإيرادات.