

تقنية التنقيب في البيانات كأداة مساعدة في إتخاذ القرارات

الأستاذ: محمودي مليك الأستاذة: زروخي صباح

أستاذ محاضر أستاذة محاضرة

كلية العلوم الإقتصادية والتجارية وعلوم التسيير

جامعة محمد بوضياف المسيلة

المخلص:

إن استخدام تقنية التنقيب في البيانات يوفر للمؤسسات في جميع المجالات القدرة على إستكشاف والتركيز على أهم المعلومات في قواعد البيانات، كما تركز تقنيات التنقيب في البيانات كذلك على بناء التنبؤات المستقبلية وتبيان السلوك والإتجاهات مما يسمح بإتخاذ القرارات الصحيحة وفي الوقت المناسب.

الكلمات المفتاحية: البيانات، التنبؤ، التنقيب في البيانات.

Abstract

The use of techniques Data mining provides for institutions in all and focus on the most important information in areas the ability to explore data mining as well as to build a the data bases, also focusing techniques future forecasts and exploration behavior and trends, which allows to take the right decisions and taken in a timely manner.

Keywords: Data, Forecasting, Data mining.

تمهيد:

أدى التطور في العلم والاقتصاد وتكنولوجيا المعلومات والاتصالات إلى زيادة كمية البيانات الرقمية، ومع هذه الكميات الهائلة من البيانات لم تعد وسائل التحليل التقليدية (الإحصائية مثلا) قادرة على التعامل معها، لذا ظهرت العديد من الدراسات منذ أواخر الثمانيات في محاولة لحل تلك المشكلات مع البحث عن حلول تجمع بين عدة تخصصات سواء كانت الإحصاءات أو قواعد البيانات والذكاء الصناعي وتمييز النماذج المختلفة والحوسبة ومن ثم تم التوصل إلى التنقيب عن

البيانات *Data mining* واكتشاف المعرفة اللذين أثبتا وجودها كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات وذلك بتحويلها من بيانات متراكمة وغير مفهومة إلى معلومات قيمة يمكن استغلالها والاستفادة منها بعد ذلك لتصبح معرفة.

لذا يعد التنقيب عن البيانات من أسرع المجالات نموا في تخصصات علم الحاسب الآلي ولقد جاءت شهرته وانتشاره من الحاجة المتزايدة لأدوات تساعد في تحليل الكميات الهائلة من البيانات وفهمها، وتنتج هذه البيانات يوميا بواسطة المؤسسات المختلفة مثل البنوك وشركات التأمين ومستودعات البيع وعلى شبكة الإنترنت وصاحب هذا الانفجار في البيانات زيادة هائلة أيضا في استخدام لحاسبات الآلية والمساحات الضوئية والكاميرات الرقمية وغيرها.

أولا: مفهوم التنقيب في البيانات

ظهر مصطلح التنقيب في البيانات في منتصف التسعينات في الولايات المتحدة الأمريكية وهو يجمع ما بين الإحصاء وتكنولوجيا الإعلام (قواعد البيانات الذكاء الاصطناعي التعلم الآلي *Machine Learning*).

وتوجد عدة تعريفات لهذا المفهوم نذكر منها:

التعريف 01 : "الاستكشاف الآلي أو المؤتمت لأنماط شائقة وغير جلية مخفية في قاعدة بيانات معينة"¹.

التعريف 02 : "إجراءات تحليل دقيقة وذكية تفاعلية وتسلسلية، تسمح لمسيري النشاطات عند استخدام هذه الإجراءات بإتخاذ قرارات والقيام بأعمال ملائمة في صالح النشاط المسؤولين عنه والمؤسسة التي يعملون بها"².

التعريف 03 : هي عبارة عن تحليلات لكمية كبيرة من البيانات بغرض إيجاد قواعد ونماذج تستخدم من طرف أصحاب القرار للتنبؤ بالسلوك المستقبلي لظاهرة ما"³.

تعريف 04 : "تحليل لمجموعات كبيرة الحجم من البيانات المشاهدة للبحث عن علاقات محتملة وتلخيص للبيانات في أشكال جديدة مفهومة ومفيدة لمستخدمها"⁴

من خلال التعريفات السابقة يمكن القول أن التنقيب في قواعد البيانات يهدف إلى استخلاص المعلومات المخبأة فيها، فاستخدامها يوفر للمؤسسات في جميع المجالات الاستكشاف والتركيز على أهم المعلومات في قواعد البيانات بالإضافة إلى كثرة البيانات الموجودة والمخزنة في قواعد البيانات حيث أصبحت موضوع تساؤل من العديد من الباحثين للاستفادة منها، ومع زيادة انتشار مستودعات التخزين الضخمة (*data warehouses*)، لذا أصبح من الضروري إيجاد تقنيات وطرق لإستخلاص المعلومات والمعرفة من البيانات المكثفة وإستغلالها في حل المشاكل وإتخاذ القرارات، بإستخدام تطبيقات الحاسوب الحديثة كالتي تعتبر تكنولوجيا حديثة ذكية قائمة علي جعل الحاسوب "يفكر كما يفكر الإنسان ويفعل كما يفعل الإنسان " جاءت فكرة الكشف والتنقيب في هذه البيانات بطرق ذكية للمساهمة في حل المشاكل وإتخاذ القرارات وتعتبر خطوة من خطوات استكشاف المعرفة من قواعد البيانات.

ثانيا: العوامل المحركة لثورة التنقيب في البيانات

يمكن تقسيم العوامل المحركة لثورة التنقيب في البيانات الى قسمين رئيسيين هما⁵:

أ. العوامل المتعلقة بجانب العرض:

- ❖ تأثيرات تطورات تكنولوجيا المعلومات خاصة تكنولوجيا تخزين البيانات و معالجتها؛
- ❖ انخفاض تكاليف الاتصالات الالكترونية مما يسر الوصول الى قواعد البيانات؛
- ❖ ظهور أساليب تحليل جديدة من أبرزها الشبكات العصبية، نظم الخوارزميات الوراثية، شجرة القرار و قواعد الاستقراء؛
- ❖ ثورة اعمار الحاسوب والتي تمخضت عن ولادة أجيال جديدة من البرمجيات الصديقة للمستخدم، مثل: مايكروسوفت ويندوز، وبرمجيات العميل- الخادم (*Client-Server*) غيرها؛
- ❖ التطور الهائل في أساليب التخزين و من أبرز هذه الأساليب تلك المتعلقة بمستودعات البيانات (*Data Warehouse*) و أسواقها.

ب. العوامل المتعلقة بجانب الطلب:

❖ تنامي الحاجة الى نتائج تحليلية سريعة من طرف المؤسسات المنافسة الشديدة و التغييرات 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؛

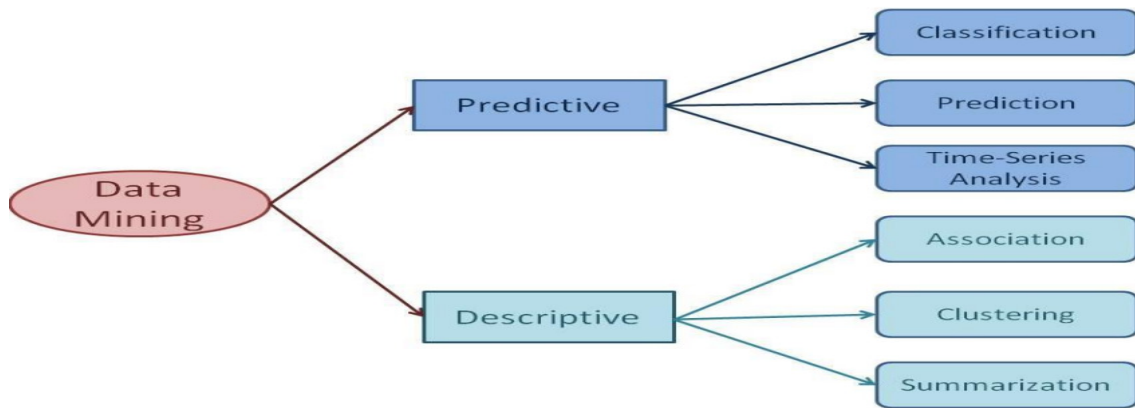
❖ ازدهار حركة التأليف و النشر خاصة في مجال تكنولوجيا المعلومات و الأنترنت و كل ما يرتبط بعصر الاقتصاد الرقمي؛

ثالثا: أدوات التنقيب في 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 فتتقسم إلى صنفين: نماذج العنقدة التي تسمح بتجميع الأفراد، و الأحداث، أو المنتجات في عناقيد، ونماذج الارتباط التي تسمح بتحديد العلاقات بينهم، والشكل التالي يوضح ذلك:

الشكل 01: أدوات التنقيب في البيانات



كما أن هناك عدة أدوات للتنقيب في البيانات، نذكر أهمها:

1. التلخيص (Summarization):

يشير التلخيص إلى أساليب تفتتت كتل البيانات الكبيرة إلى مقاييس موجزة، توفر وصفا عاما للمتغيرات وعلاقاتها⁶، و من الأمثلة على أساليب التلخيص نذكر: المتوسطات، والمجاميع،

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 تهمل تفاصيل أخرى ذات أهمية كبيرة تتعلق بسلوك المستهلك خصوصا.

2. التصنيف (Classification):

يتمثل التصنيف في تفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى. هذه الخاصية هي عموما كيفية⁷، 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 العصبية.

وكأمثلة عن طرق التصنيف المستعملة كجزء من تطبيقات استكشاف المعرفة التي تتضمن تصنيف اتجاهات الأسواق المالية، والتحديد الآلي للأشياء المهمة في صورة كبيرة من قواعد البيانات.

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 و التحليل التمييزي، أما الأساليب الجديدة فتشتمل على قواعد الارتباط و شجرة القرار والشبكات العصبية والخوارزميات الوراثية⁸.

4. العنقدة أو التجزئة (Clustering) :

يتمثل التجميع العنقودي أو التجزئة إلى قطاعات في البحث عن مجموعات متجانسة في مجتمع من الأفراد⁹، 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.

وتتم أساليب التجميع العنقودي غالبا بمساعدة أساليب التحليل العنقودي الإحصائية والأساليب المستندة إلى شجرة القرار، والشبكات العصبية والخوارزميات الوراثية.

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 شراؤهم إياها إذا ما تم تقديم عروض خاصة لهم أو إذا تم تعريفهم بهذه السلع¹⁰.

6. الكشف عن التغيرات أو الانحرافات (*Change and deviation detection*):

يرتكز على استكشاف التغيرات المهمة جدا في البيانات من خلال قياسات سابقة أو قيم معيارية.

رابعاً: مراحل عملية التنقيب في البيانات

يمكن تلخيص مراحل و خطوات عملية التنقيب في البيانات كما يلي¹¹:

1. فهم طبيعة الأعمال (*Business Understanding*): يعتبر المطلب الأول لاكتشاف المعرفة هو فهم

المشاكل و المسائل التي تواجهها الأعمال. و بمعنى آخر، كيف يمكن تحقيق المنفعة الأعظم من التنقيب في البيانات، مما يتطلب وجود صيغة واضحة و محددة لأهداف الأعمال.

2. فهم البيانات (*Data Understanding*): تعتبر مسألة معرفة ماهية وطبيعة البيانات عامل مهم

في نجاح عملية التنقيب في البيانات و اكتشاف 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 و الكفاءة 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.

ويمكن تلخيص الخطوات الضرورية لعملية فهم البيانات كآتي:

❖ **تجميع البيانات (Data Collection):** و هي الخطوة الموجهة نحو تحديد مصدر البيانات في

الدراسة بما في ذلك استخدام البيانات العامة الخارجية مثل الضرائب و غيرها.

❖ **توصيف البيانات (Data Description):** و هي الخطوة التي تركز على توصيف محتويات

الملف الواحد من الملفات أو الجداول.

❖ **جودة البيانات و تحقيقها (Data Quality and Verification):** هذه الخطوة تحدد ما إذا كان

تقليل أو إهمال بعض البيانات غير الضرورية أو كونها رديئة الجودة و قد لا تنفع في

الدراسة. لأن النموذج الجيد يحتاج إلى بيانات جيدة مما يتوجب أن تكون البيانات صحيحة و

ذات مضمون دقيق.

❖ **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. و تعتبر هذه الخطوة مهمة و ضرورية لأنها تركز على تطوير الفرضيات

المتعلقة بالمشكلة قيد الدراسة.

3. **تهيئة البيانات (Data Preparation):** و تشمل الخطوات التالية:

❖ **الاختبار (Selection)** و تعني اختيار المتغيرات المتوقعة و حجم العينة.

❖ **صياغة المتغيرات و تحويلها (Construction and Transformation Variables)** حيث يجب دائما

أن تصاغ المتغيرات الجديدة لبناء النماذج الفعالة.

❖ **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.

❖ **تصميم و تنسيق البيانات (Data Formatting)** حيث تتعلق هذه الخطوة في إعادة ترتيب حقول البيانات كما يتطلب في نموذج التنقيب في البيانات.

4. **صياغة نماذج الحل و ثبوتها (Model Building and Validation):** إن بناء و صياغة نموذج الحل السليم و الدقيق يتم من خلال عملية الخطأ و الصواب، حيث كثيرا 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.

5. **التقييم و تحليل 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. و تؤدي هذه المقارنة أو المفاضلة إلى التحقق من دقة النموذج.

6. **نشر وتوزيع النموذج (Model Deployment):** حيث تشمل 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.

خامسا: مجالات تطبيق التنقيب في قواعد البيانات في منظمات الأعمال

لقد عرف التنقيب في البيانات نجاحا كبيرا و ازدادت تطبيقاته في منظمات الأعمال وذاك راجع إلى النمو الكبير في البيانات خاصة الموجودة في قواعد البيانات و مخازن البيانات، و المنافسة الشديدة في السوق، مما يدفع هذه المنظمات إلى الاستفادة القصوى من البيانات التي تمتلكها باستخدام التنقيب في البيانات.

وقد بدأ استخدام التنقيب في البيانات في منظمات التوزيع، ثم انتقلت بعدها سريعا إلى البنوك، ومؤسسات التأمين، و شركات الهاتف (خاصة الهاتف النقال)، ومؤسسات المياه والكهرباء، ومؤخرا مؤسسات النقل الجوي والنقل بالسكك الحديدية... الخ. وكانت تطبيقاتها الأولى في ميدان إدارة العلاقة مع الزبون من خلال تحليل سلوك الزبائن من أجل الوصول إلى جعلهم أكثر وفاء و اقتراح منتجات وفق رغباتهم.

عموما، هناك عدة ميادين لتطبيق التنقيب في البيانات في منظمات الأعمال نذكر منها¹²:

1. **التسويق:** استخدمت الشبكات العصبية الاصطناعية في دراسات التسويق المستهدف بما في ذلك الحصص السوقية. و قد ساعدت هذه الأساليب التسويق على استخدام نهج تخصيص الزبائن وفقا إلى الحقائق الديمغرافية (السكانية) الأساسية مثل الجنس والعمر والمجموعات وكذلك أنماطهم الشرائية.
2. **التجزئة:** لقد استخدمت أساليب التنقيب في البيانات بصورة فعالة في التنبؤ بالمبيعات حيث أخذت العديد من المتغيرات في الدراسات مثل متغيرات السوق المتعددة، قدرات الزبائن المستندة على العادات المتبعة في الشراء. كما ساعدت أساليب مثل تحليل السلة الشرائية أو السلة السوقية كثيرا على إيجاد أي من المنتجات التي يمكن أن تشتري سوية من قبل الزبائن.
3. **البنوك:** لقد أثبتت تنبؤات الأعمال والمالية على أنها الأساليب الممتازة في تطبيقات أساليب التنقيب في البيانات. وقد استخدمت هذه الأساليب في إيجاد الأسعار المضمونة وتنبؤات السعر المستقبلية وأداء الأسهم. كما وقد حققت استخدامات مثل هذه الأساليب النجاح في تطوير أنظمة القياس الرقمية في تحديد مخاطرة القروض والاحتمالات المالية.
4. **التأمين:** لقد استخدمت أساليب التنقيب في البيانات كذلك وبصورة واسعة في مجال شركات التأمين وخاصة في تخصيص مجاميع الزبائن بهدف تحديد أسعار البوليصا وذبذبات المطالبات المتوقعة مستقبلا، وتحديد المطالبات المزيفة.

5. **الاتصالات:** حيث استخدمت وبصورة فعالة أساليب التنقيب في البيانات مثل الشبكات العصبية في محاولة تقليل - الخفضة- التي تظهر عندما تفقد المنظمة هؤلاء الزبائن خلال المنافسة وذهابهم إلى المنافسين الآخرين.

6. **إدارة العمليات:** حيث استخدمت الشبكات العصبية في عمليات التخطيط والجدولة وإدارة المشاريع بالإضافة إلى إدارة الجودة.

خاتمة:

لقد أدى إدماج مزايا تكنولوجيا المعلومات مع الطرق الإحصائية والخوارزميات إلى توفر الإمكانيات اللازمة للتنبؤ بالسلوك المستقبلي ومن ثم وضع الحلول المناسبة للمشكلات قبل وقوعها في حال إمكان حدوثها، أو من باب التنبؤ بهدف التطوير والتحديث بشكل عام في شتى المجالات كل ذلك باستخدام تقنيات التنقيب في البيانات التي تشكل مرحلة من مسار أشمل هو استكشاف المعرفة في قواعد البيانات والتي أصبحت من الهموم الكبيرة التي تقع على عاتق الدول بكافة مؤسساتها بشكل عام.

هوامش الدراسة

¹ بشير عباس، العلق، الإدارة الرقمية: المجالات و التطبيقات، مركز الإمارات للدراسات و البحوث الإستراتيجية، ابوظبي، 2005، ص 84.

²Bazsalica M., Naim P., Data mining pour le Web, éd. Eyrolles, Paris, 2001, P. 61.

³ عبد الستار العلي، عامر إبراهيم قنديلجي، غسان العمري، المدخل إلى إدارة المعرفة، دار المسيرة للنشر و التوزيع و الطباعة، الطبعة الأولى، عمان، 2006، ص. 157.

⁴ Hand d., Mannila H., Smyth R., Principles of Data Mining, MIT Press, London, 2001, p. 01.

⁵ بشير عباس، العلق، مرجع سابق، ص. ص 89-91.
⁶ المرجع نفسه.

⁷ Bazsalica M., Naim P., Op. Cit., pp.68-69.

⁸ 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.

⁹ Bazsalica M., Naim P., Op. Cit., pp.66.

¹⁰ Michael, J., A. Berry and Gordan S. Linoff, (2000), Mastering Data Mining. John Wiley & Sons, Inc,

¹¹ عبد الستار العلي، عامر إبراهيم قنديلجي، غسان العمري، مرجع سابق، ص. ص 100-102.

¹² عبد الستار العلي، عامر إبراهيم قنديلجي، غسان العمري، مرجع سابق، ص. ص 98-99.