دور تنقيب البيانات Data Mining

في زيادة أداء المنظمة ((دراسة تحليلية في المصرف الصناعي))

أ. م. د. زكريا مطلك الدوري
جامعة بغداد/ كلية الادارة والاقتصاد

داليا عبد الحسين احمد
م.م./ الجامعة التكنولوجية

تمهيد

غالبا ما يكون تعامل المنظمات المالية والمصرفية مع الزبائن بشكل أساسي مما يتطلب منها جمع كميات هائلة من البيانات عن هؤلاء الزبائن هذا بالإضافة الى ما يرد اليها يوميا من بيانات يجعلها أمام أكداس كبيرة من البيانات تحتاج الى جهود جبارة تحسن التعامل معها والاستفادة منها بما يخدم المنظمة.

ان التعامل اليدوي مع مثل هذه البيانات دون استخدام تقنيات حديثة يبعد المنظمة عن التطور والارتقاء الى مستويات اداء افضل. ولا يكفي إدخال الحاسبات الى العمل فقط، بل من الأفضل استخدام تقنيات وبرمجيات تخدم الادارة وتقدم لها ما يمكن ان تستفيد منه دون اضاعة للوقت والجهد.

لذلك فان تنقيب البيانات Data Mining يقدم الحل لهكذا مشاكل فهي تكنولوجيا جديدة تهدف الى الحصول على معلومات غير معروفه من قواعد البيانات لغرض استخدامها في عملية اتخاذ القرار . انها تحسن من فهم البيانات المتراكمة واستخلاص ثروة المعلومات الموجودة فيها من اجل استثمارها بالشكل الذي يؤدي الى اتخاذ قرار سليم يجنب المنظمة أي خسائر محتملة ويرفع من مستوى ادائها .

المبحث الاول/ منهجية البحث

اولا" – مشكلة البحث

تتلخص مشكلة البحث في ثلاث نقاط أساسية هي: وجود عدد كبير من المقترضين في المصرف الصناعي ووجود كم هائل من البيانات عنهم وعدم وجود تقنية تمكن المصرف من اتخاذ قرار سريع .

كما يمكن صياغة بعض التساؤلات مثل:

1. ما مدى اهتمام المصرف في التوثيق الكامل للبيانات؟
2. كيف تستخدم الأساليب العلمية والتكنولوجية في التعامل مع هذه البيانات؟
3. هل تبنى اتخاذ القرارات على ما يستخلص من البيانات ام على الحدس؟
4. هل يجري تقييم هذه البيانات وإكمالها وتحديثها باستمرار؟
5. هل ان استخدام تنقيب البيانات يزيد من اداء المصرف؟

ثانيا" – أهمية البحث

1. تناول البحث لمتغيرين مهمين تنقيب البيانات DM والاداء يمكن ان يجعل البحث مساهمة متواضعة في هذا المجال خاصة بعد ما لوحظ من نقص كبير جدا في المصادر العربية الخاصة بالمتغير الاول والذي يعد من الموضوعات الحرجة والحديثة جدا" حيث يلاحظ ان اقدم مصدر في هذا الموضوع قد يعود الى عام 1996 كحد أدنى.
2. يساعد البحث المصرف على تجنب مخاطر حقيقية ناتجة عن عدم التسديد والتي تشكل عبئا" على المصرف لكونها قروضا" هالكة.
3. تقديم نماذج تسهل عملية تنقيب البيانات في المصرف حيث تتنبأ بالمقترضين وتصنفهم حسب مخاطر التسديد لكي يتمكن المصرف من التركيز على المقترضين الموثوق بهم فقط سعيا منه لزيادة الاداء.
4. يقدم البحث للمصرف قاعدة بيانات يفتقر اليها تخص القروض والمقترضين كما يقدم نظام متكامل لادارة قواعد البيانات يتمثل بالبرنامج الحاسوبي المستخدم.

ثالثا" – أهداف البحث

1. دراسة مشكلة تراكم البيانات ومحاولة إيجاد الحل الأمثل لها .
2. محاولة تحسين مستوى استخدام المعلومات من قبل الادارات من اجل زيادة الاداء.
3. بناء نماذج تنقيب البيانات التي يمكن ان تستخدم في التنبؤ بالمقترضين الموثوق بهم واستبعاد غير الموثوقين منهم لتقليل تكاليفهم.
4. محاولة إرساء أسس سليمة تكون البداية لانطلاق بحوث لاحقة تحاول التطبيق في مجالات اخرى جديدة.

رابعا" – فرضية البحث

للبحث فرضية أساسية هي " ان تطبيق النماذج الناتجة عن خوارزميتي شجرة القرار Decision Tree والتجميع Clustering سيمكن الادارة من تقليل خسائر المصرف "

خامسا" – منهج البحث

يعتمد البحث على تطبيق برنامج حاسوبي جاهز ومن ثم الوصف والتحليل لنتائج التطبيق من نماذج لاستخلاص المؤشرات الأساسية التي تدعم عملية اتخاذ القرار والرجوع مرة اخرى للتحليل للوقوف على امتدادات نتائج التطبيق.

سادسا" – مجتمع البحث

يعود اختيار المصرف الصناعي العراقي ليكون مجتمعا للبحث الى:

1. تخصص المصرف بمنح القروض الصناعية فضلا" عن الأنشطة الأخرى مما يوفر للبحث ما يحتاج اليه من بيانات تتعلق بالقروض.
2. أهمية المصرف في دعم الاقتصاد الوطني في مجالات تنمية وتطوير الصناعات والاستثمار في المشاريع من خلال فروعه المنتشرة .

سابعا" – عينة البحث

شملت عينة البحث 277 قرضا" من أصل 1530 قرضا" أي انها شكلت 18% من العدد الكلي وتشمل فروع المصرف الخمسة جميعها ولمختلف الصناعات والقروض ويمثل هذا العدد كل

ما أمكن الحصول عليه بسبب إتلاف الكثير من هذه البيانات وملفاتها بعد التقادم الزمني وانتهاء تسديد القروض.

ثامنا" – الحدود الزمنية للبحث

يركز البحث عند تطبيق البرنامج على القروض التي يمنحها المصرف على مدى 19 سنة للمدة 1985 - 2003 . اما تطبيق وتحليل مؤشرات اداء المصرف فيعتمد على بيانات خمس سنوات للمدة 1997 - 2001 لأغراض الحسابات والمقارنة.

تاسعا" – أسلوب جمع البيانات

اعتمد جمع البيانات بصورة أساسية على التدوين اليدوي لبيانات كل ملف من ملفات الاقتراض وسجلات ونشرات المصرف فضلا" عن المقابلات الشخصية وتوجيه الاسئلة الى المسؤولين في المصرف بهدف الحصول على اجابات تدعم التحليل.

عاشرا" – الوسائل المستخدمة في البحث

تم استخدام وسيلتان للتعامل مع البيانات:

الاولى: برنامج حاسوبي هو Microsoft SQL Server 2000 لتطبيق خوارزميتين هي شجرة القرار Decision Tree والتجميع Clustering والحصول على نماذج تنقيب البيانات .

الثانية: استخدام المؤشرات المالية بهدف الوصول الى ما يعكس اداء المصرف ومن خلال مؤشرات تحقيق الأهداف ومؤشرات النشاط ومؤشرات السيولة والربحية.

إحدى عشر – صعوبات البحث

واجهت مسيرة البحث صعوبات جمة منها :

1. اغلب ما كتب عن موضوع تنقيب البيانات هي كتابات نظرية لا تخوض في تفاصيل التطبيق العملي.
2. صعوبة تفهم موضوع تنقيب البيانات من قبل بعض الافراد مما يعيق توضيح أسباب طلب بعض البيانات.
3. تعقيدات الحصول على الموافقات الرسمية المطلوبة للحصول على البيانات.
4. ضعف استخدام الحاسوب في حفظ البيانات الخاصة بالاقتراض والمقترضين مما أوجب التدوين اليدوي لتفاصيل كل ملف من ملفات الاقتراض.
5. عدم تنظيم البيانات الخاصة بكل ملف من ملفات الاقتراض وتطبيق البرنامج يتطلب قاعدة بيانات كاملة وواضحة ودقيقة مما أوجب تكوين قاعدة بيانات جديدة تضم بيانات كل القروض.
6. تضارب بعض انواع البيانات مما تطلب المزيد من الوقت للتحقق منها .
7. اختلاف قيم مبالغ القروض من سنة لأخرى وبسبب امتداد البحث لمدة 19 سنة، استلزم استخدام اسلوب احصائي في تقسيم مبالغ قروض كل سنة الى 3 فئات بمستويات مختلفة من اجل توحيد مبالغ القروض على اساس واحد.

المبحث الثاني/ تنقيب البيانات Data Mining

اولاً- ماهية تنقيب البيانات Essence Of Data Mining

لابد هنا من التميز بين ثلاثة مصطلحات هامة هي البيانات Data والمعلومات Information والمعرفة Knowledge ، حيث يقصد بالأولى تلك البيانات المقترنة بالعمليات الوظيفية للمنظمة أي ما يتعلق بالأفراد والمواقع والصفقات وغيرها (Wu , 2000 :1) . اما المعلومات فهي بيانات متماسكة تعطي رسالة ما (Seiner, 2002: 2) في حين ان المعرفة تضم العامل البشري فهي استنتاج يرسم من المعلومات بعد ربطها مع معلومات اخرى ومقارنتها مع ما هو معروف سابقا". (Daft , 2001 :258)

تجمع المنظمات مقدارا" كبيرا" من البيانات، لكن الذي تريده حقا هو المعلومات، لذلك فأن التقنية الأحدث والأسرع للتعامل مع كم البيانات هذا والإجابة على أسئلة العمل هي تنقيب البيانات DM التي تشبه عملية تنقيب الذهب Gold Mining فتنقيب الذهب عملية غربلة ما بين كميات من المعدن الخام لإيجاد كتلة من معدن ثمين ذي قيمة، وكذلك عملية تنقيب البيانات التي هي عملية غربلة ما بين كميات من البيانات لإيجاد المعلومات المفيدة لاتخاذ القرار. (Noonan , 2000 : 6) ويرجع سبب زيادة الاهتمام بتنقيب البيانات الى اهتمامها بالاكتشاف والتحليل من قواعد بيانات كبيرة جدا تحوي قيما" خفية فضلا" عن مساعدتها في الكشف عن نماذج وقواعد ذات مغزى تساهم في تحسين فهم الزبون وحل العديد من مشاكل العمل (Soni & Tang & Yang , 2002 : 1) .

ومن أجل فهم أفضل لتنقيب البيانات Data Mining لا بد من التمييز بينها وبين اكتشاف المعرفة في قواعد البيانات (KDD) Knowledge Discovery in Databases فعرف Fayyad الأخيرة بأنها عملية غير عادية لتحديد نماذج صحيحة وجديدة وذات فائدة محتملة ويمكن فهمها بالنهاية. اما تنقيب البيانات فهي أحد خطوات عملية اكتشاف المعرفة وتتكون من تطبيق خوارزميات تحليل واكتشاف البيانات التي تقدم عدد خاص من النماذج عبر البيانات. (438 :(Houston & Others, 1999) (Zaiane,1999 : 3) .

ان انجاز تنقيب البيانات DM يمكن ان يكون من قبل ثلاث فئات من المستخدمين هم :
(Information Discovery,Inc. 2000 : 4) (Ramachandran,2001 : 2)

1- التنفيذيون Executives الذين يحتاجون الى ادراكات من مستوى عال ويقضون وقتا" اقل مع الحاسبات من المجموعات الاخرى.
2- المستخدمون النهائيون End Users وهم افراد البيع وباحثو السوق والعلماء والمهندسون والفيزيائيون وغيرهم.
3- المحللون Analysts الذين قد يكونون محللين ماليين او اجصائيين او استشاريين او مصممي قواعد بيانات .

لقد عرفت تنقيب البيانات بانها عملية تحليلية لتحويل البيانات الى معلومات عمل يمكن استخدامها لزيادة اداء المنظمة من خلال بناء نماذج رياضية تساعد المنظمات على اتخاذ قرارات عمل افضل (Saarenvirla , 2001: 1) لكنها ليست كأنظمة دعم القرار (DSS) Decision Support System التي تعمل على تحليل البيانات لتوليد المعلومات التي تساعد في حل المشاكل المكتشفة من قبل المستخدم أي انها رد فعل (Reactive) للتحديد الصحيح للمشكلة، أما تنقيب البيانات فهي اداة تهيؤ مسبق (Proactive) اذ تنقب ذاتيا" عن العلاقات المحتملة والشاذة وبذلك تحدد المشاكل التي لم تحدد بعد من قبل المستخدم. (Rob & Coronel, 2000:609) .

ان تنقيب البيانات DM هو اسلوب يمكن من خلاله الوصول الى المعلومات المخزونة في مستودع البيانات Data Warehouse (DW) ويتضمن استخدام التحليل الاحصائي لاكتشاف العلاقات الخفية في البيانات. (Romney & Steinbart ,2000:599) كما يعتبر تنقيب البيانات أحد تكنولوجيات الذكاء الاصطناعي (AI) Artificial Intelligence فضلا" عن الانظمة الخبيرة والشبكات العصبية ويهدف الى تمكين المنظمة من الاستغلال الأمثل لبياناتها فهي تحاول إيجاد المعلومات في مجاميع البيانات الكبيرة التي قد لا تعلم المنظمة بوجودها وكذلك إيجاد العلاقات وعمل التنبؤات. (Avison & Shah ,1997:327).

يتضح من ذلك ان تنقيب البيانات عبارة عن عملية استكشاف وتحليل كميات كبيرة من البيانات لغرض الحصول على علاقات ونماذج خفية تساعد في استخلاص المعلومات المفيدة والساندة لاتخاذ قرارات عمل استراتيجية كفيلة بزيادة اداء المنظمة.

ثانيا"-انواع تنقيب البيانات Types Of Data Mining

هناك العديد من وجهات النظر في تصنيف انواع تنقيب البيانات فمن وجهة النظر الموجهة بالعملية تم تحديد ثلاثة انواع هي :

(Information Discovery,Inc., 2000 : 2-3) (Ramachandran, 2001:1)

1- الاكتشاف Discovery

وهو عملية النظر في قاعدة البيانات لإيجاد النماذج من دون ان تكون هناك فكرة محددة عن ما قد تكون عليه .

2- النمذجة التنبؤية Predictive Modeling

فيه تستخدم النماذج المكتشفة من قاعدة البيانات للتنبؤ بالمستقبل أي تخمين القيم غير المعروفة بالاعتماد على نماذج سابقة مكتشفة من قاعدة البيانات .

3- التحليل المبرهن Forensic Analysis

وهو عملية تطبيق النماذج المستخلصة لإيجاد عناصر البيانات الشاذة او غير العادية أي انه يبحث في حالات محددة وغير عادية .

ومن وجهة نظر اخرى، هناك نوعين اساسيين من تنقيب البيانات هما :

(Ahola & Runsala , 2001:3)

1- التحليل الاستكشافي Exploratory Analysis

وهو فهم مجموعة البيانات لتكوين نماذج مهمة معقولة و جديدة .

2- التحليل التنبؤي Predictive Analysis

أي التنبؤ بقيمة المتغير الناتج في المستقبل بالاعتماد على أمثلة الماضي ويتم ذلك عن طريق تكوين نموذج تنبؤي لمتغير الهدف مع واحد او اكثر من المتغيرات الداخلة . ويعد النموذج التنبؤي حلقة الوصل بين ما هو معروف وما هو غير معروف فتتنبأ بحدث المستقبل كدالة لما هو معروف الان وتبنى باستخدام امثلة الماضي لحدث المستقبل وتكون اكثر فاعلية عندما تعتمد على مقادير كبيرة من البيانات النقية والموثوقة . (Lehman , 2001:7)

أما من حيث وجهة النظر المرتبطة بطبيعة نشاط تنقيب البيانات فتكون على ثلاثة انواع :
(Information Discovery,Inc. 2000 :5) (Ramachandran, 2001: 2)

التنقيب العارض Episodic Mining

أي النظر الى البيانات من عارض معين وينجز هذا النوع من قبل المحللين.

2- التنقيب الاستراتيجي Strategic Mining

وفيه يكون النظر الى مجموعات اكبر من البيانات الكلية للحصول على فهم شامل لمقاييس مثل الربحية.

3- التنقيب المستمر Continuous Mining

وهو يحاول فهم كيفية تغير العالم ضمن مدة محددة من الزمن ومحاولة فهم العوامل التي سببت التغير .

ثالثا" - تقنيات تنقيب البيانات Techniques Of Data Mining

تستخدم عملية تنقيب البيانات تقنيات عديدة تتمكن من خلالها اكتشاف الاتجاهات والنماذج الخفية في مقادير كبيرة من البيانات، ويمكن استخدام واحدة أو أكثر من هذه التقنيات وهي كالآتي:
(Brand & Gerritsen ,1998 :1-3) (Edelstein ,1997: 3)
(Ramachandran , 2001: 3-5) (Atre , 2001 : 2) (Tow Crows, 1999 : 6-15)

1- التصنيف Classification

فيه يتم تحليل مجموعة من البيانات لتكوين مجموعة من القواعد المتجمعة التي يمكن ان تستخدم لتصنيف بيانات المستقبل أي ايجاد المعلومات التي تتعلق بالخصائص المشتركة. وللتصنيف ادوات عديدة مثل شجرة القرار Decision Tree والمجاور الاقرب Nearest Neighbor والانحدار Regression .

2- الاقتران Association

وهي القاعدة التي تتضمن علاقات اقتران ثابتة بين مجموعة من الاشياء في قاعدة البيانات.أي الاقتران بين حدوث حدث ما وحدوث حدث اخر. وهي غالبا ما تسمى بتحليل سلة السوق Market Basket Analysis .

3- تحليل التتابع Sequential Analysis

وهو يشبه الاقتران ويوضع تحت مسمى تحليل الربط Link Analysis لكنه مرتبط بالزمن فيبحث عن نماذج تحدث بتتابع أي يتعامل مع البيانات التي تحدث في حالات منفصلة.

4- التجميع Clustering

وهي تقنية وصفية تجمع الكيانات المتشابهة سوية وتضع الكيانات غير المتشابهة في مجموعات مختلفة وتعتمد بصورة أساسية على قياس المسافة مثل استخدام تقنية المجاور الأقرب ومن ثم قد تكون هناك نتائج مختلفة لاثنين من منقبي البيانات الذين يعملون على نفس البيانات ويختلف التجميع عن التصنيف بانه لا تعرف ما ستكون عليه التجمعات عند البدء او بأية صفة ستتجمع البيانات. ويستخدم في التجميع ادوات مثل متوسطات K (K- Means) والشبكات العصبية Neural Networks .

ان الادوات التي تستخدمها هذه التقنيات عديدة ومتنوعة ولكل واحدة منها دور ويخدم غرض معين، ومن هذه الادوات: (Tow Crows , 1999 :10-15)

أ- اشجار القرار Decision Trees

وهي مشتقة من الاحصاء والذكاء الاصطناعي وتجد الارتباطات في البيانات وتستخدم في الاستدلال على قواعد العمل وتعد أساس بناء النموذج التنبؤي كما يمكن ان تستخدم اشجار القرار في تحديد المتغيرات الواجب استخدامها كمدخلات للشبكات العصبية.

ب- الشبكات العصبية Neural Networks

وهي قريبة من اشجار القرار لكنها اصعب فهما" منها ولا توضح نتائجها وتستغرق وقتا" أطول وتقدم نماذج ذات قوة تنبؤية افضل .

تتكون الشبكة العصبية من طبقات تبدأ بطبقة المدخلات Input Layer حيث متغير التنبؤ وترتبط عقد المدخلات بعدد من عقد الطبقة الخفية Hidden Layer التي ترتبط بدورها بعقد بعقد طبقة خفية اخرى او بطبقة المخرجات Output Layer حيث تضم واحد او اكثر من المتغيرات التابعة .

جـ- الانحدار Regression

يستخدم الانحدار في التنبؤ بالقيم الجديدة بالاعتماد على القيم الموجودة ويستخدم الانحدار الخطي للحالات البسيطة أما الحالات المعقدة التي يصعب التنبؤ بها فيستخدم الانحدار النسبي لانها تعتمد على تفاعلات معقدة لمتغيرات متعددة .

د - السلاسل الزمنية Time Series

وتتنبأ بالقيم المستقبلية غير المعروفة بالاعتماد على سلاسل تغير الزمن للمتنبآت فيؤخذ بالحسبان الخواص المميزة للزمن كتدرج المدد والموسمية مثلا".

هـ- استنتاج القاعدة Rule Induction

فيها يتم اشتقاق مجموعة من القواعد المستقلة وعلى خلاف اشجار القرار فهي لا تأتي من شجرة وقد لا تغطي القواعد المكونة كل الحالات الممكنة .كما انها قد تتعارض في تنبؤاتها .

و- المجاور الاقرب K – Nearest Neighbor

يستخدم المجاور الأقرب K (K – NN) فكرة ان حل المشاكل الجديدة يكون عن طريق ملاحظة حلول مشاكل مشابهة تم حلها مسبقا .

ز- التحليل التمييزي Discriminant Analysis

وهي اداة تصنيف تجد السطوح المتعددة التى تفصل الفئات ويكون النموذج الناتج سهل التغيير لان كل ما على المستخدم هو تحديد جانب الخط الذي تقع عليه النقطة.

ح- الاسناد Boosting

ويعتمد على اخذ عينات عشوائية متعددة من البيانات وبناء نموذج التصنيف لكل منها ويتغير وضع البناء بالاعتماد على نتيجة النماذج السابقة ويكون التصنيف الأخير هو الفئة الأكثر تخصيصا" من قبل النماذج .

ط- الخوارزميات الجينية Genetic Algorithms

سميت كذلك لانها تتبع نموذج نشوء الاحياء الذي يتنافس فيه اعضاء النشوء الواحد (من النماذج) لتتقدم خصائصها في النشوء اللاحق من النماذج الى ان يتم ايجاد النموذج الافضل.

رابعا" – عملية تنقيب البيانات Data Mining Process

هنالك العديد من الآراء التي تناولت عملية تنقيب البيانات D M تفصيلا" أم اختصارا" لكنها عموما" كانت تصب في نفس المضمون، لذلك سنأتي على ذكرها بالشكل الآتي :

(Brand & Gerritsen, 1998 :3) (Tow Crows , 1999 :22)

(Saarenvirta , 2001: 6) (Skalak , 2001 : 1)

1- تعريف مشكلة العمل Define business problem

يقصد هنا فهم العمل وبياناته والتحديد الواضح للاهداف الذي يقود الى بناء نماذج مختلفة جدا" .

2- بناء قاعدة بيانات تنقيب البيانات D M Build data mining database

ويكون ذلك من خلال جمع البيانات ووصفها وتنقيتها وتوحيدها واكمالها والمتابعة والمراقبة. وقد لا يفضل استخدام مستودع البيانات DW لهذا الغرض بل يفضل تكوين سوق البيانات Data Mart الذي يختص بجانب محدد. وتشكل هذه الخطوة مع الخطوتين اللاحقتين جوهر تهيئة البيانات التي تأخذ وقتا" وجهدا" اكثر من كل الخطوات الأخرى مجتمعة حيث تصل نسبتها 50 % - 90 % .

3- استكشاف البيانات Explore data

فيها يتم تحديد الحقول الأكثر اهمية في التنبؤ والقيم التي قد تكون مفيدة باستخدام تحليل الربط وغيرها من التقنيات .

4- تهيئة البيانات للنمذجة Prepare data for modeling

وهي الخطوة الاخيرة من تهيئة البيانات التي يتم فيها اختيار المتغيرات التي تكون افضل متنبآت واختيار الصفوف وايجاد متغيرات جديدة من البيانات الخام وكيفية تمثيلها باستخدام عمليات رياضية او جبرية ومن ثم اختيار الاداة التي من الممكن ان تكون أكثر فائدة في بناء النموذج .

5- بناء نموذج تنقيب البيانات Build model D M

يتطلب بناء النموذج اتخاذ القرار الخاص بنوع التنبؤ المراد صنعه (كالتصنيف والانحدار) ومن ثم اختيار نوع النموذج لصنع التنبؤ الذي يمكن ان يكون شجرة القرار او الشبكة العصبية او غيرها، ومن المهم ان يتم بناء نماذج عديدة بديلة لايجاد النموذج الأكثر منفعة في حل مشكلة العمل وقد يتطلب ذلك اجراء بعض التغييرات للبيانات. هذا مع ملاحظة وجوب ملائمة النماذج لابعاد متعددة وقدرتها على التعامل مع بيانات رقمية وغير رقمية.

ومن أجل ضمان دقة ومتانة التنبؤ لا بد من تمرين النموذج على نسبة من البيانات ومن ثم اختباره واثباته على ما تبقى من البيانات، فبعد ان يتم تكوين النموذج باستخدام قاعدة بيانات التمرين، يستخدم للتنبؤ بقاعدة بيانات الاختبار ويكون معدل الدقة الناتج تقدير جيد لكيفية اداء النموذج.

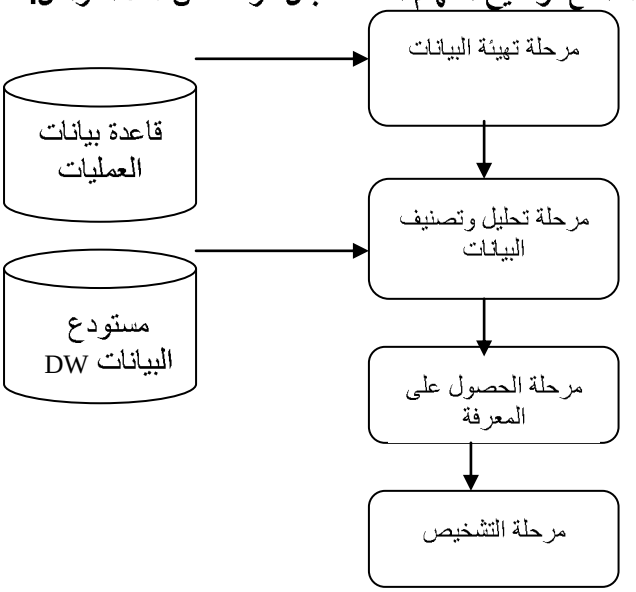
6- تقييم النموذج Evaluate model

مهما كان النموذج دقيق فليس هناك ما يضمن انه يعكس العالم الواقعي فقد تكون البيانات المستخدمة لبناء النموذج قد لا تلائم العالم الواقعي مما يقود الى نموذج غير صحيح. لذلك من المهم اختبار النموذج في العالم الواقعي أي تجربته على مجموعة صغيرة من الحالات لتقليل مخاطر النموذج غير الصحيح والتأكد من نتائج النموذج.

7- تنفيذ النموذج والنتائج Deploy model and results

يكون ذلك بأحد اسلوبين: أما توصية المحلل بالافعال معتمدا على الملاحظة البسيطة للنموذج ونتائجه أو تطبيق النموذج لمجموعات بيانات مختلفة، هذا مع استمرار رقابة الاداء التي قد تتطلب اعادة بناء النموذج ومن مدة الى اخرى.

ويمكن توضيح عملية تنقيب البيانات بالشكل (1) الذي يضعها بأربع مراحل تبدأ بتهيئة البيانات ومن ثم تحليلها وتصنيفها تليها مرحلة الحصول على المعرفة ومن ثم مرحلة التشخيص التي يتم فيها التنبؤ من خلال النموذج المبني. هذا مع توضيح المهام الخاصة بكل مرحلة من هذه المراحل.



- تحديد مجموعة البيانات
- تنقية مجموعة البيانات
- اكمال مجموعة البيانات
- تحليل التصنيف
- تحليل التجميع والتتابع
- تحليل الربط
- تحليل الاتجاه والانحراف
- اختيار الخوارزميات وتطبيقها
- الشبكات العصبية
- المنطق الاستنتاجي
- اشجار القرار
- شجرة التصنيف والانحدار
- المجاور الاقرب
- التكهن
- التنبؤ
- النمذجة

الشكل (1) مراحل تنقيب البيانات Data Mining
Source : Rob, Peter & Coronel, Carlos "Data Base Systems Design, Implementation and Management" Fourth Edition Course Technology, 2000, P.611 .

خامسا" – تطبيقات تنقيب البيانات Data Mining Applications

امتدت تطبيقات تنقيب البيانات الى مجالات عديدة وواسعة ولا زالت مستمرة في الدخول الى مجالات جديدة تضاف الى ما سبق وهي كالآتي :
(Avison & Shah , 1997 :328) (Ramachandran , 2001 : 3)
(Wu , 2002:2) (Tow Crows , 2002:1)

1- الأعمال المصرفية Banking : كتحليل مخاطر القروض
2 – المالية Financial : كتحديد الغش في تبادل الاسهم
3- الاتصالات Telecommunications : كتحديد الاستخدام المزيف لخدماتها
4- التسويق Marketing : كايجاد العلاقة بين الخصائص الديموغرافية للزبون
5- التامين ورعاية الصحة Insurance and Health Care : كتحليل الدعاوى
6 - الطب Medicine : كتحديد العلاجات الطبية الناجحة للأمراض المختلفة
7- النقل Transportation : كتحديد جدول التوزيع بين المنافذ
8- البيع بالتجزئة Retailing : كتقدير فاعلية الترويج
9- ادارة علاقات الزبون Customer Relationship Management : كاتخاذ الفعل اللازم للاحتفاظ بالزبون لانه اقل كلفة من كسب زبون جديد.
10- رقابة الجودة او تحليل الخطأ Quality Control or Error Analysis : كتحديد الخصائص المحيطة بالمنتجات المتضررة ومن ثم تحسين الجودة من خلال اجراء التغيرات.
11- الاستخدام Hiring : كتحديد خصائص أفراد الأداء العالي
12- التجارة الإلكترونية Electronic Commerce
13- تحليل قائمة الخدمة في المطاعم Food – Service Menu Analysis
14- تحليل الضمان Warranty Analysis
15- التعليم لاستقطاب واستبقاء الطالب Student Recruiting and Retention

سادسا" – استراتيجيات نجاح تنقيب البيانات Strategies of Data Mining Success

لكي يتم الحصول على تنقيب البيانات ناجح، لا بد من اتباع عدد من الاستراتيجيات التي تضمن ذلك ومنها:

1- توجيه العمل نحو هدف محدد يوضح المشاكل المراد حلها (Noonan , 2000 :4) (Skalak , 2001:2)
2- توفير بيانات ذات كمية ومحتوى عال يزيد من قوة التنبؤ . (Hermiz , 1999:3) (Small , 1997:6 فكلما كانت البيانات افضل، كانت النتائج افضل أي ذات تحليل شامل واكثر دقة.
3- بناء عدة نماذج باستخدام عينات من البيانات بدلا" من بناء نموذج واحد باستخدم كل البيانات (Small , 1997:7) .
4- التهيئة الصحيحة للبيانات (Skalak , 2001:2) والاسراع في استخدام تنقيب البيانات (Noonan , 2000:3) تساهم بشكل كبير في انجاح تنقيب البيانات .
5- التواصل المستمر بين مدراء العمل والخبراء مع امكانية احتواء متخذ القرار (White , Cross 2001:5) من أجل اعادة تعريف او تصميم اهداف العمل وايجاد النتائج الاولية وتكرار ذلك مع ازدياد ادراك البيانات

6- التخطيط للتعلم من عملية تنقيب البيانات وتعليمها الى من له صلة مباشرة باستخدامها في اتخاذ القرارات . (Hermiz , 1999:4) وعدم اهمال أي نتائج تحت ضغط الوقت والكلفة (Skalak , 2001:2).
7- التكامل بين عدد من الادوات بضمنها القديمة والجديدة . (Smyth,2001:5)

المبحث الثالث/ الأداء *Performance*

اولا" – مفهوم الاداء

يمثل الاداء معيار الرشد في استخدام الموارد البشرية والمادية والمعلومات المتاحة بالشكل الذي يحقق أعلى المردودات منها وكذلك اشباع حاجات ورغبات الافراد العاملين ورفع الروح المعنوية لديهم ليعزز رغبتهم واندفاعهم للعمل. (الشماع وحمود، 2000 : 330)

وقد يتداخل مع مفهوم الاداء او تقويم الاداء مفاهيم اخرى مثل:

1- تقييم المشاريع او تقييم الجدوى الاقتصادية التي يقصد بها إقامة دراسة من نوع خاص يمكن من خلالها تكوين صورة متكاملة عن ما سيكون عليه المشروع في المستقبل خلال حياته المتوقعة بحيث يمكن اتخاذ قرار بإقامة المشروع من عدمه مع امكانية ترتيب المشاريع حسب الاولويات. (عبد الكريم وكداوي، 1999 : 211)
2- مفهوم الانتاجية productivity التي يقصد بها كمية الانتاج منسوبة لعنصر او اكثر من عناصر الانتاج ، لكنها لا تكفي في اعطاء حكم متكامل على اداء المنظمة بسبب وجود جوانب اخرى من نشاط المنظمة لم تشملها. (الكرخي، 2001 : 42).
3- الفاعلية Effectiveness وتعني تحقيق المدراء لأهدافهم المنظمية، فالاداء لا يكفي لوحده بل هناك حاجة الى تحقيق الاهداف من خلال الانشطة. ان الاداء يتعلق بوسائل عمل الاشياء في حين ان الفاعلية تتعلق بالنهايات او نيل الاهداف المنظمية. (Robbins & Coulter , 1999:9)

يعد اداء المصرف دافع قوي ونتيجة مهمة لإجراء التغييرات في العمليات المصرفية وغالبا ما تضع المصارف الاداء العالي هدفا لها كحافز او دافع أساسي لتثبيت موقعها والذي يكون من خلال تقديم مستوى من الخدمات يعظم الارباح بأقل كلفة ممكنة. (Saunders, 1997:264) وطالما كانت المعلومات هي أحد أهم الموارد بالنسبة للقطاع الخدمي بصورة عامة وللمصارف بصورة خاصة فان حسن استغلالها بصورة مثلى من الممكن ان يؤدي الى مستويات متقدمة من الاداء التي يسعى اليها المصرف .

ثانيا" – أسس الاداء

ان دراسة الاداء هي عملية شاملة ومتكاملة بطبيعتها نظرا لتعدد الانشطة والعمليات في المنظمة والترابط الوثيق بينهما، لذلك فان التعرف على اسباب التغير في الاداء الكلي وتحليله والحكم عليه يتطلب التعرف على الاسس او القواعد الاساسية التي يستند عليها الاداء وهي كالآتي: (عبد الكريم و كداوي، 1999: 208-211) و (الكرخي ،2001:48-50)

1- تحديد الاهداف
أي تحديد الاهداف الرئيسة والثانوية للمنظمة والتي يجب ان تكون واضحة ومحددة لكل من المسؤولين والعاملين ودقيقة وشاملة لكل أجزاء المنظمة .
2- تحديد الخطط التفصيلية لانجاز الفعاليات
يجب ان تكون هذه الخطط واقعية ومتناسقة مع طبيعة الاهداف فضلا" عن مرونتها.
3- تحديد مراكز المسؤولية

أي تحديد الجهة المسؤولة عن اتخاذ القرارات الكفيلة بتنفيذ هذا النشاط في حدود الموارد الموضوعة .

4- تحديد مؤشرات الاداء

وهي من أهم الأسس وأكثرها صعوبة في نفس الوقت حيث يتم اختيار المؤشر المناسب باختلاف الاهداف المرسومة وطبيعة العملية الأنتاجية.

ثالثا" – مراحل قياس الاداء

تمر عملية قياس الاداء بمراحل عدة وهي كالآتي :

(حمد ، 1986: 77) (Rambaldi & Bautista,2000:14) (الكرخي، 2001: 50-52)

1- جمع البيانات والمعلومات الاحصائية التي تتطلبها عملية القياس ويمكن الحصول عليها من حسابات الانتاج والارباح والخسائر والميزانية العمومية وغيرها من المعلومات المتوفرة للسنة الحالية والسنوات السابقة وفي منظمات متشابهة في نفس وغيره لأغراض المقارنة.
2- تحليل ودراسة البيانات والمعلومات الاحصائية لحساب المؤشرات اللازمة لعملية قياس الاداء.
3- اجراء عملية القياس لجميع انشطة المنظمة باستخدام المؤشرات المالية الملائمة.
4- وضع النتائج وتحديد الانحرافات واسبابها والحلول اللازمة لها.
5- متابعة العمليات التصحيحية للانحرافات وتزويد الجهات المسؤولة بالمعلومات والبيانات اللازمة للاستفادة منها مستقبلا".

رابعا" – مؤشرات الاداء

المؤشر هو أي وسيلة للقياس يمكن الاستعانة بها في اتخاذ قرار حكم موضوعي على حالة معينة وقد يأخذ شكل قاعدة قانونية او اقتصادية او اجتماعية او سياسية او رياضية. (الكرخي، 76:2001) أما مؤشرات الاداء فهي تمثل اشارات لتحذير المدراء نحو القضايا التي تحتاج الى اهتمام اكثر. (Kunstelj & Others,2001:10)

وهناك الكثير من المؤشرات التي يمكن استخدامها لقياس الاداء ، منها ما هو موضح في الملحق (1) الذي يبين اسم المؤشر وطريقة احتسابه. لكن التوصل الى رقم معين لا يعني شيئا بحد ذاته بل يتطلب مقارنته بغيره من الارقام كأن تكون: (خان وغرايبة، 77 1995:) (شاكر وآخرون،2000: 54-79)

1- المؤشرات التاريخية Historical Standard

أي دراسة النسب لنفس المنظمة في السنوات السابقة.وتعود أهمية هذه المعايير الى إعطائها صورة عن اتجاه المنظمة ووضعها فيما اذا كان يسير بتحسن مستمر او يتجه الى الأسوأ. لكن بالمقابل لا يمكن مقارنتها المنظمة مع منظمات اخرى وهي غير دقيقة اذا توسعت المنظمة او تغيرت طبيعة عملها .

2- مؤشرات الصناعة Industry Standard

أي مقارنة نسب المنظمة مع منظمات اخرى مساوية لها بالحجم وطبيعة العمل لكنها غير دقيقة لصعوبة وجود منظمات متشابهة تماما من حيث الحجم وطبيعة الاعمال.

لذلك فأن من المفضل ان يتم استخدام المؤشرين للحصول على صورة شاملة لأوضاع المنظمة.

وفيما يخص القطاع المصرفي فقد استخدم الباحثون ارباح المصرف في قياس الاداء حيث يتطلب الاداء الكامل من المصرف تحقيق ارباح عالية من الخدمات وتقليل كلفها. (Saunders , 1997:264)

وبذلك يتضح من ذلك ان المهمة المرماة على عاتق الادارة ليست باليسيرة فهي لا تتوقف عند قياس الاداء بل استخدام هذا القياس بصورة موضوعية ومقارنته بالسنوات السابقة تارة ومع

المنظمات الاخرى المماثلة تارة اخرى لمعرفة موقعها بين المنظمات وما تحتاج اليه من تطوير من اجل العمل به مستقبلا من اجل غد افضل.

المبحث الرابع/ تحليل النتائج

اولا" – التطبيق

تم اجراء عملية تنقيب البيانات على بيانات القروض للمصرف الصناعي العراقي لغرض مساعدة الادارة على اتخاذ القرار الذي يمكنها من الحكم على الزبون فيما اذا كان التعامل معه يشكل خطرا" على المصرف أم لا. وبذلك تتمكن من التركيز على الزبائن الموثوق بهم واستبعاد من هو غير موثوق به لأنه يكون السبب فيما يتحمله المصرف من خسائر ناجمة عن اقراضه بسب عدم التسديد.

ومن اجل التطبيق استخدم برنامج حاسوبي يدعى خادم لغة الاستفسار المهيكلة SQL Server 2000 (Structured Query Language Server 2000) الذي يمكن تعريفه بانه نظام ادارة قاعدة البيانات Database Management System الذي تم تصميمه للمساعدة في ادارة بيانات المنظمة فهو يقدم طرائق لمسح وتحديث وادخال البيانات التي يمكن من خلالها الحفاظ على قاعدة بيانات متجددة. (Gunderloy & Jorden,2000:263)

يقدم البرنامج خوارزميتين أساسيتين لتحليل البيانات هما :

(Tiedrich, 2000:14) (Soni & Others, 2001:2) (Bloor Research, 2001:102)

1- شجرة القرار Microsoft Decision Tree (MDT)
2- التجميع Microsoft Clustering

ان زيادة عدد المقترضين في المصرف، يجعل من الصعب على الادارة ان تستخلص القواعد التي تحدد المخاطر الجيدة والسيئة يدويا، لذلك فان زيادة الاهتمام بتنقيب البيانات Data Mining ناتج عن كونها تهتم بالاكتشاف والتحليل عن طريق وسائل مؤتمتة أو شبة مؤتمتة.

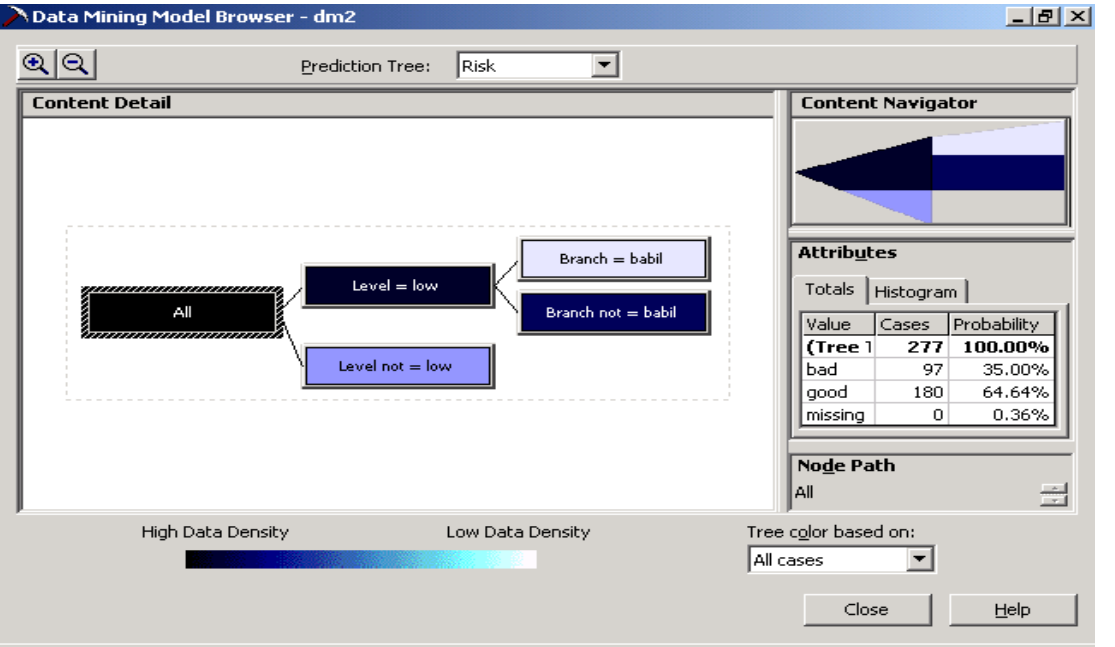
ومن تصفح اشجار القرار والتجمعات الناتجة من نماذج تنقيب البيانات من قبل المستخدم، تسمح له بإلقاء نظرة سريعة على ناتج عملية التنقيب وبالتالي يتمكن من اتخاذ قرار سريع وحكيم مستند على نماذج واضحة ومفهومة مقرونة بتفصيلات بيانية مع النسب.

ثانيا" – تحليل نماذج شجرة القرار

يستخدم نموذج شجرة القرار بصورة أساسية لغرض التنبؤ ويمكن ان يساعد ايضا" في توضيح كيفية توزيع البيانات (Seidman, 2001: 114) أي انه يسهل رسم استنتاجات عامة حول مجتمع البيانات.

وبعد اختيار الخطر Risk أي خطر القرض كمتغير للتنبؤ، ينتج الشكل (2) الذي تظهر فيه شجرة القرار وقد وجدت ان مستوى القرض Level هو الصفة الاكثر أهمية في التنبؤ وتفرع الى فرعين هو واطئ Low وغير واطئ وكانت الحالات أكثر كثافة عند المستوى الواطئ نسبة الى اللون الغامق، ومن ثم وجدت ان الفرع Branch هو الصفة الاكثر أهمية الثانية وقسمته الى فرعين ايضا اما ان يكون فرع بابل Babil او لا يكون كذلك وكانت الحالات أكثر كثافة عندما لا يكون الفرع بابل.

كما يبين الشكل في الجهة اليمنى العليا ملاح المحتوى Content Navigator الذي يزودنا بنظرة عامة عن شكل الشجرة وباستخدام تدرج الالوان كما هو الحال في شجرة القرار حيث يدل اللون الغامق على مدى كثافة الحالات في العقد أي ان العقدة التي لونها غامق تضم أعلى نسبة من الحالات المطابقة لقيمة الصفة وهكذا مع تدرج اللون الذي يبينه المستطيل الصغير في أسفل يسار الشاشة.



الشكل (2) نموذج شجرة القرار على اساس خطر القرض Risk

كما يبين الشكل ايضا" لوحة الصفات Attributes التي تقع في وسط يمين الشاشة التي تعكس بيانات كل عقدة من خلال الضغط على أي من عقد الشجرة، وفيما يخص الشكل (2) فتبين لوحة الصفات بيانات الشجرة ككل وفيها عدد حالات الخطر السيئة bad 97 حالة بنسبة احتمالية 35 % وعدد حالات الخطر الجيدة 180 حالة وبنسبة احتمالية 64.64 % ولا توجد هناك حالات مفقودة حتى ان احتماليتها 0.36 %.

اما فيما يتعلق بمسار العقد Node Path الذي يقع تحت لوحة الصفات فيعرض وصف للقواعد التي تؤثر في تضمين الحالات داخل العقدة. وهو في هذه الشجرة يشمل كل القواعد.

وبذلك يكون قد نتج عن استخدام خوارزمية شجرة القرار نموذج تنبؤي يتنبأ بمخاطر القروض فيما اذا كانت سيئة أم جيدة .

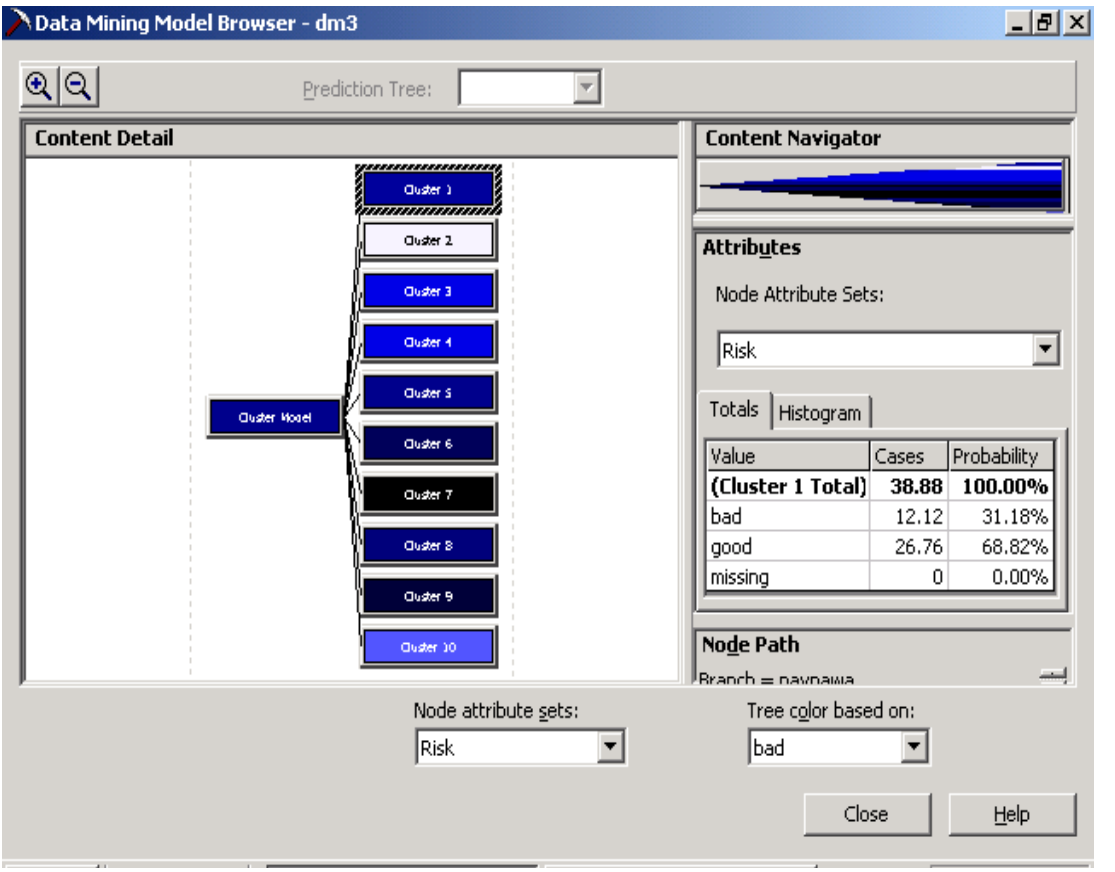
ثالثا" – تحليل نموذج التجميع

تعد خوارزمية التجميع مفيدة جدا" في ايجاد القيود التي تمتلك صفات مشتركة مع غيرها، وهي افضل خوارزمية عندما تكون هناك كمية كبيرة جدا من البيانات ذات درجة عالية من التركيب المنطقي والعديد من المتغيرات فهي تسمح بتصور العلاقات وتوضيح الشواذ وايجاد عينات لخوارزميات اخرى في تنقيب البيانات . (Seidman, 2001: 146)

يتم تمثيل التجمعات بنفس اسلوب شجرة القرار تماما عدا ان الشجرة هنا تكون دائما بمستوى واحد. ولاختيار صفة معينة كصفة الخطر، يكون ذلك من قائمة مجموعة صفات العقدة Node Attribute Set في وسط أسفل الشاشة. اما أسماء العقد فهي غير واضحة (Cluster 1، Cluster 2 ، ..... وهكذا) لكن لكل تجمع من هذه التجمعات مجموعة من القواعد التي تصف القيود الموجودة داخل التجمع، ويمكن رؤية هذه القواعد من خلال الضغط على العقدة والنظر الى محتويات مسار العقدة Node path في اسفل يمين الشاشة.

وباختيارنا لصفة الخطر Risk والضغط على التجمع الاول Cluster 1 واختيار الحالات السيئة bad لتبنى على اساسها لون التجمعات، يكون الشكل (3) حيث يتضح ان عدد الحالات في هذا التجمع ذو اللون الغامق هو 39 حالة منها 12 حالة ذات مخاطر قرض سيئة bad بنسبة احتمالية 31.18 % و 27 حالة ذات مخاطر قرض جيدة good بنسبة احتمالية 68.82 % و 0 حالة مفقودة بنسبة احتمالية 0 % . اما القواعد التي تصف التجمع الاول في Node path فهي ان مخاطر القروض جيدة لفروع المصرف في نينوى و بغداد والبصرة ولصناعات هي الجلود والثلج والملح والغذائية والنجارة والطحين والبلاستك والكيمياوية والجنس هو ذكر ومستوى القرض واطئ واصحاب القروض هم افراد منهم متزوجون واخرون غير متزوجين وهكذا بالنسبة لبقية التجمعات التي تكون مرتبة تنازليا" من حيث عدد القيود التي تضمها .

وبذلك يكون قد نتج عن استخدام خوارزمية التجميع نموذج وصفي يضع المقترضين في مجموعات حسب مخاطر القروض .



الشكل (3) نموذج التجميع على اساس خطر القرض Risk

رابعا" – تحليل مؤشرات الاداء

على الرغم من وجود العديد من المؤشرات التي تستخدم في قياس الاداء، الا ان المصارف ومنها المصرف الصناعي تركز على عدد منها معتمدة محليا" ودوليا" منها :

1- مؤشرات تحقيق الاهداف وتضم :
   أ- مؤشر تحقيق خطة الائتمان
   ب- مؤشر تحقيق خطة الاستثمار
2- مؤشرات النشاط وتضم :
   أ- مؤشر كفاية رأس المال
   ب- مؤشر الائتمان النقدي/ الحسابات الجارية والودائع
   ج- مؤشر الاستثمار/ الحسابات الجارية والودائع
3- مؤشرات السيولة والربحية وتضم :
   أ- مؤشر النقد في الصندوق ولدى المصارف/ الحسابات الجارية والودائع
   ب- مؤشر صافي الربح/ اجمالي الموجودات (عدى الائتمان التعهدي)
   ج- مؤشر صافي الربح/ رأس المال

ان جميع المؤشرات السابقة الذكر يمكن ان تؤشر مدى قدرة الادارة في الحفاظ على نمو العوائد التي تأتي أساسا" من قروض المصرف واستثماراته واجرة خدماته. (Rose, 1999:159) وقد جرى تطبيقها في المصرف الصناعي للسنوات الممتدة من 1997 الى 2001 .

كما قد لوحظ ان نسب تسديد القروض متدنية جدا" حيث ان أعلى نسبة تسديد للقروض كانت 50.7% في سنة 2001 هذا يعني ان نصف المبالغ المستحقة تقريبا اصبحت مصاريف يتحملها المصرف فكيف هو الحال مع اقل نسبة 18.8% في سنة 2000 .

ان اجراء عملية تنقيب البيانات على بيانات الزبائن والوصول الى نماذج تصف جميع حالات الزبائن ونماذج اخرى تتنبأ بمخاطر القروض استنادا" على حالات زبائن مختلفة يساعد كثيرا" في الحكم على طلبات الاقتراض واستبعاد ما تنطبق عليها النماذج كونه زبونا" خطر التعامل معه أي ان تسديده غير موثوق به بل انه يحمل المصرف أعباء ثقيلة، وباستبعاد طلبات الاقتراض هذه والتركيز على مقترضين موثوق بهم يمنح المصرف فرصة أكيدة في الارتفاع بنسب التسديد وتقليل المصاريف التي تشكل عبئا" على المصرف وتعيق الحفاظ على نمو عوائده.

وبذلك تكون فرضية البحث الأساسية قد تحققت وهي ان تطبيق النماذج الناتجة عن خوارزميتي شجرة القرار والتجميع سيمكن الادارة من تقليل خسائر المصرف .

من هنا يمكننا الاجابة على كل التساؤلات الواردة حيث ان المصرف لم يهتم الاهتمام الكافي بتوثيق البيانات فهي تفتقر الى الكثير من البيانات التي يمكن الاستفادة منها في تكوين قاعدة بيانات متكاملة تمتد لفترات زمنية طويلة. كما أن المصرف لم يتعامل مع هذه البيانات باساليب علمية وتكنولوجية فهي مركونة لحين الحاجة اليها بدلآ من استخدامها في بناء القرارات ولا يجري تقييمها لغرض اكمالها وتحديثها باستمرار. وعلى العكس من ذلك فأن استخدام تنقيب البيانات يوفر للمصرف الفرص المناسبة لاتخاذ قرارات بمستوى عالي من الدقة تزيد من اداء المصرف وتجنبه المخاطر الناتجة عن القرارات الخاطئة التي تؤدي الى تدني نسب تسديد القروض وبالتالي زيادة الاعباء المالية على المصرف التي يتحملها مصاريف.

أن النماذج الناتجة عن استخدام خوارزميتي شجرة القرار والتجميع تلخص لادارة المصرف بشكل واضح مجتمع البيانات وتسهل رسم الاستنتاجات منه ،فالخوارزمية الاولى بينت ان على الادارة ان تعيد النظر بالقروض ذات المستوى الواطيء ولفروع المصرف عدا فرع بابل لان هذه القروض تشكل نسبة الخطر الاكبر كقروض غير مسددة أي أن على الادارة ان تكون حذرة اكثر تجاه هؤلاء المقترضين وتحاول التقليل منهم والتركيز على القروض ذات المستوى الاعلى .في حين صنفت الخوارزمية الثانية المقترضين في مجموعات وبشكل تنازلي وحسب تدرج اللون من الغامق الى الفاتح بحيث يسهل التعرف على المجموعات بنظرة واحدة ومعرفة خصائصها كالتجمع الاول الذي ضم اكبر عدد من القروض غير المسددة في فروع نينوى وبغداد والبصرة ولصناعات الجلود والثلج والملح والغذائية والنجارة والطحين وللمستويات الواطئة من القروض أي يتوجب على الادارة الحذر والتشديد على المقترضين الذين يحملون هذه الخصائص لان قرارت الاقراض الصائبة ستبعد المصرف عن الخسائر الناتجة عن عدم التسديد وتساهم بشكل اساسي في نمو عوائده .

المبحث الخامس/ الاستنتاجات والتوصيات

اولا"- الاستنتاجات المتعلقة بالجانب النظري

1. توفر عملية تنقيب البيانات ميزة التعامل مع كميات هائلة من البيانات وكلما كانت كمية البيانات كبيرة كانت اكثر قدرة ودقة في الوصول الى نماذج صحيحة مفيدة.
2. يختلف تنقيب البيانات عن أنظمة دعم القرار كونه قد يكشف عن علاقات ابعد من ان يفكر بها المستخدم.
3. تعتمد نتائج تنقيب البيانات بصورة اساسية على مدى دقة وصحة وشمولية البيانات المستخدمة.
4. تستخدم في تنقيب البيانات تقنيات وادوات عديدة ومتنوعة لكنها تصب في اتجاهين أساسيين هما الاستكشاف والتنبؤ.
5. لا يمكن تحديد خوارزمية معينة باعتبارها الافضل لبناء نموذج تنقيب البيانات بل يفضل ان تستخدم خوارزميات متعددة ويتم بناء نماذج متعددة الى ان يتم الوصول الى النموذج الافضل.
6. في حالة وجود قاعدة بيانات كبيرة جدا، يكون خيار بناء نماذج عديدة مبنية على عينات افضل من خيار بناء نماذج قليلة مبنية على كل البيانات لكي يكون النموذج اكثر دقة ومتانة.
7. تتطلب عملية قياس الاداء اختيار مؤشرات واضحة سهلة الاستخدام وقليلة الوقت والجهد تتلاءم مع طبيعة عمل المصارف.

ثانيا"- الاستنتاجات المتعلقة بالجانب التطبيقي

1. كلما كان عدد المتغيرات او الاعمدة في قواعد البيانات اكثر كان نموذج التنقيب الناتج اكثر واقعية في التعامل مع مشاكل العمل.
2. يوفر برنامج SQL Server 2000 سهولة تنقيب البيانات من مصادر بيانات علائقية ومتعددة الابعاد ومن خلال خوارزميتين هما شجرة القرار والتجميع.
3. يمكن اجراء عملية التنقيب على بيانات تضم متغيرات رقمية وغير رقمية وتظهر في النموذج بالضبط كما صيغتها في قاعدة البيانات مثل مستوى القرض الذي يحمل تقديرات مثل واطىء ومتوسط وعالي.
4. تكون النماذج الناتجة عن عملية التنقيب وخاصة من البرنامج المستخدم سهلة الفهم حيث يستطيع المستخدم مثل المحلل المالي ان يفسر النموذج ويقارنه مع افكاره الخاصة. ان سبب ذلك يعود الى ان البرنامج يعمل على اختبار نماذج عديدة ويقدم في النهاية النموذج الافضل.
5. يوفر تطبيق البرنامج عامل السرعة حيث يمكن الوصول الى نموذج التنقيب في غضون دقيقة او اثنتين كحد أقصى مع السماح بتحديث هذه النماذج باستمرار. ان هذا التحديث والسرعة يسهل على ادارة المصرف اتخاذ القرار الخاص بالتعامل مع طالبي القروض بالرفض او القبول.
6. نتج عن استخدام خوارزمية شجرة القرار نموذج تنبؤي حدد فيه مستوى القرض **Level** الصفة الأكثر أهمية في التنبؤ بمخاطر القروض ويكون مستوى القرض هنا اما واطئا" او غير واطئ. واختارت الخوارزمية بعد ذلك الفرع **Branch** الصفة الأكثر أهمية الثانية وهو اما ان يكون فرع المصرف في بابل او لا يكون كذلك. ويصاحب هذا النموذج وصف واضح ومبسط لكل اجزائه.
7. ينتج عن استخدام خوارزمية التجميع نموذج وصفي حدد عشرة تجمعات مرتبة تنازليا حسب عدد القيود من التجمع الاول وحتى التجمع العاشر، حيث يضم مثلا التجمع السابع اكبر عدد من الحالات السيئة للقروض تصل الى 28 حالة خطر سيئة مقابل 2 حالة خطرة جيدة فقط

مع نسب احتمالية كل منها.وتنحصر مخاطر القروض السيئة في فروع المصرف في اربيل والسليمانية والمستثمرة في الصناعات الانشائية والميكانيكية ومحطات تعبئة الوقود والنجارة وتكون مستويات القروض متوسط وعالية والمقترضين هم افراد من كلا الجنسين ذكور واناث.

8. ثبوت صحة الفرضية الاساسية حيث ان تطبيق النماذج الناتجة عن خوارزميتي شجرة القرار والتجميع ، يمكن الادارة من رفع نسبة ما مسدد الى ما هو مستحق من قروض ويؤدي ذلك الى تقليل الكلف وزيادة صافي الربح وهو ما يظهر جليا في معيار صافي الربح / رأس المال حيث حقق المصرف اعلى المعدلات في سنة 2001 ووصل الى 699% تقابلها نسبة تسديد عالية نوعا ما مقارنة ببقية السنوات ووصلت الى 50.7 %.

ثالثا" – التوصيات الخاصة بالمصرف الصناعي

1. ضرورة توفير عناية اكبر بعملية جمع البيانات من زبائن المصرف وادخالها على الحاسوب الالكتروني بدلا من تركها في ملفاتها وما يتطلبه ذلك من توفير اجهزة الحاسوب الحديثة. ان عمل ذلك يسهل على المصرف الاستفادة من هذه البيانات بشكل عملي ويوفر الوقت لأي تطبيقات لاحقة.
2. لا تقتصر فائدة اجهزة الحاسوب على ادخال بيانات الزبائن فقط ، بل يتعداه الى تكوين قاعدة بيانات خاصة بزبائن المصرف ككل او قاعدة بيانات خاصة بالقروض والمقترضين او التسهيلات المصرفية، ومن الافضل ان يكون ذلك من خلال تكوين مستودع البيانات (DW) الذي يعمل على حفظ البيانات وتنقيتها وتحديثها باستمرار لضمان استرجاع أي جزء منها في أي وقت.
3. العمل على زيادة نشر الوعي في المصرف وخاصة في المستويات التنفيذية لتقنية تنقيب البيانات توضيح مدى الحاجة الملحة لاتخاذ القرارات المستندة على المعرفة لما لذلك من دور كبير في زيادة اداء المصرف من خلال تقليل التكاليف الناجمة عن عدم تسديد المقترضين وارتفاع عدد القروض الهالكة.
4. التوسع في حجم ونوع البيانات المطلوبة من الزبائن من اجل إضافة متغيرات جديدة لعملية التنقيب كمستوى دخل المقترض مثلا سواء كان شهريا" أم سنويا" وبشكل تقريبي وعدد الاطفال ان وجدوا.
5. التركيز على تدريب وتطوير متخذي القرار على مثل هكذا تقنيات حديثة لكي يسهل عليهم استخدام أو تطبيق البرنامج والتوصل الى المعلومات المطلوبة في اتخاذ قراراتهم دون الحاجة الى انتظار وصول المعلومات من جهة اخرى مما يضيف ميزة اخرى هي الحفاظ على سرية المعلومات الخاصة بالزبائن واقتصارها على اشخاص محددين.
6. تقديم الدعم المالي المستمر لكل ما من شأنه ان يبقي المصرف مواكبا للتطورات التقنية الحديثة من أجهزة وبرامج من اجل الوصول الى المعلومات المفيدة التي تعتبر من الموارد المهمة للمصرف التي يتوجب استغلالها بالشكل الأمثل كبقية الموارد.

ثانيا" – التوصيات الخاصة بالدراسات المستقبلية

1. اجراء دراسة مماثلة وفي المصرف ايضا لكن على الجانب الاخر من الائتمان النقدي وهو التسهيلات المصرفية.
2. اجراء دراسة مماثلة في قطاعات اخرى مثل شركات التامين حيث يمكن اجراء تنقيب البيانات في مجال تحليل الدعاوى التي تعتبر ارضية خصبة لمثل هكذا دراسات.
3. تناول تقنيات اخرى حديثة جدا" ذات ادوات تحليلية أعمق وأكثر توسعا" من تنقيب البيانات مثل Insightful Miner (I – Miner) .

المصادر References

اولا" – العربية

1. الشماع، خليل محمد حسن وحمود، خضير كاظم، "نظرية المنظمة" الطبعة الاولى، دار المسيرة للنشر والتوزيع، الاردن، 2000.
2. الكرخي، مجيد عبد جعفر، "مدخل الى تقويم الاداء في الوحدات الاقتصادية" الطبعة الاولى، دارالشؤون الثقافية العامة، العراق، 2001.
3. حمد، موفق خزعل "تقويم الاداء الصناعي لمنشاة القادسية العامة للصناعات الكهربائية للفترة 1979-1983 " رسالة ماجستير، 1986.
4. خان، محمد يونس وغرايبة، هشام صالح، "الادارة المالية" الطبعة الثانية، مركز الكتب الاردني، 1995.
5. شاكر، منير و اسماعيل، اسماعيل ونور، عبد الناصر "التحليل المالي – مدخل صناعة القرارات" الطبعة الاولى، عمان، 2000 .
6. عبد الكريم، عبد العزيز مصطفى وكداوي، طلال محمود "تقييم المشاريع الاقتصادية دراسة في تحليل الجدوى الاقتصادية وكفاءة الاداء" الطبعة الثانية، دار الكتب للطباعة والنشر، جامعة الموصل،1999.

ثانيا" – الانكليزية

أ – المجلات Journals

1. Ahola, Jussi & Rinta-Runsala, Esa: Data Mining Case Studies in Customer Profiling, “Research Report” TTE1-2001-29, Version 1.0, 2001.
2. Atre,Shaku : Defining Today’s Data Mining, “Executive Update”, Business Intelligence Advisory Service , CUTTER , Vol.1 , No.4 , 2001.
3. Brand, Estelle & Gerritsen, Rob: Data Mining Solutions “DBMS Magazine”, 1998.
4. Edelstein, Herb: Mining For Gold “Information Week” April, 1997.
5. Hermiz, Keith B.: Critical Success Factors For Data Mining Projects, “DM Review Magazine”, ES Media Group Feb. 1999.

6. Houston, Andrea L. & Others: Medical Data Mining on the Internet: Research on a Cancer Information System "Artificial Intelligence Review" 13:437-466, 2000.
7. Information Discovery, Inc.: A Characterization of Data Mining Technologies and Processes "DM Review Magazine", EC Media Group, 2000.
8. Lehman, J.T.: Future Tense "Intelligent Enterprise Magazine", Oct. 2001.
9. Noonan, Jack: Data Mining Strategies "DM Review Magazine", EC Media Group, July 2000.
10. Rambaldi, Giacomo & Bautista, Mike : Monitoring and Evaluation: Beyond Record-Keeping "Special Reports – Suhay" July-September, 2000.
11. Saarenvirta, Gary: Operation Data Mining " DB2 Magazine", Summer 2001.
12. Seiner, Robert S. : The IRM Test: How Did "IT" Get This Way? A Self-Help Test "DM Review Magazine", March 2002.
13. Skalak, David: Data Mining Blunders Exposed! "DB2 Magazine", Summer 2001.
14. Small, Robert D.: Debunking Data Mining Myths " Information Week", CMP Media, Jan. 1997.
15. Wu, Jonathan: What is Data Mining?, "DM Review Magazine", EC Media Group, August 2000.
16. Wu, Jonathan: The Value in Mining Data, "DM Review Magazine", Feb. 2002.

ج- الانتـــرنت Internet

1. Bloor Research : Databases on evaluation & comparison, 2001.
2. Kunstelj, Mateja & Leben, Anamarija & Vintar, Mirko : Influences Of Information Technology On The Quality Of Public Services , NISP Acee Annual Coference , 10-12/5/2001 , Jurmala , Latvia .
3. Ramachandran M, Pushpa: Mining for Gold Wipro Technologies, December 2001. http://www.Wipro.com
4. Smyth, Padhraic: Data Mining At The Interface of Computer Science And Statistics, 2001.
5. Soni, Sanjay & Tang, Zhohui & Yang, Jim : Performance Study of Microsoft Data Mining Algorithms, Microsoft Corp., 2001.
6. Tiedrech, Alan H.: Microsoft Corp. SQL Server 2000 Analysis Services, Datapro Information Services, Gartner Group, Inc., November 2000.

7. **Two Crows Corporation: Introduction to Data Mining and Knowledge Discovery, Third Edition, 1999.**
8. **Two Crows Corporation: Data Mining Applications, 2002. http://www.Towcrows.com**
9. **White Cross: Mining Very Large Databases to Support Knowledge Exploration, Version 1, January 5, 2001. http://www.Whitecross.com**
10. **Zaiane, Osmar R.: Introduction to Data Mining , CMPUT 690 Principles of Knowledge Discovery in Databases , University of Alberta, 1999 .**

د - الكـــتب **Books**

1. **Avison, David & Shah, Hanifa "The Information Systems Development Life Cycle" : McGraw-Hill, UK, 1997.**
2. **Daft, Richard L. "Organization Theory and Design" Seventh Edition, South-Western College Publishing , U.S.A , 2001 .**
3. **Gunderloy, Mike & Jorden, Joseph L. "Mastering SQL Server 2000" SYBEX Inc., U.S.A, 2000.**
4. **Hempel, George H. & Simonson, Donald G. "Bank Management, Text and Cases" Fifth Edition, John Wiley & Sons,Inc.U.S.A,1999.**
5. **Revsine, Lawrence & Collins, Daniel W. & Johnson,W. Bruce "Financial Reporting & Analysis " Prentice Hall ,Inc. ,U.S.A,1999.**
6. **Rob, Peter & Coronel, Carlos "Data Base Systems Design, Implementation and Management" Fourth Edition Course Technology, 2000.**
7. **Robbins, Stephen & Coulter, Mary "Management", Sixth Edition, Prentice-Hall, USA, 1999.**
8. **Romney, Marshall B. & Steinbart, Paul John "Accounting Information Systems" Eighth Edition Prentice-Hall, Inc. USA, 2000.**
9. **Rose, Peter S. "Commercial Bank Management" Forth Edition, McGraw-Hill, Inc., Singapore, 1999.**
10. **Saunders, Anthony "Financial Institutions Management" Second Edition, McGraw-Hill, Inc., USA, 1997.**
11. **Seidman, Claude " Data Mining with Microsoft SQL Server 2000 " Microsoft Corp., U.S.A, 2001.**

الملحق (1) المؤشرات المالية

النوع	المؤشر	طريقة حسابه
أ- نسب السيولة	1- نسبة التداول	الموجودات المتداولة / المطلوبات المتداولة
	2- نسبة السيولة	الموجودات المتداولة – (المخزون – المدفوعات المقدمة) / المطلوبات المتداولة
ب-نسب المديونية وراس المال	1- نسبة الديون الى راس المال	الديون الطويلة الاجل / حق الملكية
	2- نسبة الديون الى الموجودات	مجموع الديون / مجموع الموجودات
ج-نسب النشاط	1- معدل دوران المخزون	المبيعات / المخزون السلعي
	2- معدل دوران الذمم	المبيعات / اجمالي الذمم
	3- معدل دوران الموجودات الثابتة	المبيعات / صافي الموجودات الثابتة
	4- معدل دوران مجموع الموجودات	المبيعات / مجموع الموجودات
د- نسب الربحية	1- نسبة هامش الربح الاجمالي	الربح الاجمالي / المبيعات
	2- نسبة هامش الربح الصافي	صافي الدخل / المبيعات
	3- معدل العائد على الاستثمار	صافي الدخل / مجموع الموجودات
	4- نسبة العائد على راس المال المستخدم	صافي الدخل / راس المال المستخدم
	5- نسبة العائد على حقوق الملكية	صافي الدخل / حقوق الملكية
	6-نسبة الارباح لكل سهم	صافي الدخل - ارباح الاسهم الممتازة / عدد الاسهم العادية المباعة

المصدر: من إعداد الباحثان بالاعتماد على

(خان وغرايبة ، 78:1995) (Revsine & Others ,1999:160-174)

(شاكر وآخرون ،2000: 54-79) (Hempel & Simonson ,1999: 67)