محاولة تسيير خطر منح القروض البنكية باستخدام عملية التنقيب في البيانات. دراسة حالة بنك الفلاحة والتنمية الريفية – وكالة تيارت –

أ.د. شريط عابد*

بوجنان خالدية**

Résumé :

Dans le present travail nous essayerons de découvrir le processus d'extraction des données par l'application dans une banque algérienne'la Banque de l'Agriculture et du Développement Rural' agence de Tiaret- afin de réduire la perte des prêts accorder pour assurer la prise de la meilleur solution de prêt, a la fin de cet article nous atteindrons que la fouille de données est l'une des méthodes qui visent à améliorer la compréhension des données accumulées et tirer la richesse pour investir dans une voie qui mène à la prise de décision optimal.

Mots clés : fouille de données, entrepôt de données, la prédiction, groupes inégaux.

المستخلص: نهدف من خلال هذا المقال إلى كشف الغطاء على عملية التنقيب في البيانات من خلال تطبيقها في أحد البنوك الجزائرية " بنك الفلاحة والتنمية الريفية" – وكالة تيارت – بهدف تخفيض خسارة القروض الممنوحة بما يضمن اختيار أفضل لقرار الإقراض، ليتم في الأخير التوصل إلى أنها أحد الأساليب التي تهدف إلى تحسين فهم البيانات المتراكمة واستخلاص الثروة الموجودة فيها من أجل استثمارها بالشكل الذي يؤدي إلى إتخاذ القرار الأمثل.

الكلمات المفتاح : التنقيب في البيانات، مستودعات البيانات، التنبؤ، المجموعات غير متكافئة.

* أستاذ التعليم العالي، جامعة تيارت – الجزائر.

** طالبة دكتوراه، جامعة تيارت – الجزائر.

تمهيد

أصبحت المؤسسات المالية تتعامل مع الزبائن بشكل أساسي مما يتطلب منها جمع كميات هائلة من البيانات عن هؤلاء الزبائن هذا بالإضافة إلى ما يرد إليها يوميا من بيانات؛ يجعلها أمام أكداس كبيرة من البيانات تحتاج إلى جهود جبارة تحسن التعامل معها والإستفادة منها. فالتعامل اليدوي مع مثل هذه البيانات دون استخدام تقنيات حديثة أفضل من شأنها أن تؤدي إلى إتخاذ قرارات غير رشيدة وبهدف إيجاد بديل لهذه الطريقة إرتئينا كشف الغطاء عن إحدى الطرق الإحصائية المعتمدة في مختلف الدول المتقدمة ومحاولة تطبيقها على واقع أحد البنوك الجزائرية" بنك الفلاحة والتنمية الريفية " – وكالة تيارت – ؛ هذه الطريقة هي التنقيب في البيانات والتي تعتمد على استخراج المعرفة من كميات هائلة من البيانات. وهي تقنية فرضت نفسها بقوة في عصر المعلوماتية، كأداة مناسبة متاحة للمؤسسات المالية للإستكشاف والتركيز على أهم المعلومات في قواعد البيانات، كما تركز تقنيات التنقيب على بناء التنبؤات المستقبلية واستكشاف السلوك والإتجاهات مما يسمح باتخاذ القرارات الصحيحة واتخاذها في الوقت المناسب، ويمكن استخدامها في إدارة المخاطر المختلفة التي تواجهها والتحوط منها، بحيث يتخصص مجموعة من الباحثين والمتخصصين في تصميم وابتكار تقنيات جديدة من خلال ربطها مع تقنيات الذكاء الإصطناعي والعمليات الإحصائية.

يستمد هذا البحث أهميته من المساهمات العلمية التي يقدمها لأنه يتناول قضية أساسية هي التنقيب في البيانات والتي تتيح للمؤسسات المالية تجنب مخاطر حقيقية ناتجة عن عدم التسديد، حيث أنها تشكل عبئا عليها لكونها قروضا هالكة هذا من جهة، ومن جهة أخرى يمكن أن تستخدم في التنبؤ بالمقترضين الموثوق بهم واستبعاد غير الموثوقين منهم لتقليل تكاليفهم .

وبغية بلوغ الهدف المتوخى من هذه المقالة سنحاول الإجابة على الإشكالية التالية:

" هل استخدام عمليات التنقيب في البيانات تزيد من قدرة البنك في تسييره للقروض الممنوحة واتخاذ القرار الأمثل ؟ "

من خلال التطرق إلى المحاور التالية والمتمثلة في:

1. مفهوم التنقيب في البيانات وأنواعه ؛

2. مراحل عملية التنقيب في البيانات واستراتيجيات نجاحها ؛

3. دور عملية التنقيب في البيانات في تسيير منح القروض لبنك الفلاحة والتنمية الريفية – وكالة تيارت.

كما أننا انهينا بحثنا بخاتمة تتضمن بعض النتائج والتوصيات.

1- مفهوم عملية التنقيب في البيانات وأدواتها.

إن التطور في العلم والإقتصاد وتكنولوجيا المعلومات أدى إلى زيادة كمية البيانات الرقمية في الآونة الأخيرة هذه الكميات الهائلة من البيانات لم تعد وسائل التحليل التقليدية قادرة على التعامل معها. لذلك ظهر التنقيب عن البيانات واثبت وجوده كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات، وذلك بتحويلها من مجرد معلومات متراكمة وغير مفهومة (بيانات) إلى معلومات قيمة يمكن استغلالها والإستفادة منها بعد ذلك.

1-1- مفهوم عملية التنقيب في البيانات

ظهر مصطلح التنقيب في البيانات في منتصف التسعينات في الولايات المتحدة الأمريكية، وهو يجمع ما بين الإحصاء وتكنولوجيات الإعلام، وتوجد عدة تعريفات لهذا المفهوم، نذكر أهمها:

حيث يمكن تعريفها بأنها : " الإستكشاف الآلي أو المؤتمت لأنماط شائقة وغير جلية مخفية في قاعدة بيانات معينة"[1].

وتعرف أيضا بأنها : " سيرورة تحليل دقيقة وذكية، تفاعلية وتسلسلية، تسمح لمسيري النشاطات عند استخدام هذه السيرورة باتخاذ قرارات والقيام بأعمال ملائمة في صالح النشاط المسؤولين عنه والمؤسسة التي يعملون بها"[2].

أو أنها : " عبارة عن تحليلات لكمية كبيرة من البيانات بغرض إيجاد قواعد وأمثلة ونماذج التي يمكن أن تستخدم تقود وتدل أصحاب القرار، وتتنبأ بالسلوك المستقبلي"[3].

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"[4].

من خلال التعريفات السابقة يمكن القول أن التنقيب في البيانات عبارة عن سيرورة لاستخراج أو اكتشاف معرفة مفيدة وقابلة للإستغلال من خلال مجموعة كبيرة من البيانات. حيث يساعد في استكشاف المعرفة المخفيّة والنماذج غير المتوقّعة، إضافة إلى استكشاف قواعد جديدة موجودة في قواعد بيانات كبيرة.

---

[1] بشير عباس، العلاق، الإدارة الرقمية: المجالات والتطبيقات، مركز الإمارات للدراسات والبحوث الإستراتيجية، ابوظبي، 2005، ص: 84.

[2] M.Bazsalica & Autres , Data mining pour le Web, Edition Eyrolles, Paris, 2001, P : 61.

[3] عبد الستار العلي وآخرون، المدخل إلى إدارة المعرفة، دار المسيرة للنشر والتوزيع والطباعة، الطبعة الأولى، عمان، 2006، ص: 157.

[4] D.Hand & Autres, Principles of Data Mining, MIT Press, London, 2001, p: 01.

1-2- أدوات للتنقيب في البيانات

وهناك عدة أدوات للتنقيب في البيانات، نذكر أهمها:

1-2-1- التلخيص

يشير التلخيص إلى أساليب تفتيت كتل البيانات الكبيرة إلى مقاييس موجزة، توفر وصفا عاما للمتغيرات وعلاقاتها[5]. ومن الأمثلة على أساليب التلخيص نذكر : المتوسطات، المجاميع والإحصائيات الوصفية التي تتضمن مقاييس النزعة المركزية مثل : المتوسط الحسابي، الوسيط والمنوال، وكذا مقاييس التشتت كالانحراف 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 الأفراد خصوصا.

1-2-2- التصنيف

يتمثل التصنيف في التفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى. هذه الخاصية هي عموما كيفية[6]. ويمكن انجاز التصنيف بالإعتماد على الأساليب الإحصائية القديمة مثل : الإنحدار والتحليل التمييزي، أو بالإعتماد على أساليب حديثة نسبيا مثل قوى الإرتباط والإستنتاج المستند إلى الحالة والشبكات العصبية الإصطناعية.

1-2-3- التنبؤ

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: الإنحدارات

[5] بشير عباس، العلاق، مرجع سبق ذكره، ص:92.

[6] M.Bazsalica & Autres , Op. Cit., p, p:68-69.

بأنواعها والتحليل التمييزي، أما الأساليب الجديدة فتشتمل على قواعد الإرتباط، شجرة القرار، الشبكات العصبية والخوارزميات الوراثية[7].

1-2-4- العنقدة أو التجزئة

التجميع العنقودي أو التجزئة إلى قطاعات عبارة عن مجموعات متجانسة في مجتمع من الأفراد.و يشير التجميع العنقودي أو التجزئة إلى قطاعات إلى عملية تشكيل مجموعات أو قطاعات مؤلفة من أفراد أو أصحاب أسر، وذلك بالإستناد إلى معلومات متضمنة مجموعة من المتغيرات التي تصفهم[8].

و تتم أساليب التجميع العنقودي غالبا بمساعدة أساليب التحليل العنقودي الإحصائية والأساليب المستندة إلى شجرة القرار، الشبكات العصبية الإصطناعية والخوارزميات الوراثية.

1-2-5- تحليل الإرتباط

يتمثل الإرتباط في البحث عن علاقات أو ارتباطات موجودة بين عدة خصائص. ويشير تحليل الإرتباط إلى مجموعة من الأساليب التي تستخدم لربط أنماط التعامل عبر القطاعات المتقاطعة أو عبر الوقت[9].

---

[7] J.Berry, Data Mining Techniques For Marketing., Edition Wiley Publishing,INC, Indianapolis, 2004, p: 10.

[8] M.Bazsalica & Autres , Op. Cit., p, p:66.

[9] بشير عباس العلاق، مرجع سبق ذكره، ص: 93.

1-2-6- الكشف عن التغيرات أو الانحرافات

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

2- مراحل عملية التنقيب في البيانات واستراتيجيات نجاحها.

سنحاول من خلال هذا المحور تسليط الضوء على مراحل عملية التنقيب في البيانات واستراتيجيات نجاحها.

2-1- مراحل عملية التنقيب في البيانات

يمكن تلخيص مراحل وخطوات عملية التنقيب في البيانات كما يلي[10]:

2-1-1- فهم طبيعة الأعمال

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

2-1-2- فهم البيانات

تعتبر مسالة 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.

[10] زكرياء مطلك الدوري، داليا عبد الحسين أحمد، مجلة العلوم الاقتصادية والادارية، عدد 48، 2007، ص:49.

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 كالآتي:

- **تجميع البيانات** : وهي الخطوة الموجهة نحو تحديد مصدر البيانات في الدراسة بما في ذلك استخدام البيانات العامة الخارجية مثل الضرائب وغيرها.
- **توصيف البيانات** : وهي الخطوة التي تركز على توصيف محتويات الملف الواحد من الملفات أو الجداول.
- **جودة البيانات وتحقيقها** : 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.
- **التحليل الاسترشادي للبيانات**: تستخدم الأساليب مثل الإظهار المرئي أو التصور أو عملية التحليل المباشر التي تؤدي إلى إجراء التحليل الأولي للبيانات. وتعتبر هذه الخطوة مهمة وضرورية لأنها تركز على تطوير الفرضيات المتعلقة بالمشكلة قيد الدراسة.

2-1-3- تهيئة البيانات

وتشمل الخطوات التالية[11]:

- **الإختيار**: وتعني اختيار المتغيرات المتوقعة وحجم العينة.
- **صياغة المتغيرات وتحويلها**حيث يجب دائما أن تصاغ المتغيرات الجديدة لبناء النماذج الفعالة.
- **تكامل البيانات** : حيث أن مجاميع البيانات في دراسة التنقيب عن البيانات من الممكن خزنها في قواعد بيانات متعددة الأغراض التي تكون بحاجة إلى توحيدها في قاعدة بيانية واحدة.
- **تصميم وتنسيق البيانات**: حيث تتعلق هذه الخطوة في إعادة ترتيب حقول البيانات كما يتطلب في نموذج التنقيب في البيانات.

2-1-4- صياغة نماذج الحل وثبوتها

إن بناء وصياغة نموذج الحل السليم والدقيق يتم من خلال عملية الخطأ والصواب، 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.

2-1-5- التقييم وتعليل نتائج النموذج

حالما يتم صياغة النموذج والتحقق من ثباته وصدقه، تجري مباشرة عملية التحقق من ثبات حزمة البيانات التي يتم تغذيتها بواسطة النموذج. وبما أن نتاجات 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.

[11] Seidman, Claude , Data Mining with Microsoft SQL Server 2000 , Edition Microsoft Corp., U.S.A, 2001,p :56.

2-1-6- نشر وتوزيع النموذج

حيث تشتمل هذه الخطوة على نشر وتوزيع النموذج داخل المنظمة لمساعدة عملية صنع القرار. وأن النموذج الصالح يجب أيضا أن يحقق الرضا لدى المستفيدين طالما أن اختيار النموذج لابد أن يتم من خلال الدراسة الاسترشادية أو نموذج مصغر من الدراسة الشاملة.

2-2- استراتيجيات نجاح تنقيب البيانات

لكي يتم الحصول على تنقيب البيانات ناجح، لا بد من اتباع عدد من الإستراتيجيات التي تضمن ذلك ومنها[12]:

توجيه العمل نحو هدف محدد يوضح المشاكل المراد حلها؛

- توفير بيانات ذات كمية ومحتوى عال يزيد من قوة التنبؤ فكلما كانت البيانات أفضل، كانت النتائج أفضل أي ذات تحليل شامل وأكثر دقة؛
- بناء عدة نماذج باستخدام عينات من البيانات بدلا من بناء نموذج واحد باستخدم كل البيانات والإسراع في استخدام تنقيب البيانات؛
- التواصل المستمر بين مدراء العمل والخبراء مع إمكانية إحتواء متخذ القرار من أجل إعادة تعريف أو تصميم أهداف العمل وإيجاد النتائج الأولية وتكرار ذلك مع إزدياد إدراك البيانات؛
- التخطيط للتعلم من عملية تنقيب البيانات وتعليمها الى من له صلة مباشرة باستخدامها في وعدم إهمال أي نتائج تحت ضغط الوقت والكلفة؛
- التكامل بين عدد من الأدوات بضمنها القديمة والجديدة.

[12] Robbins & Autres, Management, Edition Prentice-Hall, U.S.A, 1999, p :23.

3- دور عملية التنقيب في البيانات في تسيير منح القروض لبنك الفلاحة والتنمية الريفية – تيارت-

قمنا في بحثنا هذا بعملية تنقيب البيانات على ملفات منح القروض بنك الفلاحة والتنمية الريفية BADR - وكالة تيارت- لغرض مساعدة البنك في اتخاذ القرار الذي يمكنه من الحكم على المقترضين، إذا كان التعامل معهم يشكل خطر على البنك أم لا؛ وبذلك يتمكن البنك من التركيز على العملاء الموثوقق بهم واستبعاد من هم غير موثوق بهم.

3-1- الدراسة التطبيقية

لإجراء الدراسة والتوصل للنتائج المبتغاة قمنا باتباع الخطوات التالية:

3-1-1- جمع المعطيات

للقيام بدراستنا كان المجتمع المستهدف مجموع المؤسسات والأفراد الذين استفادوا من قرض على الأقل لدى بنك الفلاحة والتنمية الريفية – تيارت- بين فترة 2000-2012.

3-1-2- عرض البيانات

تم اختيار العينة بطريقة عشوائية من خلال أرشيف البنك الخاص بلملفات منح القروض، حيث بلغ عدد الملفات المقدمة 365 ملف وتشمل العينة المستهدفة على:

- 522 طلب قرض مقبول.
- 113 طلب قرض مرفوض.

الهدف من الحصول على هذه القروض هو تمويل المشاريع الإستثمارية.

3-1-3- دراسة طبيعة المتغيرات

تمت عملية جمع المتغيرات من خلال الدراسة الميدانية التي قمنا بها على مستوى مصلحة القروض بالبنك محل الدراسة وصنفت هذه المتغيرات حسب طبيعتها إلى متغيرات نوعية ومتغيرات كمية. (انظر الجدول رقم 01)

لتبسيط الإجراءات قمنا بتصفية الضمانات غير المادية (الأمن، التأمين، عدم سحب فوائد...)، كما قمنا بتقدير الضمانات المادية بالدينار الجزائري.

يمكن تقسيم المتغيرات التفسيرية إلى ثلاثة فئات وهي:

- **المتغيرات التي تصف المقترضين**:الجنس، السن، الحالة الإجتماعية، مستوى التعليم.
- **المتغيرات المرتبطة بالنشاط**:مجال النشاط، الربحية، قيمة القرض.
- **أنواع أخرى من المتغيرات**: نوع العميل، قيمة الضمان، السنة.

من خلال التحليل أحادي التباين وثنائي التباين سنحاول وضع وصفا للمتغيرات من أجل تحديد المعلومات الأساسية وأهم الروابط بينها والتي ستساعدنا كثيرا في مرحلة معالجة البيانات.

3-1-4- التحليل الوصفي للمعطيات

من أجل التنبؤ بالمقترضين الموثوق بهم واستبعاد غير الموثوقين منهم لتقليل تكاليفهم قمنا بالتحليل الوصفي للمتغيرات، وذلك باتباع المراحل التالية:

3-1-4-1- اختبار كاي تربيع

بعد دراستنا لملفات القروض الخاصة بالعنية المدروسة ومن خلال تصنيف البنك لها إلى طلبات مرفوضة ومقبولة، قمنا باختبار كاي تربيع باستخدام التحليل ثنائي التباين، وذلك للتعرف على جودة التوافق بين المتغير التابع "القرار" والمتغيرات التفسيرية على النحو التالي:

الفرضيات:

- H_0: عدم وجود علاقة بين اتخاذ القرار والمتغيرات التفسيرية.
- H_1: وجود علاقة بين اتخاذ القرار والمتغيرات التفسيرية.

نقوم باستخراج القيمة الإحتمالية المستخرجة من برنامج التحليل الإحصائي SPSS اعتمادا على اختبار كاي تربيع ومقارنتها بمستوى الثقة والذي يمثل 05%، لمعرفة ما إذا كانت هذه القيمة أكبر من مستوى الثقة تقبل فرضية العدم، أما إذا كانت القيمة الاحتمالية أصغر من مستوى الثقة ترفض فرضية العدم. (انظر الجدول رقم 02).

يتضح من خلال النتائج المتحصل عليها في الجدول رقم 02 وبعد إجراء إختبار كاي مربع أن هناك استقلالية بين المتغير التابع "القرار" والمتغيرات التفسيرية المتمثلة في : مجال النشاط، الحالة الإجتماعية، الجنس.

أما فيما يخص باقي المتغيرات التفسيرية المتبيقية نجد أن النتائج تشير إلى وجود تبعية بينها وبين عملية اتخاذ القرار المناسب.

و منه نستنتج أن منح القروض لا يعتمد على نوع الجنس أو الحالة الإجتماعية أو مجال النشاط، في حين أن باقي المتغيرات التفسيرية فهي تلعب دورا أساسيا في اتخاذ القرار الرشيد لمنح القروض، وعليه فإن البنك سيتجنب مخاطر حقيقية ناتجة عن خطر عدم التسديد والتي تشكل عبئا عليه لكونها قروضا هالكة.

3-1-4-2- معامل الإرتباط

لقياس شدة العلاقة بين المتغير النوعي والمتغيرات الكمية، نرى أنه من الضروري حساب قوة الارتباط بينها. والواقع أن معامل الإرتباط هو نتيجة مباشرة لمقدار التباين المشترك للمتغيرات.

معامل الإرتباط هو قيمة يتراوح مجال قراءاتها ما بين 1- وبين 1+، وهو يبيّن لنا إذا تحققت علاقة ما بين متغيرين وما مدى قوّة وإتّجاه هذه العلاقة. (انظر الجدول رقم03)

نلاحظ من خلال النتائج المتحصل عليها من الجدول رقم 03 أن نسبة الإرتباط على مقربة من قيمة 0 للمتغيرات الثلاثة، مما يعني عدم وجود قوة إرتباط بين المتغيرات التفسيرية الكمية والمتغير النوعي التابع .

3-1-4-3- تحويل المتغيرات

المتغيرات التفسيرية التي نحن بصدد دراستها تتخذ شكلين كمية أو نوعية، ومن أجل تطوير نموذج تنبؤي فمن الضروري التحقق من دقة البيانات المتوفرة لدينا، وذلك من خلال تحويل المتغيرات الكمية إلى متغيرات نوعية بعد تقسيمها إلى فئات.

يرفض عادة طلب الحصول على قرض بسبب عاملين : العامل الأول مرتبط بخصائص المقترض والمشروع، أما العامل الثاني فيتعلق بإدارة وتوجيه البنك.

لذلك، ومن أجل بناء نموذج تنبؤي يميز بين المقترضين الموثوق بهم واستبعاد غير الموثوقين منهم، يجب علينا أن نستبعد الطلبات المرفوضة من طرف البنك لأسباب إدارية التي يقدر عددها بـــ 31 طلب قرض مرفوض، وهو ما يعني أن عدد المقترضين في النمذجة يقدر بـــ 604مقترض، 82 طلب قرض مرفوض و522 قرض مقبول.

و بالإضافة إلى ذلك، لا يمكن إدراج السنة كمتغير في النموذج التنبؤي لأنه يقتصر على الزمن الماضي .لهذا، لا بد من استبعاده بعد تقسيم المتغيرات الكمية :عمر المقترض، قيمة القرض وقيمة الضمانات تم الحصول على المتغيرات النوعية الترتيبية على النحو التالي:

- عمر المقترض: (1) $]19، 27]$ (2) $]27، 32]$ (3) $]32، 36]$ (4) $]36، 66]$
- قيمة القرض: (1) $]13985، 1054011]$ (2) $]1054011، 1415473]$

(3)]1415473 ،2093714] (4)]2093714 ،20147331]

- قيمة الضمان: (1)]0 ،1500898] (2)]1500898 ،2183093]

(3)]2183093 ،3694068] (4)]3694068 ،9895016165].

3-1-4-5- استخراج البيانات

لغرض تحقيق أهداف البحث اعتمدنا على تصميم تجريبي يطلق عليه إسم المجموعات المتكافئة، والذي يتكون من مجموعتين الأولى تجريبية والثانية ضابطة (أصلية).

أخذ العينات هو خطوة أساسية في التطبيق بما في ذلك التنبؤ أو التصنيف والعديد من التقنيات الممكنة، ولكن حجم البيانات المتوفرة لدينا 604 طلب للحصول على قرض صغير نسبيا، لذلك سنقوم بتقسيم المقترضين إلى عينتين، عينة تجريبية تمثل 75 ٪ من المقترضين، أي453 طلب قرض وعينة أصلية (ضابطة) تمثل 25٪ من المقترضين، أي151 طلب قرض.

يتطلب هذا التصميم تهيئة مجموعتين متكافئتين في عدد من المتغيرات التي قد تؤثر في المتغير التابع وقد يتطلب البحث ضبط المتغيرات من خلال تكافؤ مجموعتي البحث. (انظر الجدول رقم04)

- **العينة التجريبية هي:**

N_{app}: حجم العينة التجريبية.

$N_{ap-accep}$:حجم العينة التجريبية المحصل عليها من طلبات القرض المقبولة.

N_{ap-ref} :حجم العينة التجريبية المحصل عليها من طلبات القرض المرفوضة.

إذن يقدر حجم العينة التجريبية المحصل عليها من طلبات القرض المقبولة بـــ:
522 × 75, 0 = 391 طلب قرض.

أما حجم العينة التجريبية المحصل عليها من طلبات القرض المرفوضة. تقدر بـــ: = 62 82 × 0 ,75 طلب قرض.

و عليه فإن حجم العينة التجريبية يقدر بـــ:

391+62=453 طلب قرض.

- **العينة الأصلية (الضابطة) هي:**

N_{val} :حجم العينة الأصلية

$N_{val-accep}$:حجم العينة الأصلية المحصل عليها من طلبات القرض المقبولة.

$N_{val-ref}$:حجم العينة الأصلية المحصل عليها من طلبات القرض المرفوضة.

إذن يقدر حجم العينة الأصلية المحصل عليها من طلبات القرض المقبولة بـــ:= 131 522 × 25 ,0طلب قرض.

أما حجم العينة الأصلية المحصل عليها من طلبات القرض المرفوضة. تقدر بـــ: = 20 82 × 0,25 طلب قرض.

وعليه فإن حجم العينة الأصلية يقدر بـــ:

131+20=151 طلب قرض.

لبناء أحسن نموذج سنقوم باستعمال أسلوب التحليل التمييزي خطوة - خطوة، واختيار مسافة Mahalanobisالتي تعتمد على تساوي المعدلات بين المجموعتين، كما سيتم انتقاء المتغيرات على أساس قيمة فيشر.

من خلال الجدول رقم05، نلاحظ أن برنامج التحليل الإحصائي SPSS قام باختيار ثلاثة متغيرات، هذه المتغيرات المتحصل عليها هي الأكثر استقصاء في النموذج.

3-1-4-6- الدالة الاستقصائية النموذجية

أحسن دالة استقصائية نموذجية تأخذ أكبر قيمة موجودة في المحور الأول. (انظر الجدول رقم 06)

أما فيما يخص الارتباط التجميعي Canonical Correlation بين الدالة الاستقصائية والمجموعتين تمثله القيمة 0.766 كما هو موضح في وهو ارتباط قوي يدل على جودة توفيق الدالة التمييزية.

أما قيمة Wilk's lambda فقد كانت 0.414 ما يدل على جودة الدالة، لأن قيمة sig < 0.05.

3-1-4-7- قياس دقة النموذج

يتضح من خلال النتائج المحصل الخاصة بالعينة التجريبية (انظر الجدول رقم07) أنه من بين 62 طلب قرض مرفوض توصل النموذج إلى أن 30 طلب قرض بقي مرفوضا أي بنسبة تصنيف صحيح تقدر بـ 48.39 ٪ في حين 32 طلب قرض تم قبوله أي بنسبة خطأ تقدر بـ51.61٪.

ومن بين 391 طلب مقبول توصل النموذج إلى أنه لم يرفض أي طلب قرض أي بنسبة تصنيف صحيح تقدر بـ 100٪.

يتضح من خلال النتائج المحصل الخاصة بالعينة التجريبية (انظر الجدول رقم 08) أنه من بين 130 طلب قرض مقبول ضمن العينة الأصلية توصل النموذج إلى أنه تم قبول 131 طلب قرض، أي بنسبة تصنيف صحيح تقدر بـ 99.24٪ وهي نسبة مرتفعة جدا تدل على مدى دقة النموذج، في حين تم رفض طلب قرض واحد أي بنسبة خطأ تقدر بـ 0.76٪.

ومن بين 20 طلب مرفوض من العينة الأصلية توصل النموذج إلى أن 09 طلب قرض بقي مرفوضا، أي بنسبة تصنيف صحيح تقدر بـ 45.00٪ في حين 11 طلب قرض تم قبوله أي بنسبة خطأ تقدر بـ 55.00٪.

- التحليل

بالإعتماد على التحليل التمييزي قمنا ببناء نموذج تنبؤي مما يسمح للبنك باتخاذ القرار المناسب في منح القروض (التمييز بين الطلبات القبولة والطلبات المرفوضة) فمن خلال النتائج المتحصل عليها تم تقدير نسبة الخطأ الإجمالية بـ 7.95 ٪ ؛ إذا فهو نموذج مقبول.

كما تم استقصاء ثلاثة متغيرات تفسيرية تؤثر على اتخاذ القرار هي : ربحية ضعيفة أو متوسطة، قيمة القرض مابين 13 و105 مليون دينار جزائري وما بين 105 و210 مليون دينار جزائري وقيمة الضمانات المقدمة.

يدل هذا على أن منح قرار منح القروض يقدم على أساس جودة المشروع من حيث الربحية والضمانات المتوفرة.

خاتمة

تعرضنا في هذا المقال لمفهوم عملية التنقيب في البيانات، وأسباب نشأتها، حيث اتضح لنا أن عملية التنقيب في البيانات من المصطلحات الجديدة التي دخلت عالم الإقتصاد، وهي تعني الإستكشاف، التحليل، والتطبيق لأدوات تحليلية مبتكرة، والصياغة لحلول إبداعية والعوامل التي أدت إلى ظهور هذا المفهوم وساعدت في الوقت نفسه على انتشاره على أرض الواقع كثيرة أهمها اتساع وتعدد البيانات، بالإضافة إلى تطور تكنولوجيا المعلومات والإتصال...

كما عرجنا من خلال المحور الثاني على المراحل التي تمر بها عملية التنقيب في البيانات، وبينا أهم استراتيجيات نحاحها.

أما في المحور الثالث فقد قمنا بتطبيق عملية التنقيب في البيانات على عينة من ملفات منح القروض لدى بنك التنمية المحلية والريفية – وكالة تيارت- وتوضح لنا أن عملية التنقيب في البيانات تركز على بناء التنبؤات المستقبلية واستكشاف السلوك والإتجاهات، مما يسمح بتقدير القرارات الصحيحة واتخاذها في الوقت المناسب.

وعلى ضوء هذا العرض استخلصنا النتائج التالية:

- توفر عملية تنقيب البيانات ميزة التعامل مع كميات هائلة من البيانات وكلما كانت كمية البيانات كبيرة كانت أكثر قدرة ودقة في الوصول الى نماذج صحيحة مفيدة؛
- يختلف تنقيب البيانات عن أنظمة دعم القرار كونه قد يكشف عن علاقات أبعد من أن يفكر بها المستخدم ؛
- تعتمد نتائج تنقيب البيانات بصورة أساسية على مدى دقة وصحة وشمولية البيانات المستخدمة ؛
- تستخدم في تنقيب البيانات تقنيات وأدوات عديدة ومتنوعة لكنها تصب في اتجاهين أساسيين هما : الإستكشاف والتنبؤ؛
- في حالة وجود قاعدة بيانات كبيرة جدا، يكون خيار بناء نماذج عديدة مبنية على عينات أفضل من خيار بناء نماذج قليلة مبنية على كل البيانات لكي يكون النموذج أكثر دقة ومتانة ؛
- يمكن اجراء عملية التنقيب على بيانات تضم متغيرات رقمية وغير رقمية ؛
- يوفر تطبيق البرنامج عامل السرعة حيث يمكن الوصول إلى نموذج التنقيب في غضون دقيقة أو اثنتين كحد أقصى مع السماح بتحديث هذه النماذج باستمرار، إن هذا التحديث والسرعة يسهل على إدارة البنك اتخاذ القرار الخاص بالتعامل مع طالبي القروض بالرفض أو القبول ؛
- ضرورة توفير عناية أكبر بعملية جمع البيانات من زبائن البنك وإدخالها على الحاسوب الإلكتروني بدلا من تركها في ملفاتها وما يتطلبه ذلك من توفير اجهزة الحاسوب الحديثة ؛
- لا تقتصر فائدة أجهزة الحاسوب على إدخال بيانات الزبائن فقط، بل يتعداه إلى تكوين قاعدة بيانات خاصة بزبائن البنك ككل أو قاعدة بيانات خاصة بالقروض والمقترضين أو التسهيلات المصرفية، ومن الأفضل أن يكون ذلك من خلال تكوين

مستودع البيانات الذي يعمل على حفظ البيانات وتنقيتها وتحديثها باستمرار لضمان استرجاع أي جزء منها في أي وقت ؛

- العمل على زيادة نشر الوعي في البنك وخاصة في المستويات التنفيذية لتقنية تنقيب البيانات توضيح مدى الحاجة الملحة لاتخاذ القرارات المستندة على المعرفة لما لذلك من دور كبير في زيادة أداء البنك من خلال تقليل التكاليف الناجمة عن عدم تسديد المقترضين وارتفاع عدد القروض الهالكة ؛

- التوسع في حجم ونوع البيانات المطلوبة من الزبائن من أجل إضافة متغيرات جديدة لعملية التنقيب.

ملحق الجداول

الجدول رقم **01**: طبيعة المتغيرات.

نوع المتغير	المتغير	تعيين المتغير
متغيرات نوعية	**القرار**	- طلب مقبول - طلب مرفوض
متغيرات كمية	**الجنس**	- ذكر - انثى
	الحالة الإجتماعية	- متزوج - أعزب
	المستوى التعليمي	- ابتدائي - متوسط - عالي
	طبيعة المقترض	- فرد - مؤسسة
	مجال النشاط	- ريفي - حضري - صناعي
	قطاع النشاط	- صناعة - تجارة - خدمات - فلاحة
	نوع القرض	- قروض بالصندوق - قروض بالإمضاء - قروض بالصندوق والإمضاء
	الربحية	- ضعيفة - متوسطة - قوية
	السنة	
	عمر المقترض	
	قيمة القرض	
	نوع الضمان	- ضمانات شخصية - ضمانات حقيقية - ضمانات شخصية+ حقيقية

المصدر : من إعداد الطالبة.

الجدول رقم **02**: نتائج اختبار كاي مربع

	النتيجة	القيمة الإحتمالية	كي مربع المجدولة	كي مربع المحسوبة
الجنس	H_0مقبولة	0,957	3,841	0,003
الحالة الإجتماعية	H_0مقبولة	0,695	3,841	0,154
المستوى التعليمي	H_0مرفوضة	0,000	5,991	17,938
نوع المقترض	H_0مرفوضة	0,019	5,991	7,957
مجال النشاط	H_0مقبولة	0,426	9,488	1,709
قطاع النشاط	H_0مرفوضة	0,001	5,991	19,682
نوع القرض	H_0مرفوضة	0,011	5,991	9,103
الربحية	H_0مرفوضة	< 0,0001	5,991	201,538
السنة	H_0مرفوضة	< 0,0001	21,026	307,792

.**SPSS** المصدر : مخرجات البرنامج الإحصائي

الجدول رقم **3**: نتائج معامل الإرتباط.

المتغيرات	قيمة معامل الإرتباط	التباين المشترك
عمر المقترض	0،0001	43،668
مقدار الائتمان	0،0953	2136788915995.620
ضمانات القيم	0،0079	6016721308655.590

.**SPSS** المصدر : مخرجات البرنامج الإحصائي

الجدول رقم **04** :حجم العينة المدروسة.

نوع العينة	حجم العينة	القرار	حجم العينة الكلي	نسبة المقترضين
حجم العينة التجريبية	391	طلب قبول	453	**0.75** ٪
	62	طلب رفض		
الأصلية حجم العينة	131	طلب قبول	151	**0.25** ٪
	20	طلب رفض		

المصدر : من إعداد الطالبة.

الجدول رقم **05** : المتغيرات الأكثر استقصاءا في النموذج.

قوة الإرتباط بين المتغيرات	المتغيرات
0,339	ربحية ضعيفة
0,112	قيمة القرض] 1415473,1054011]
0,220	قيمة الضمان] 2183093,1500898]

المصدر : مخرجات البرنامج الإحصائي**SPSS**.

الجدول رقم **05** : تمثيل المتغيرات على المحور الأول.

	المحور الأول
طلب مقبول	0,473
طلب مرفوض	-2,982

.**SPSS** المصدر : مخرجات برنامج

الجدول رقم **06**: نتائج العينة التجريبية.

العينة التجريبية	طلب مرفوض	طلب مقبول	المجموع	نسبة التصنيف الصحيح٪
طلب مقبول	0	391	391	**٪100.00**
طلب مرفوض	30	32	62	**٪48.39**
المجموع	30	423	453	**٪92.94**

المصدر : من إعداد الطالبة.

الجدول رقم **07**: نتائج العينة التجريبية.

العينة الأصلية	طلب مرفوض	طلب مقبول	المجموع	نسبة التصنيف الصحيح٪
طلب مقبول	1	130	131	**٪99.24**
طلب مرفوض	9	11	20	**٪45.00**
المجموع	10	141	151	**٪92.05**

المصدر : من إعداد الطالبة.

المراجع:

بشير عباس، العلاق، الإدارة الرقمية: المجالات والتطبيقات، مركز الإمارات للدراسات والبحوث الإستراتيجية، ابوظبي، 2005.

عبد الستار العلي وآخرون، المدخل إلى إدارة المعرفة، دار المسيرة للنشر والتوزيع والطباعة، الطبعة الأولى، عمان، 2006.

زكرياء مطلك الدوري، داليا عبد الحسين أحمد، مجلة العلوم الاقتصادية والادارية، عدد 48، 2007.

D.Hand & Autres, Principles of Data Mining, MIT Press, London, 2001.

J.Berry, Data Mining Techniques For Marketing., Edition Wiley Publishing,INC, Indianapolis, 2004.

Seidman, Claude , Data Mining with Microsoft SQL Server 2000 , Edition Microsoft Corp., U.S.A, 2001.

Robbins & Autres, Management, Edition Prentice-Hall, U.S.A, 1999.

M.Bazsalica & Autres , Data mining pour le Web, Edition Eyrolles, Paris, 2001.