Statistical analysis to the data of consumer credit to evaluate credit scoring

Summary:
Credit scoring (creditworthy) is the term used by the credit industry to describe methods used for classifying applicants for credit into two risk classes according to their likely repayment behavior: "default" and "non-default". Accurate classification is of benefit both to the creditor (in terms of increased profit or reduced loss) and to the loan applicant (avoiding over-commitment). The research aims to evaluate the Lending Club database for consumer credit by examining the historical data of consumer loans of size 42539 during the period 2007-2011 using logistic regression and decision trees and to predict whether a new applicant paid off or defaulted upon their loan using automatic models. The data set consists of the financial attributes of each customer and included a mixture of loans that the customers paid off or defaulted upon for both accepted and rejected loans, which eliminates the bias suffered by the most credit assessment studies due to limiting themselves to the study of loan data only accepted. The paper compares the correct classification accuracy rates for logistic regression and decision tree algorithms. The logistic regression outperformed the decision tree methods. The most important variables were: total payments, amount, grade, term, policy, purpose and intercept.

Keywords: Credit Scoring, Logistic Regression, Decision Trees.
ملخص البحث

1-1 مقدمة

اتولدت محددة نمو البنوك العقارية في السعودية في العشر سنوات الأخيرة بشكل ملحوظ، وقد صاحب ذلك زيادة كبيرة أيضاً في حجم المخصصات وهو ما شجع البنوك (الأجنبية والوطنية) على دخول السوق السعودية لتشددي من هذا الوفر الذي لا تتفعل عليه في الغالب أي فرود للمخزون. ومن ثم فقد ساعد البنوك في تقديم برامج تسويقية مختلفة ساهمت إلى جانب ارتفاع نسبة المعيشة والنمو الاقتصادي الاجتماعي - في زيادة نسبة طالبي الترويج وقد أرتفعت القروض الشخصية من 178 مليار ريال في 2007 إلى 285.4 مليار ريال في الرابع من عام 2012 (مؤسسة النقد العربي السعودي، التقرير السنوي، 19 فبراير 2013) بزيادة 60% في 5 سنوات، وهو ما جعل السعودية تحتل المركز الأول على مستوى الخليج بالنسبة للقرود الشخصية.

وعلى الجانب الآخر، فقد واجب هذا النمو الاقتصادي تطور صناعة الخدمات المالية أيضاً بشكل سريع، إذ تجاوزت البنوك في تقديم عروض شعبية على قروض المستهلك بمعدل فائدة يترواح بين 6% إلى 3% (حسب السجل الإداري خلال), ويفترض سداد تصل إلى 5 سنوات، ويشير البنك (المصرف الشهير في السعودية): تحويل الديون، والتاريخ الإداري الجيد للعمل، وأن تكون جهة العمل ضمن الجهات الموقعة على التعامل مع البنك. وكانت أعراض خفض معدل الفائدة على الإشراف: تحريك السوائل المبنية لدى البنوك التي لا تقدر عليها أي تكاليف، ومحاولة الفوز بأد من العملاء وخاصة بعد انتهاء سداد القروض المورغمة بقرض الأسهم، وتحويل الرواتب للبنك.

ولا تتغيير الجدارة الإدارية على الأفراد، بل أنها مصطلح يشمل أيضاً الدول ومنشآت الأعمال، غير أن اختلاف المتغيرات الحاكمة لكل من الحالات الثلاثة جعلها تعلق في البحوث التطبيقية كل على حدده وبرس الباحث

الحالة الأولى المعروفة باسم القروض الشخصية (قرض المستهلك/قرض الأفراد/قرض التجزئة).

2-1 ملخص البحث

في ضوء العرض المتزايد لخدمات التمويل في البنوك ومع التغيرات الملحوظة لسعر الفائدة على الإشراف في السنوات الأخيرة وارتفاع نسبة المعينة، ينكمس آلاف الأفراد يومياً للبنوك بطلب الحصول على قروض جديدة. فيما يليس أن ينتهي العمل من سداد قروض القديم (وفي بعض الأحيان نصفها)، إلا أن تتحمل البنك بخصوص حصول على قرض جديد، ومع تحقيق البنوك الأرباح من رواء هذه العملية، إلا أن خسائرها أيضاً كانت جسيمة بسبب عدم قدرة بعض المقررين على السداد نتيجة التقسيط الخاطئ للإفلاف. فقد ارتفعت خسائر البنوك السعودية بسبب القروض الشخصية من 10% (10 مليارات ريال) في عام 2008 إلى 11% (47 مليار ريال) عام 2014.

وتتاح البنوك إلى نظام تلبية تلك الطلبات أو رفضها. وعلى الرغم من أن غالبية البنوك العالمية تستخدم الآن الأنظمة الإلكترونية في تقديم الالتزام استجابةً للرقع التي أقرتها لجنة بانكس للبنوك المركزية، إلا أن هذا التقييم في بعض البلدان العربية مازال يعتمد على الأحكام والمعرفة الشخصية والطرق الإحصائية.
التحليل الإحصائي لبيانات طالبي القروض لتقديم جهودهم الادمائية

التقليدية. وليستابة لذلك، يجب على الجهات المانحة للائتمان استخدام وتطوير نماذج تقييم الائتمان إلى جانب الطرق الإحصائية التقليدية لدعم قراراتها الإدارية بهدف زيادة الدقة المتعلقة بقرار منح القروض، بحيث يمنح الائتمان المتقدمين الأكثر جدارة (وهو ما يؤدي لزيادة الربح)، ويحجب عن المتقدمين الأقل جدارة (وهما يؤديتا لخسارة).

ويعد تقييم قرارات خطر الائتمان عملية معقدة بطبيعتها بسبب العلاقات غير الخطية بين المتغيرات المستقلة التي تتفاعل مع بعضها البعض إلى جانب أشكال الخطر المختلفة التي تتضمنها كل عملية. وعلى الرغم من أن هذا التقييم قد يتطلب بعض المصاعب التي يصعب قياسها (النزاعات المفاجئة أو الإصابة)، إلا أن بعض القروض السيئة يمكن تجنبها باستخدام أساليب تقييم تتوقف فيها قدرة أكثر على تميز خلل الحالت العاطفي، وهو أمر غريب للغاية.

3- أهداف البحث

يهدف البحث في تقدير نسبة التعلم في سداد القروض لأنه يتناول نماذج تقييم الائتمان وفق أحدث طرق التنبؤ في البيانات. يمنع تطبيقها على معظم البنوك بعد، ولا شك أن ذلك سيؤثر بالرغم -في حالة إقرار تلك الطرق- على من:

- المقرض، يوضح أسباب رفض قروضه بدل من الموافقة عليه على الرغم من عدم جدارة، وتعرضه، واستدلال القرض القديم بأي جزء منفقه قدره في الدائرة، كما يساعد هذا التفسير أيضًا في تحديد أوضاعه عند إصداره ثانية بطلب آخر للحصول على قرض.

3- الهدف: يخصص قرار الائتمان استعدادًا لأحدث طرق التنبؤ في البيانات بدلاً من الاعتماد على الأحكام الشخصية والطرق الإحصائية التقليدية.

ويتأخر البحث مع الجهود الحالية للشركة السعودية للمعلومات الإدارية لمؤسسة النقد العربي السعودي التي تقوم بحساب الجدارة الإدارية لطالب القروض من قائمة بيانات موحدة للقرض الشخصية.

4- أسئلة البحث

1. لماذا يتم الموافقة للبعض على الائتمان في عراض دقيق بعد الانتهاء من تسجيل بياناتهم على الإنترنت، بينما ترفض طلبات البعض الآخر؟

2. وماذا يحول شخص ما على قرض معدل فائدة 7%؟ بما يُحول على شخص آخر معدل بأضعاف ذلك؟ وما هي أحدث طرق تقييم الائتمان؟ كيف يمكن التفكيك بين واقع البيانات التجريبية لتحقيق أفضلية إيجادا على القروض الأخرى؟ وكيف يتم تحديد حساب الدرجة الإدارية لطالب القروض الجديد كمثلكة تصنيف تنبؤية؟ وهل تحتوي قاعدة البيانات الموحدة على كافة المتغيرات التي تفيد في تقنيات الائتمان وتقدير النموذج؟

5- أهداف البحث

يهدف البحث إلى:

1. المقارنة بين بعض قواعد بيانات القروض الشخصية الموحدة لتعزيز على المتغيرات التي يجب دراستها.

المجلة المصرية للسكان وتنظيم الأسرة - المجلد رقم 48 - الحد الأول - يونيو 2015
التنقيب في بيانات طلابي الفروض لنادي الإقرارات من خلال طرق الإحصاء اللوجستي وشجرة التصنيفات.

وتقدم نموذج أثاث متكامل يتم من خلاله التنبؤ بالأداء الإثاثي لطلاب الفروض.

1-3 تنظيم البحث

نُظم البحث في 7 فصول كالتالي، خُصص الفصل الأول لقائمة البحث والثاني عبر تحضير التصوير، والثالث لوصف البيانات والمتغيرات، والرابع لسرد الدراسات السابقة والخليجية، والخامس لعرض النتائج التجريبية للبحث، في حين خُرجت الخلاصة والتوصيات في الفصل السابع والأخير.

(2) تصميم البحث

يدرس البحث ويقارن فعالية الإحصاء اللوجستي و3 نماذج نشاطة الفروض على (chi squared, entropy) (reduction, and Gini reduction) بسماح البديل إذا كان طلاب الفروض الجيد سيساعد له ميعادل في سياقه. وتحتوي مجموعة البيانات على 460 ملتزم، وتعتمد عليها الفروض الجيدة في 74% حالة (80%) والروض السيئة في 11% حالة (20%). وقد أجريت المعايير التكيفية لتقييم مجموعة البيانات الكاملة إلى مجموعة التدريب والاختبار، وقد تمت مجموعة التدريب 410 حالة (70%) منها 2329 قرض جيد و821 قرض سيء، كما تمت مجموعة الاختبار 190 حالة (30%) منها 1437 قرض جيد و453 قرض سيء. وسُجِّل النتائج لـ 2 نقاط فصل احتمالية probability cutoffs: 0.0, 0.3, 0.7, 0.9. أظهرت النتائج أن جزء الفروض قوي وكفاء في تعريض كل من الفروض الجيدة والسيئة في مجال الإثاث الإثاثي في ظل اختبار التلبس أن كل الفروض السليمة في مجموعة البيانات الجيدة ومستديهي تمديد الإثاث. ويتم البحث جزئياً ليشمل التحليل وتفسير النتائج ورسم النماذج، ويساوي في الإثاث والملاحظات العامة النتائج التي تظهر في الدراسات السابقة.

وتعطي نتائج النموذج الخلاصة الإثاثية ما يعرف باسم بطاقة الدرجة scorecard، وهي النتائج من عينة الحاصلين بالفعل على قروض سواء سدتها أم تعرروا في سياقه. ولحاسب الدرجة الخاصة بطلاب الإثاث الجيد، تُذكِّر بيانات التحليل (المتغيرة الجدد) لتصنيف في أحد الأحيان السريعة المشاهدة سيساعد/سستر) ومن ثم تحديد درجة الالتباسية، ويؤخذ النموذج الخلاصة الإثاثية لكل خاصية مقاسة لطالب الفروض، ثم يتم جمع تلك الدوافع للحصول على درجة كلية تغطي بناء عليها -المقارنة مع عتبة معينة بحددها threshold البكالوريا. ومع ذلك يجب تجاهل المتغيرات متعددة التصنيفات والمتغيرات المستمرة إلى تركيب ثاني لقياس نفس الأنماط. ثم وفقاً للتحليل الاستثناوي لمتغيرات الإثاث، مثبناً بالأختبار متعدد المتغيرات (الانحدار اللوجستي وشجرة التصنيفات). ويعود بناء النموذج (تربيته) من داخل العينة، ويعتبر ويان نتائج معمل للتصنيف الصحيح، لتعتبر العملية مع اختباره من خارج العينة.

(3) وصف الفروض والمتغيرات

نُشرت نتائج التصنيف التتابع (تصنيف ما إذا كان طلاب الإثاث الجيد: جيد بالإثاث أم غير جيد به)، يجب أولاً تتبع صورة عامة عن المتغيرات المستقلة التي يحتوي تأثيرها عليها. وقد تم عرض ذلك في القسم الفرعي 1-2 قبل أن يتملألانن لوصف قاعدة بيانات الدراسة في القسم الفرعي 1-2 مورداً بشروط قواعد البيانات المحددة في المادة المصرفي للكيماويات ومثلية الإلثام والمجلة للسنة 2015 -عدد الأول- يوليو لمنية 48.
تم مراجعة 7 دراسات أجريت على تقييم اثنان المستهلك في أوقات مختلفة وأهمكم مختلفة (وهما ما يحقق الهدف الأول)، و-boot جدول 1: المتغيرات التي يُحمل تأثيرها على الجدارة الاجتماعية لقروض المستهلك وفئاتها ومعقدها وعمقها والمعلومات والبيانات التي تم جمعها.

<table>
<thead>
<tr>
<th>الدراسات</th>
<th>المتغيرات</th>
<th>القيم</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>C</td>
<td>B</td>
</tr>
<tr>
<td>- 1</td>
<td>حرص الحساب الجاري متغير كمي يحوَّل لمتغير تصنيفي</td>
<td></td>
</tr>
</tbody>
</table>
| - 2 | فترة الوداع بالشهير متغير كمي بتقسيم Swagger 
| - 3 | الالتزام بالدائنين مناقشة كل حالة التقول |
| - 4 | تعزية الدورانية أو العلاج أو النقل أو التعليم أو ... متغير كمي |
| - 5 | قيمة المخاطر/ihanab الأداء |
| - 6 | الأصول المتاحة متغير كمي |
| - 7 | سنوات الخبرة في الوظيفة المالية |
| - 8 | نسبة الخصم من الدخل الاجمالي متغير كمي |
| - 9 | النوع/الحالة الاجتماعية |
| - 10 | مدة الإقامة في المكان الحالي |
| - 11 | العمر |
| - 12 | وجود اثاث آخر |
| - 13 | نوعية السكن |
| - 14 | الديون المتباقية |
| - 15 | عدد الديون السابقة |
| - 16 | نوعية الديون |
| - 17 | عدد الدفعات |
| - 18 | وجود الديون الأخرى |
| - 19 | الدخل الشهري |
| - 20 | الفائدة السكنية |
| - 21 | مبلغ القرض |
| - 22 | الائتمان |
| - 23 | دخل عائلة |
| - 24 | مبلغ القرض الشهري |
| - 25 | المصروفات الشهرية |
| - 26 | مبلغ القرض |
| - 27 | درجة تطبيقية |
| - 28 | الدخل الائتماني |

المجلة المصرية للفضاءة والتنظيم الإداري - المرجع رقم 418 العدد الأول - يونيور لسنة 2015

- تشير إلى دراسة: (2007)
- Koh et al. (2006)
- Dinh and Kleimeir (2007)
- Crook et al. (1992)

- تشير إلى دراسة: (2007)

- وقد قيل (2007)...

- متغير تصنيفي واحد (كما هو الحال في معظم الدراسات) هو الأفضل. كما أن مستوى التعليم الذي يظهر في نفس الدراسة (كيوثر للدخل) يعد مراقبة نوعية وممارسة في الدراسات الأخرى، ويُفضل أن يعتبر البنك على متغير نوعية الدولة (كيوثر أفضل الدخل) بدلاً من اعتماده على متغير مستوى التعليم، لأنه من الممكن أن يحمل شخصون درجة الدكتوراه وهو عاطل عن العمل، وأقل الربح متغير حساب الإدار (النواب المرتبطة) لأن ذلك أمر محدود للغاية في المجتمع السعودي.

وينضج من العرض السابق أن دراسات تقييم المشاركة قد اتفقت على معنى تأثير الخمس متغيرات الأولى فقط، لكنها لم تتلقى بشأن الأمور وعدد المتغيرات المستقلة التي يجب تحت تأثيرها على متغير حالة الفقر، وقد كان ذلك طبيعيًا لأن تلك المتغيرات تتوافر بدورها على الجوانب السلوكية والظروف الاقتصادية للبلد التي أجريت فيها تلك الدراسات والقيود الاقتصادية المفرطة. ويفيد جدول 1 في التعرف على المتغيرات الأكثر استجابة في البنوك التي تطبق نظام تقييم المشاركة وهو ما يخدم إنشاء قاعدة بيانات موحدة (أو تحديد قاعدة البيانات الحالية).

2- منافع قواعد البيانات الموحدة

- يحقق وجود قاعدة بيانات موحدة منافع التالية:

- الحد من مشكلة القيمة المقيدة: يُرجع وجود القيمة المقيدة إلى عدم كفاية نظام المعلومات ببعض النروج، و/أو إعمال بعض الموظفين في إدخال تلك القيم للنظام، و/أو عدم رغبة بعض المتقاضين في الإجابة عن الأسئلة المتعلقة بتلك القيم. ويقوم بعض الباحثين بحل تلك المشكلة بحلف السجلات التي تحتوي على تلك القيم وهو ما يؤدي لاختيار جم العينة والوصول لنتائج غير موثوق فيها، أو باستهداف كافة الحلول (المتغيرات) التي تحتوي على القيم المقيدة وهو ما يؤدي لنقص دقة تصنيف النموذج، إلا أن وجود قاعدة بيانات موحدة حديثة - يتم التسجيل فيها عبر الإنترنت وتشتهر ماه جميع الحلول لิกا النتائج - يقضي على هذه المشكلة تماماً.

- التضاء على مشكلة التحيز: تنتج مشكلة التحيز في قواعد البيانات القديمة في السجلات الخاصة بالبنوك بسبب عدم تمثيل العينة للمجتمع لاحتراءها على بيانات الطرق المقيدة فقط دون المعرفة. وفي حالة وجود قاعدة بيانات موحدة، فإن بيانات طالبية الطرق الذين رفضت طلباتهم سوف تكون متناحرة للدراسات القادمة إلى جانب بيانات الطرق المقيدة وهو ما يقضي على مشكلة التحيز.

- وضع درجات مدرسية أساتذة طالبي الطرق: بدلاً من تصنيف كل تلك السمات بإمكانيتها درجات بالطريقة التي يراها مناسبة.
3-3 قاعدة بيانات نادي الإقراض

تحتوي قاعدة البيانات الخام لنادي الإقراض (خلال فترة الدراسة) على 4936 حالة و 101 متغير. وبعد استبعاد السجلات والحوال ذات القيم المفقودة والمتبقيات التعريفية (كالعنوان ورقم التعريف العميل)، فقد قُثص حجم مجموعة البيانات إلى 2490 حالة مسجلة على 22 متغير. وقد استُخدم متغير حالة الفقر الذي يحتوي على 6 تصنيفات - فروض مستقرة (11.1%)، ومتأخرة (1%)، default (4.4%)، ومتأخرة أقل من شهر (3.3%)، ومتأخرة أكثر من شهر (1.3%)، وجارية (38.5%)، ومقداراً بالكامل (48.3%) - لتوزيع الفروض إلى جيدة وشائبة. وكما هو مبين في شكل 1، وجدت 2 و 3، فقد صنف الفقر على أنه سيء إذا وقع في أحد التصنيفات الأربعة الأولى (5600 حالة بنسبة 13.5%)، كما صنفت الفقر على أنه جيد إذا وقع في أحد التصنيفات الأخيرين (3870 حالة بنسبة 88.8%). ومعنى آخر، قد حُول متغير حالة الفقر في جدول 2 إلى متغير ثاني في جدول 3 ليصبح الأخير هو المتغير التابع للدراسة. وقد طُبقت المعايير العرفية وخصوص 50% من العينة الكليّة لمجموعة التدريب والمجموع الشريفي لمجموعة الاختبار. وتمّت مجموعة الاختبار والاختبار أيضاً، إضافة إلى فروض جيدة وفروض سيئة، وقد بُنيت النماذج من مجموعة التدريب، واختبر أدائها من مجموعة الاختبار.

يلتزم بالقواعد المنشئة للأعمال وخدمات المصارف، والشروط الأخرى، وقد تمّ تدريس النماذج المذكورة في هذا الدراسة.

المنحة التقديرية للدراسات العليا والبحث العلمي، والموضوعات ذات الصلة.

المجلة المصرية للدراسات التربوية، مجلة الآمنة، vol. 48، fasc. 1، 2015.
المتغيرات المستقلة 21 مقياس متغير تابع واحد تصنيفي ثنائي. ولم يمكن دراسة آثار بعض الخصائص الشخصية مثل: النوع والدالة الاجتماعية ومستوى التعليم والعمل والجنسية - على الرغم من أهميتها المحتملة في المنطقة العربية - على الجدارة الاجتماعية. لفروض المستهلكين نعد ووجدوا في قاعدة البيانات.

(3) الدراسات السابقة

كان طبيعياً أن تركز الدراسات الأولى لتقدير الالتزام (مع بداية التسعينات) على تأسيس قواعد البيانات وتحديد المتغيرات المؤثرة على الجدارة الاجتماعية باستخدام الطرق الإحصائية التقليدية للتحليل متعدد المتغيرات. فقد اعتمدت بعض هذه الدراسات في تقييم الالتزام على تنفيذ التمثيل الذي يطلب فروضًا إحصائياً صعب قبوله في الواقع المعملي. لذلك، فقد استدلال الدراسات التالية لتحليل التمثيل بالالتزام بالالتزام لإنها أقل تقييدًا. كما اعتمدت هذه الدراسات أيضاً - ولكن بدرجة أقل - على الخوارزميات الجينية وطريقة أقرب الحيران والرموز الخطية والأنظمة (Crook et al., 2000) ومن أقدم وأشهر الدراسات التي تتعلق بهذا الموضوع جدارة الإثارة دراسة (Thomas, 1992) ودراسة (Fahrmeir and Tutz, 1994) (بما تحدد أساليب تعلم الآلة (machine learning)) على شكل تقييم الالتزام وتقييم المخاطر الاجتماعية.


فحص آداء النماذج المختلفة للجدارة الاجتماعية باستخدام بيانات 37 متغير.
ولم يعثر الباحث على أي بحوث إحصائية تناقش مشكلة تقسيم الامتنان في المنطقة العربية في أي من قواعد البحث، وهو ما يضفي على هذا البحث أهمية خاصة. كما يتميز البحث بتجنبه لمشكلة التحيز الذي وقع فيه معظم دراسات تقسيم الامتنان - بسبب اعتمادها على بيانات تقرير المنتظم فقط دون المعرفة، حيث تعزز قاعدة البيانات من امتنان أيضاً لمن لم تتضمن خصائصهم مع السياسة الامتنانية لسيك.

(5) منهجية البحث

ابتداء الدراسة، وهو جزء من البحوث السابقة (استرداد القاعدة) من أشهر وأحدث طرق التنبؤ في الدراسات عند دراسة مشكلة تقسيم الامتنان بخصوص تقييم طالب التقرير الجديد في أربع فئات: جدري بالانتمان، أقوى جدري، ويجيب هذا القسم على سؤال البحث الثالث.

1- الإعداد الโลجستي

استخدام تقسيم الدراسة المناسبة لتقييم الامتنان القدرة المستهدفة من بين طرق الامتنان المتعدد وتحليل التفاوت والخوارزميات الجينية وطرق أخرى للجزء وال الصفحة الخفية، وتقوم المتكافئة مع الامتنان الโลجستي في هذا البحث وفقاً للمنهجية التالية:

- تحديد المتغيرات وتكوينها: بدأ تحليل الامتنان الโลجستي تحديد المتغيرات النابع والمتغيرات المستقلة. ولا يوجد خلاف في دراسات تقييم التقرير الشخصية على أن المتغير التتابع هو "حالة التقرير" الذي يأخذ شكلين ثانياً، حيث تُحسب حالات التكرار بأنها رجعي (B) حيث أن عدد الحالات المعزولة يُحسب من خلال عينة المتغيرات الفعلية بمراعاة التزايد لمتسددا، وكم هو مبين بجداول 2، فقد تم تكوين المتغيرات التصنيفية لإعطاء فئاتها درجات من الأعلى (حسب عدد فئات المتغير) إلى الأدنى (الدرجة 3)، مع تخصيص الدوارة الأعلى (الدرجة 1) للفرع ذات الجدارة الأعلى (التي تحتوي على أعلى عدد من المتغيرات). وتم التوصل في هذه المرحلة إلى مجموعات متغيرات متميزة، حتى مع البنوك التي لا تطبق نماذج تقسيم الامتنان، لأن تلك البنوك لن تتمكن من صنع قرار الامتنان في حالة عدم وجود قواعد بيانات موحدة - مما يجعل كبار القروض في معظم البنوك من مجموع المتغيرات و/أو يوجد التحيز.

- التموئج: ولكن احتمال التكرار في السداد لطالب التقرير الجديد يكون غير قابل للمشاهدة قبل حدوثه، ويستطع أسلوب الامتنان الโลجستي أن يتغلب على ذلك بتقدير هذا الاحتمال كما التالي:

\[
\pi_j = \frac{1}{1 + e^{-x_j}}
\]

حيث تشير \(\pi_j\) إلى الدرجات (دقيقة للتغييرات المستقلة المشروعة لخصائص طالب التقرير مرجحة مع المعدلات التالية:

\[
z_j = \beta_0x_1 + \beta_1x_2 + \ldots + \beta_kx_k
\]

المجلة المصرية للسلاسل وتتولأم، للسلاسل 48 العدد الأول - ببليو بسنة 1285
وفقاً للمعادلات السابقتين، فإن الدرجات الكبيرة تعني احتمالات تعترب صاحبة.

- اختيار المتغيرات ودقة التنبؤ: يتم إدخال جميع المتغيرات الميدانية للمعادلة 1 وتشفيتها للحصول على أقصى دقة تنبؤية للنموذج بتطبيق الاتجاه المتدرج للأمام (forward stepwise) وإضافة المتغيرات للنموذج (ب الخدمات الاتجاه المتدرج للأمام) ونحوه. في هذا الاختبار، يتم تطبيق الاتجاه المتدرج لرضا المتغيرات من النموذج وがかか من المتغيرات اليها. ونتيجة هذه المرحلة، يتم تعديل المتغيرات الحالية والوصول لتقديرات لمعالجتهم./!

- اختبار الفكرة التنبؤية: ويتخطى العملية بالفقه التنبؤية للتسامح، ويعني أن يكون ذلك من خارج العينة -outسعتي مستويات دقة مبالغ فيها. ويتضمن ذلك مقارنة:

**جدول 2: مصفوفة التباين نظرية**

<table>
<thead>
<tr>
<th></th>
<th>المتغير (سيء)</th>
<th>غير متغير (جيد)</th>
</tr>
</thead>
<tbody>
<tr>
<td>المتغير المشاهد</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_0$ للقرض الجيدة (خطأ النوع الأول)</td>
<td>$G_0$</td>
<td>$G_0$</td>
</tr>
<tr>
<td>$B_0$ للقرض السيئة (خطأ النوع الثاني)</td>
<td>$B_0$</td>
<td>$B_0$</td>
</tr>
<tr>
<td>الكل (G0+B0)</td>
<td>$B_0/(G_0+B_0)$</td>
<td>$G_0/(G_0+B_0)$</td>
</tr>
</tbody>
</table>

الحساسية = $B_0/(G_0+B_0)$
العديدية = $G_0/(G_0+B_0)$

وتقارن المصفوفة التشغيل المشاهد والمتوافق من خلال 4 تصنيفات:

- إثبات صحيحة على القطر الرئيسي $G_0$: تصنيف النموذج للقرض على أن جيد وهو جيد بالفعل. $B_0$
- إثبات خاطئة على القطر الرئيسي $G_0$: إذا كان النموذج الجيدة التي صنفها النموذج (بالتالي) $G_0$ إلى عدد القروض الجيدة التي صنفتها النموذج (بالتالي) على أنها سيئة أو خطأ من النوع الأول، كما يشير $B_0$ إلى عدد القروض السيئة التي صنفتها النموذج (بالتالي) على أنها جيدة أو خطأ النوع الثاني.

التحليل الإحصائي لبيانات طابع الفروض لتقديم جداولهم التفصيلية

10

4-5 شجرة القرار

بينما ينتج الليمسي الدرجات أولًا ثم يصنفها استنادًا لقاعدة طبيعية ما، فإن نماذج شجرة القرار

تبدأ بتصنيف المشاهدات في مجموعات ثم ينتج درجات كل مجموعة (100، 2003، p. 3)، وتتضمن نماذج

الشجرة إلى: شجرة الليمسي (ويكون فيها المتغير التابع كيميائي)، شجرة التصنيف (ويكون فيها المتغير التابع

تصنيفي)، وهي حالة هذا البحث.

ويعدّ تصنيف شجرة القرار من أبسط أساليب التصنيف، كما أنه من أكثرها قبولاً وأسرعها (أحد أهم

استخداماً بالنسبة لمشاكل تقسيم الائتمان. إذ أن تحليل شجرة القرار للمتغيرات الأكثر أهمية في التأثير على المتغير

التابع (تقنية الائتمان)، يمكن الباحثين من استخدامها كتقنية لاختيار المتغيرات إلى جانب استخدامها كتقنية لبناء

الشجرة (Linoff and Berry، 2011، p. 199)، ولحلة الزياد ونماذج أخرى من النماذج

والعدالة الداخلية، والعدل الخارجي (الأطراف أو الأوراق). وتكون عقدة الجذر في الأعلى بلا حفوف داخلية، والأوراق

بالأسفل بلا حفوف خارجية. وفي الشجرة الثنائية، يخرج من عقدة الجذر فرع ينفصل إلى عقدتين داخليتين، وينفرع

من كل عقدة فرعين إلى أن يصل للعقد الطرفية. فكل عقدة داخلية يدخل عليها فرع واحد وخرج منها فرعين، وكل

عقدة خارجية يدخل عليها أيضاً فرع واحد ولكن لا يخرج منها أي فرع. وتنمو كل عقدة طفيفة عنوان فئة ما، أما

الفرع فيحتوي على حالات اختيار السمة التي تقسم لاحالات منفصلة بخصائص مختلفة، وتوزع الخوارزميات الكتلة

استراتيجيات البحث للوصول إلى دقة معقولة، بحيث يتم التوصل للحجم الأمثل للشجرة بلا نمو زائد أو نقص بأسرع

ما يمكن.

وتركز خوارزميات بناء شجرة القرار على معالجة مسأليتين: إنتاج أشارات التكرار، ومتبيني وقف هذا

التقسيم. ينصح أفضل انتظام على درجة الاضطراب أو عدم الالتباسية، ومن أمثلة مقاييس الاضطراب (عدم الانسجام):

\[ Entropy(t) = -\sum_{i=0}^{c-1} p(i \mid t) \log_2 p(i \mid t) \]  

\[ Gini(t) = 1 - \sum_{i=0}^{c-1} [p(i \mid t)]^2 \]  

\[ Classification\; error(t) = 1 - \max_i [p(i \mid t)] \]

حيث تشير 0 إلى عدد الفئات، 0 = 0 في حسابات الإحصائي، و 0 في حسابات الإحصائي، فإذا خُف حت الزخم t، فإن الكسر يُست killings

عدد العقدة t. وإذا خُف حت الزخم t، فإن الكسر يُست killings p_t = 1 - p_0

حيث (p_0، p_t)

إذا تم التوصل للتقسيم الأمثل النهائي، فإن شجرة الليمسي تنتج قيمة مقدرة ل marketer المشاهدة في المتغير

التابع / يُساوي متوسط قيمة المتغير التابع في المجموعة التي تنتمي إليها المشاهدة رقم t:

\[ \hat{y}_t = \frac{1}{n_m} \sum_{m=1}^{n_m} y_{lm} \]
 حيث تشير $m$ إلى رقم المجموعة التي يجب لها القيمة الموقعة، وتشير $\pi_i$ إلى حجمها. أما في حالة شجرة التصنيف، فتُحسب تلك القيم بمعلومية الاحتمالات المقدرة لانتشار المشاهدة لمجموعة معينة. ويمكن احتمال النجاح في حالة التصنيف التالي:

$$\pi_i = \frac{1}{m_n} \sum_{l=1}^{d_m} y_{im}$$

وتأخذ المشاهدة $y_{im}$ القيمة 0 أو 1، لذلك فإن الاحتمال المقدر يناظر نسبة النجاح في المجموعة $m$. ومن الجدير بالذكر أن كلاً من $\pi_i$ يُعد ثابتاً لكل المشاهدات.

ومعبرة عن جودة شريط الاختبار، تكون درجة الاضطراب في العقد الأصلية child nodes (العقد المنتشرة منها)، وكلما أرتفعت قيمة المكسب $\Lambda$، كلما كان التخفيض أفضل.

$$\Delta = I(\text{parent}) - \frac{k}{N} \sum_{j=1}^{k} I(v_j)$$

حيث تشير $I(\text{parent})$ إلى مقياس الاضطراب لعقدة ما، $N$ إلى عدد الحالات الكلية في العقد الأصلية $A$، و$N(v_j)$ إلى عدد الحالات المقترنة بالعقدة الطفل $v_j$.

وقد تم بناء شجرة القواعد في هذا البحث (أو السماح لها بالنمو) بثلاث طرق:

1. خوارزمية CHAID: أي واكتشاف الفاعل التقني باستخدام مربع كاي
   وتشتهر هذه الخوارزمية المترتبات المستقلة صاحبة أقوى تفاعل مع المتغير التابع.

2. خوارزمية CHAID: أي شجرة الإحصائية السريعة غير المحدودة الكثيرة.
   ويتم فيها التصورات إلى شرائح متراجعة بقدر الإمكان بالنسبة للمتغير التابع، بحيث تحتوي العقدة الطرفية على نفس
   القمية للمتغير التابع.

Quick, Unbiased, Efficient

3. خوارزمية QUEST: أي شجرة الإحصائية السريعة غير المتغيرة الكثيرة.
   وهي تتجنب تعيين الطرق الأخرى للمترتبات المستقلة صاحبة أكثر عدد من الفئات.

Statistical Tree

وترتكز معايير التقسم المبنية على قياس $\chi^2$ (اللبانة). إذ تُعد إحصاء الاختبار $\chi^2$

للانتشار بين قنوات المتغير التابع والعقد الفروع. ويمكن استخدام هذه الإحصاءات في الحكم على جدارة التقييم بقياس

الفرق بين تكاثرات الخصائص المتوقعة والمتوقعة بغض استقلال عند الفروع عن العقد المستهدفة. وقد تُفرض في هذا

البحث أن مستوى المعيون 0.0500. ولمزيد من التفاصيل حول شجرة القواعد، يمكن الرجوع إلى واحد أو

Linoff and Berry, 2011; Giudici, 2003; dtreg Software: http://www.dtreg.com;

andSAS Enterprise Miner: http://www.sas.com

المجلة المصرية للسكان وتقييم الإسرة - المجلد رقم 48 العدد الأول - يوليو سنة 2015
التحليل الإحصائي لبيانات طالبي القروض لتقديم جزائهم الاجتماعي

(1) النتائج التجريبيّة

يدرس البحث أثر 21 متغير مستقل (16 كمي، 5 تصنيفي) على متغير تابع واحد تصفيفي (أي حالة القرض). ووفقًا للمنهجية البحث، فقد بدأ التحليل بعمل التحليل الاستكشافي بغربيّة دراسة أثر كل متغير مستقل على المتغير التابع من خلال التحليل ثنائي المتغيرات وتصنيف مجموعة البيانات. استخدم السجلات ذات القيم الماركة والمقدّمة. تمّ تجاوز التغيرات المعنوية وادخالها لنموذج التحليل متعدد المتغيرات، دفعة واحدة.

6.1 التحليل الاستكشافي

أجري التحليل الاستكشافي باستخدام التحليلات تحليلات متعددة لأن مختلف البيانات ووصف أنماط الدراسة. كما أجري التحليل ثنائي المتغيرات لبحث شدة الروابط الممكّنة بين كل متغير مسرّ والمتغير التابع. ويدعو اختيار المتغيرات.

6.1.1 التحليل أحادي المتغيرات

- مبلغ القرض، ومجال الدفعات، ومبلغ الصلب من أصل القرض.

يتجزّر مبالغ القرض الشخصية في مجموعة بيانات الدراسة بين $500،000 و$3،000،000. وقد خُصصت سجلات القرض الكبيرة (التي تزيد عن $3،000،000) من التحليل لأنها تكون تأثيرًا عن دارة القرض الشخصية، ولكن عندنا محدودة لأنها تمتّع قيمًا. ويفصل الحكم على مجمل نشاط البنوك في فترة الدراسة؛ حيث أن المبالغ المقدّمة الكاملة لم تتجاوز 48،000،000، في حين أن 4،000% من القروض لا زالت جارية. 13% من القروض التي استُخدِمت أُنتجت على أنها متعطّرة/دراستها تلك المبالغ القروضية والمتقاربة مع نظائرها في الدراسة الكلية، اتضح أن إجمالي القروض الخاص بالمتعرّبين بلغ $6،400،000، بضعة أضعاف ($6،400،000) من إجمالي مبالغ القرض، منذ منها 2،500،000؛ لن تكون نسبة الخسائر 40% (من إجمالي أصول القروض المتعطّرة)، و 40% (من إجمالي أصول القروض ككل)، ولم يكن ذلك بسبب جمع البيانات المقدّمة التي ظهر أمام متغير الصلب من أصل القرض في الجدول الأولين بالمخرج، لأن التسنين السابقين لم تتجاوز بشكل جوهري، بعد حذف السجلات الخاصة بتلك البيانات المقدّمة (أصبحت 72،4%) على الترتيب، وباستثناء مبالغ القروض نسبيًا من الأصل، بعد تحويلها إلى الشكل التصنيفي على النحو المبين بجدول 4- بحسب حالة الصلب لمجموعة البيانات ككل، اتضح أن التغيّر يتناسب عكسيًا مع مبالغ القرض، فهو أكبر للقرض السعيد المعادي في القروض المتوسطة، ثم الكبيرة.

• الدخل السنوي، ونسبة الدخل إلى الدخل، والتحقيق من الدخل

يحل الدخل السنوي للمعتزرة، وطالب القرض أهمية كبيره في دراسة تقدم الانتظام. إذ يُعدّ ذلك الدخل إجازة لمنح القرض لطالبه، ومؤشرًا لاحتمال الصلب المترافق للمعتزرة. يتراوح الدخل السنوي لأفراد عينة الدراسة بين $1،896،000 و$750،000 بمتوسط 94،594. وقد تم حذف بيانات 11 حالة عند هذه المرحلة بسبب دخولهم في التحقيق المتزنقة التي تتجاوز 500،000، فإنها كأنها فشلت على ذقة التحليل. ويدعو أيضًا نسبة الدخل إلى الدخل من المؤشرات الهامة لمنح الانتظام، أي نسبة الدفع المبهر إلى الدخل الشهري الخاضع للتصريف. إذّما انسحبت تلك النسبة، كما شكل ذلك فاقدًا جيدًا لمنح الانتظام، وقد تراوح نسبة الدخل إلى الدخل بين 10% إلى 30% لكل من العينة الكلية وعينة المتعرّبين على حد سواء.
وبلغت نسبة الذين لم يحققوا الفائدة من دخولهم السنوي $8.3٪، ثالثاً، من القروض السيئة، وهو ما يفسر
سبب الخسائر الكبيرة التي تعرض لها البنك.
• معدل الفائدة، والضلع الشهري، وأجل القرض، والتصنيف الائتماني، والسياسة الائتمانية

تُفرض معدل الفائدة على القرض حسب تصنيف الائتماني (الدرجة الائتمانية) وأجله (مدته) ومبلغ القسط
ومدى التوافق بين خصائص العمل والسياسة الائتمانية التي يفرضها البنك. عند استكشاف بيانات متغير القسط
الشهري، فإن هناك 7 قيم مرتفعة (تتجاوز مبلغ القرض)، وقد تم استبعاد السجلات الخاصة بتلك الحالات
حتى لا تؤثر على دقة التحليل.

وقد تراوح القسط الشهري الذي يدفعه الم particulière بين 15.7٪واربطة 532.5٪$ ب المتوسط 322.5٪$ (بعد أن كان
قبل الحفظ)، كما تراوح معدل الفائدة بين 5٪ و 34.7٪. وبلغت الفائدة متوسطة الأجل لمدة 3 سنوات
من حجم العينة، وناتج كان لمدة 5 سنوات، وكان المتوسط معدل الفائدة 3 سنوات 11.8٪، وارتفاع متوسطه
بزيادة 4.1٪ من سنوات إلى 14.9٪.

وتتلاقى معظم البنوك -إيجابًا على سائل البحث الأول- على تلبية القروض للعملاء ذوي الدرجات الائتمانية
العليا، والذين أي أن إجابته تقتصر على خصائص هذه القنلة فقط، وهو ما يعرض نتائج نمذجاء لمشكلة التخيز
المتمثلة في عدم شمول خصائص العملاء ذوي الدرجات المرتفعة في البيانات. وتختلف مجموعة البيانات الدراسة
هنا كشرط ورغم خصائص المعاملة، وهو ما يواجهها على مشكلة التحيز المركز، ويتسبب معدل الفائدة عكسًا مع
التصنيف الإحصائي للعملاء، إذ تبلغ متوسطة التصنيف 8 والتي تشير بيانات ل huyện برامج التدريب unfold
3٪ وارتفاع تعريفيًًا بانخفاض التصنيف إلى أن
وصول إلى 0.4٪ للتصنيف 6. كما ينطلق معدل الفائدة أيضاً على مدى متابعة خصائص المدفوعة للسياسة
الائتمانية للعملاء، حيث تبلغ 38.2٪ في المتوسط لم يظهر عليهم الشروط، وتبتكر فعالية التعامل السابقية، كان من الطبيعي أن يكون معدل الفائدة للقرض السيء أعلى منه
للقرض الجيد، حيث تبلغ متوسطها 43٪ مقابل 21٪ على القريب،
والخالصة، أن معدل الفائدة يرتفع بانخفاض التصنيف الائتماني، وعدم متابعة خصائص العملاء مع
السياسة الائتمانية للبنك، وطول أجل القرض، وكبير مبلغ القرض وبالتالي مبلغ القسط، وهو ما يجب على سائل
البحث الثاني.
• الخبرة الوظيفية، وامتثال منزل، وعرض القرض

شكل فترات الخبرة الوظيفية (الذين زادت خبرتهم عن 10 سنوات، والتي قبضت خبرته عن سنة) أكثر الفئات
التي استندت من القروض (23٪)، وبلغت نسبتها من القروض السيئة 43٪. بما مثلاً، فإن الفئات المستأجرين والرهن
العقاري 39٪ من الحالات على القروض في متغيرات المنسق بنسبة 49٪ أيضاً من القروض السيئة. أما
بنسبة من متغير عرض القرض، فقد تعزز المودية أكثر الفئات خطراً على الجدارة الائتمانية، حيث مثلت هذه
الفئة فورها 24٪.43٪.

عدد الاستعلامات، وعدد السجلات العامة المهنية، وعدد خطوط الائتمان الجارية والكلية، والرصيد المتجدد،

وعدد الاستعدادات من الخط المتجدد
على الرغم من تراجع عدد الاستعدادات عن المفترضين بين 0 إلى 33، إلا أنه لم تتجاوز 2 في 95٪ من المفترضين.
كما لم تتجاوز عدد السجلات العامة المهنية
المجلة المصرية للسكان وتخطيط الأسرة - المجلد رقم 48 العدد الأول - يونيو لسنة 2015
يتيح التحليل التالي المتغيرات معرفة كفاءة كل متغير مفسر في تحديد العملاء غير الموثوق فيهم (0-2)، وهو ما يفيد في اختيار المتغيرات فيما بعد في مرحلة التحليل متعدد المتغيرات. وقد تم عمل ذلك باستخدام نسب الأرجحية odds ratios المبينة تنامتها في جدول 2 التالي:

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>type</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Odds ratios</th>
<th>95% C.I for EXP(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>installment</td>
<td>continuous</td>
<td>1563.926</td>
<td>1</td>
<td>0.000</td>
<td>0.975</td>
<td>0.972 - 0.978</td>
</tr>
<tr>
<td>2</td>
<td>total_pymnt</td>
<td>continuous</td>
<td>638.217</td>
<td>1</td>
<td>0.000</td>
<td>1.001</td>
<td>1.001 - 1.001</td>
</tr>
<tr>
<td>3</td>
<td>grade</td>
<td>categorical</td>
<td>580.154</td>
<td>6</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>grade(1)</td>
<td>categorical</td>
<td>379.473</td>
<td>1</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001 - 0.003</td>
</tr>
<tr>
<td></td>
<td>grade(2)</td>
<td>categorical</td>
<td>420.231</td>
<td>1</td>
<td>0.000</td>
<td>0.003</td>
<td>0.002 - 0.005</td>
</tr>
<tr>
<td></td>
<td>grade(3)</td>
<td>categorical</td>
<td>505.005</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000 - 0.000</td>
</tr>
<tr>
<td></td>
<td>grade(4)</td>
<td>categorical</td>
<td>567.103</td>
<td>1</td>
<td>0.000</td>
<td>0.011</td>
<td>0.007 - 0.016</td>
</tr>
<tr>
<td></td>
<td>grade(5)</td>
<td>categorical</td>
<td>531.629</td>
<td>1</td>
<td>0.000</td>
<td>0.033</td>
<td>0.025 - 0.044</td>
</tr>
<tr>
<td></td>
<td>grade(6)</td>
<td>categorical</td>
<td>415.111</td>
<td>1</td>
<td>0.000</td>
<td>0.122</td>
<td>0.100 - 0.149</td>
</tr>
<tr>
<td>4</td>
<td>total_rec_prncp</td>
<td>continuous</td>
<td>529.789</td>
<td>1</td>
<td>0.000</td>
<td>1.001</td>
<td>1.001 - 1.001</td>
</tr>
<tr>
<td>5</td>
<td>int_rate</td>
<td>categorical</td>
<td>479.268</td>
<td>1</td>
<td>0.000</td>
<td>1.583</td>
<td>1.502 - 1.627</td>
</tr>
<tr>
<td>6</td>
<td>purpose</td>
<td>categorical</td>
<td>108.660</td>
<td>13</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>loan_amnt</td>
<td>continuous</td>
<td>51.662</td>
<td>1</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td>8</td>
<td>atm</td>
<td>categorical</td>
<td>67.031</td>
<td>2</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>atm(1)</td>
<td>categorical</td>
<td>56.113</td>
<td>1</td>
<td>0.000</td>
<td>0.236</td>
<td>0.162 - 0.344</td>
</tr>
<tr>
<td></td>
<td>atm(2)</td>
<td>categorical</td>
<td>68.931</td>
<td>1</td>
<td>0.000</td>
<td>0.322</td>
<td>0.246 - 0.423</td>
</tr>
<tr>
<td>9</td>
<td>inq_last_6mths</td>
<td>continuous</td>
<td>51.096</td>
<td>1</td>
<td>0.000</td>
<td>0.889</td>
<td>0.861 - 0.918</td>
</tr>
<tr>
<td>10</td>
<td>pub_rec</td>
<td>continuous</td>
<td>28.331</td>
<td>1</td>
<td>0.000</td>
<td>0.662</td>
<td>0.559 - 0.771</td>
</tr>
<tr>
<td>11</td>
<td>revol_bal</td>
<td>continuous</td>
<td>28.001</td>
<td>1</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td>12</td>
<td>annual_inc</td>
<td>continuous</td>
<td>15.933</td>
<td>1</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td>13</td>
<td>credit_policy(1)</td>
<td>continuous</td>
<td>8.584</td>
<td>1</td>
<td>0.033</td>
<td>0.732</td>
<td>0.594 - 0.902</td>
</tr>
<tr>
<td>14</td>
<td>total_acc</td>
<td>continuous</td>
<td>5.703</td>
<td>1</td>
<td>0.017</td>
<td>0.999</td>
<td>0.980 - 1.008</td>
</tr>
<tr>
<td>15</td>
<td>open_acc</td>
<td>continuous</td>
<td>4.219</td>
<td>1</td>
<td>0.040</td>
<td>1.019</td>
<td>1.001 - 1.037</td>
</tr>
<tr>
<td>16</td>
<td>is_inc_v(1)</td>
<td>continuous</td>
<td>4.617</td>
<td>1</td>
<td>0.032</td>
<td>0.691</td>
<td>0.801 - 0.990</td>
</tr>
<tr>
<td>17</td>
<td>term(1)</td>
<td>continuous</td>
<td>9.990</td>
<td>1</td>
<td>0.002</td>
<td>7.747</td>
<td>6.600 - 9.098</td>
</tr>
<tr>
<td>18</td>
<td>emp_length</td>
<td>categorical</td>
<td>17.765</td>
<td>10</td>
<td>0.059</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>cli</td>
<td>continuous</td>
<td>3.369</td>
<td>1</td>
<td>0.066</td>
<td>1.007</td>
<td>1.000 - 1.014</td>
</tr>
<tr>
<td>20</td>
<td>home_ownership</td>
<td>categorical</td>
<td>6.249</td>
<td>4</td>
<td>0.181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>revol_util</td>
<td>continuous</td>
<td>0.343</td>
<td>1</td>
<td>0.855</td>
<td>1.000</td>
<td>0.988 - 1.002</td>
</tr>
</tbody>
</table>
ويتعدى عدد هذه المرحلة أن المتغيرات المشتركة للتحليل متعدد المتغيرات كانت 12 (بعد حذف إحدى
 chicago متغير مبلغ القرض المستمر أو التصنيف)، وأن المتغيرات التي غادرت التحليل كانت الأربعة الأخيرة.

2- التحليل متعدد المتغيرات

قدَّمت مجموعة البيانات الكاملة (2422 حالة) إلى مجموعتين: مجموعة التدريب (50%) التي تم منها تدقيق النماذج، ومجموعة الاختبار (50%) التي تم منها التحقق من مصداقية تلك النماذج، وتم تدقيق الانحدار

للوجستي وضرورة القرار.

* الانحدار الوجستي

طبق الانحدار الوجستي على المتغيرات بإتخاذ متغير حالة القرض كمتغير ثابع، (0 سوء أو متطرف، 1 جيد
 أو غير متطرف) وباقي المتغيرات كمتغيرات مستقلة، واستخدمت طريقة الانحدار المدرج الأحادي المبنية على نسبة
 الإمكان بتقاطع (.0, 0, 7، 0 على الترتيب). ويساوي نتائج 7 تفاعلات المقارنة بين تلك النماذج الثلاثة، وعلى

إذاً أن القيم غير المعتمدة لاختيار Hosmer and Lemeshow

الانحدار الوجستي لوصف البيانات عند نقاط الفصل الثلاثة، إلا أن أفضل النماذج كان عند نقطة القطع 0.3

(صاحب أعلى معدل دقة تصنيف للقيود السبعة).

جدول 4: ملخص المقارنة بين نماذج الانحدار الوجستي عند نقاط الفصل الثلاثة

<table>
<thead>
<tr>
<th>Model</th>
<th>Cut value</th>
<th>Nagelkerke R²</th>
<th>Hosmer and Lemeshow</th>
<th>Classification rates %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.757</td>
<td>good 69.8 bad 65.9 Total 92.1</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.494</td>
<td>0.757</td>
<td>good 99.0 bad 51.3 Total 93.1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.494</td>
<td>0.757</td>
<td>good 99.6 bad 35.3 Total 91.6</td>
</tr>
</tbody>
</table>

وبينما جدول 5 أن أهم المتغيرات المؤثرة على الجدارة الإثنائية حسب النموذج الأول كانت 7، هي: الدفاعات
الكاملة، وملغ القرض، ودرجة التصنيف الإثنائي، وأجل القرض، وملاحظة خصائص العمل بالنسبة الإثنائية للبنك،
والخاضع من القرض، وعدد النقاط على الترتيب.

قدَّمت مجموعة البيانات الكاملة (2422 حالة) إلى مجموعتين: مجموعة التدريب (50%) التي تم منها تدقيق

النماذج، ومجموعة الاختبار (50%) التي تم منها التحقق من مصداقية تلك النماذج، وتم تدقيق الانحدار

للوجستي وضرورة القرار.

* الانحدار الوجستي

طبق الانحدار الوجستي على المتغيرات بإتخاذ متغير حالة القرض كمتغير ثابع (0 سوء أو متطرف، 1 جيد
 أو غير متطرف) وباقي المتغيرات كمتغيرات مستقلة، واستخدمت طريقة الانحدار المدرج الأحادي المبنية على نسبة
 الإمكان بتقاطع (.0، .0، 7، 0 على الترتيب، ويوافق نتائج 7 تفاعلات المقارنة بين تلك النماذج الثلاثة.

الملف المصدر مستمر للسكان وتوزيع الأزمة - المجلد رقم 48 العدد الأول - يونيو لسنة 2015
على الرغم أن القيم غير المعنوية لاختبار Hosmer and Lemeshow لنموذج النماذج الثلاثة، إلا أن أفضل النماذج كان عند نقطة الفصل الثلاثة 

(صاحب على معدل دقة تصنيف الفروض السيئة).

<table>
<thead>
<tr>
<th>Model</th>
<th>Cut value</th>
<th>Nagelkerke R²</th>
<th>Hosmer and Lemeshow Test</th>
<th>good</th>
<th>bad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.494</td>
<td>0.757</td>
<td>95.8</td>
<td>65.9</td>
<td>92.1</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.494</td>
<td>0.757</td>
<td>99.0</td>
<td>51.3</td>
<td>93.1</td>
</tr>
<tr>
<td>3</td>
<td>0.7</td>
<td>0.494</td>
<td>0.757</td>
<td>99.6</td>
<td>35.3</td>
<td>91.6</td>
</tr>
</tbody>
</table>

وبينما جدول 4 ي塞尔ل التفاعل الإحصائي بين نماذج الإحصاء التصنيف عند نقاط الفصل الثلاثة، والكال، ومبلغ الفروض، ودرجة التصنيف الإحصائي، وأجل الفروض، ومتابعة خصائص العمل للسياسة الإنتاجية للبنك، والغرض من الفروض، ووحد الفئات على الترتيب.

**Table: Variables in the Equation**

<table>
<thead>
<tr>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>term(1)</td>
<td>1.686</td>
<td>.115</td>
<td>215.489</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>credit_policy(1)</td>
<td>.689</td>
<td>.124</td>
<td>30.588</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(1)</td>
<td>2.699</td>
<td>.353</td>
<td>58.426</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(2)</td>
<td>2.479</td>
<td>.244</td>
<td>103.401</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(3)</td>
<td>2.033</td>
<td>.164</td>
<td>154.702</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(4)</td>
<td>1.809</td>
<td>.126</td>
<td>205.726</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(5)</td>
<td>1.352</td>
<td>.115</td>
<td>137.154</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>grade(6)</td>
<td>.781</td>
<td>.112</td>
<td>48.405</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>purpose</td>
<td>32.146</td>
<td>.022</td>
<td></td>
<td>13</td>
<td>.002</td>
</tr>
<tr>
<td>purpose(1)</td>
<td>-474</td>
<td>.514</td>
<td>848</td>
<td>1</td>
<td>.357</td>
</tr>
<tr>
<td>purpose(2)</td>
<td>-667</td>
<td>.519</td>
<td>1653</td>
<td>1</td>
<td>.199</td>
</tr>
<tr>
<td>purpose(3)</td>
<td>-809</td>
<td>.530</td>
<td>2332</td>
<td>1</td>
<td>.127</td>
</tr>
<tr>
<td>purpose(4)</td>
<td>-143</td>
<td>.533</td>
<td>72</td>
<td>1</td>
<td>.788</td>
</tr>
<tr>
<td>purpose(5)</td>
<td>-527</td>
<td>.526</td>
<td>1006</td>
<td>1</td>
<td>.316</td>
</tr>
<tr>
<td>purpose(6)</td>
<td>-880</td>
<td>.529</td>
<td>2766</td>
<td>1</td>
<td>.096</td>
</tr>
<tr>
<td>purpose(7)</td>
<td>-546</td>
<td>.538</td>
<td>1035</td>
<td>1</td>
<td>.309</td>
</tr>
<tr>
<td>purpose(8)</td>
<td>-600</td>
<td>.553</td>
<td>1177</td>
<td>1</td>
<td>.278</td>
</tr>
<tr>
<td>purpose(9)</td>
<td>-1011</td>
<td>.580</td>
<td>3255</td>
<td>1</td>
<td>.071</td>
</tr>
<tr>
<td>purpose(10)</td>
<td>-735</td>
<td>.557</td>
<td>1743</td>
<td>1</td>
<td>.187</td>
</tr>
<tr>
<td>purpose(11)</td>
<td>-140</td>
<td>.562</td>
<td>662</td>
<td>1</td>
<td>.803</td>
</tr>
<tr>
<td>purpose(12)</td>
<td>-333</td>
<td>.608</td>
<td>301</td>
<td>1</td>
<td>.583</td>
</tr>
<tr>
<td>purpose(13)</td>
<td>-1019</td>
<td>.583</td>
<td>3051</td>
<td>1</td>
<td>.081</td>
</tr>
<tr>
<td>loan_amnt</td>
<td>.001</td>
<td>.000</td>
<td>1089.983</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>total_pymnt</td>
<td>-0.001</td>
<td>.000</td>
<td>1249.633</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.868</td>
<td>.532</td>
<td>29.063</td>
<td>1</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Variable(s) entered on step 6: purpose.*

*صنف الحالة في "آزم" إذا كان احتمالها أدنى بـ 0.05 باستخدام المرجح بزيادة قيم النقطة المحددة.*

المجلة المصرية لطبقرائومة الأسرة - المجلد رقم 48 العدد الأول - يونو سنة 2015
شرارة القرارات

أجريت محاكاة الكمبيوترية لثلاث طرق متعددة لشرارة القرارات (chi squared, entropy reduction, Gini reduction) واختبرت دقة تصنيفها للقرور الجيدة والسلبية. وسمت معدلات دقة التصنيف الصحيح على المستوى الكلي، والقرور المستدقة، والقرور المختارة لثلاث نقاط قبل اعتدال (0.7، 0.6، 0.5). إذ قد يتم اختيار أفضل نموذج على قسط العينة الإحصائية التي يستخدمها البنك. ولأن الحد المستهدف هو اكتشاف التخلف عن سداد التحصوي؛ فإن التفعيل عند الاحتمال 0.3 يعني أن كل مادة الخطأ (الناجحة عن منح التحصوي وهو لا ينبغي أن يمنح) تكون أعلى مقدار (0.3 متابعًا). من كلفة رفضه عندما ينبغي أن يمنح، وبمعنى آخر، فإن التفعيل عند 0.3 سوف يمنع للبنك بحثみな المعامل الذين يفضل تعثرهم في سداد التحصوي ومنح الاختلاف للعملاء الأكثر جدارة اقتصاديا. كما يعني التفعيل 0.5 أن كلفة الخطأ (الناجحة عن منح التحصوي وهو لا ينبغي أن يمنح) تساوي كلفة رفضه عندما ينبغي أن يمنح، وأخيرًا، فإن التفعيل 0.7 يعني أن كلفة الخطأ (الناجحة عن منح التحصوي وهو لا ينبغي أن يمنح) تكون أقل من كلفة رفضه عندما ينبغي أن يمنح، وبمعنى أن في حالات التحصوي الصغيرة غير المؤمنة (التي لا تستطع أي ضمانات)، فإن التفعيل عند 0.5، و0.7، و0.1 ينسب التحصوي الأكبر (التي يأخذ عليها ضمانات) مثل التحصوي بسرعة سريرًا مشبوهًا.

وقد تبين عدم وجود اختلاف معنوي بين معدلات دقة التصنيف الصحيح للخوارزميات الثلاث عند نقاط الذروة 0.03، 0.1، 0.5، سواء للقرور الجيدة (المستدقة) أو السينة (المنسية) أو على المستوى الكلي. وكانت خوارزمية CRT هي الأفضل في التصنيف الصحيح على المستوى الكلي لأن معدل دقة التصنيف لها بلغ 92.6%، بينما كانت خوارزمية CHAID هي الأفضل في التصنيف الصحيح للقرور المستدقة لأن معدل دقة التصنيف لها بلغ 50.1%.

كما أُدرست الأهمية السيمية للمتغيرات المستخدمة في بناء النماذج الثلاث للحالة القرارات، وتبين أن النماذج الثلاث قد اقتطع على أن متغير "نسبة الدين إلى الدخل" كان أهم المتغيرات في التنبؤ بنتائج المتغير الهيدف (حالة التحصوي)، حيث بلغت الأهمية السيمية للمتغير في النماذج الثلاث 0.1. كما اقتطع النماذج الثلاث أيضًا على أهمية متغيرات أخرى في التنبؤ بحالة التحصوي، وكانت هذه المتغيرات هي: عدد خطوات الاتصال الأخيرة (الأقسام المتضمنة)، واليوم بالشهر من أقدم خط التأخر (القرور المتضررة) وقيمة الممتلكات الحالية. وقد بلغت الأهمية السيمية المتوسطة للنماذج الثلاث: 0.30، 0.27، 0.24 على التوالي.

وقد اقتصر عرض الشرارة فيها على 3 سنوات، وبالمثل من أن معدلات دقة التصنيف للطرق الثلاث كانت chi squared and entropy غير معنوية إحصائيًا عند نقاط التحليل الثلاثة، إلا أنه يبدو أن طريقة Gini reduction كانت أفضل من طريقة ROC هذهيا، كما أنه تألقت بنتائج أفضل، واتبعت ROC للتحليل الشامل ل معدلات التصنيف، يُصبح بشكل طريقة الصفر، حيث كانت نتائج التحليل جيدة أكثر، وكان على الصفر، كما كان أداء النموذج أفضل، واتبعت ROC للتحليل الشامل ل معدلات التصنيف، يُصبح بشكل طريقة الصفر، حيث كانت نتائج التحليل جيدة أكثر، وكان على الصفر، كما كان أداء النموذج أفضل، واتبعت ROC للتحليل الشامل ل معدلات التصنيف، يُصبح بشكل طريقة الصفر، حيث كانت نتائج التحليل جيدة أكثر، وكان على
Table 6: Classification - tree1

<table>
<thead>
<tr>
<th>Sample</th>
<th>Observed</th>
<th>Predicted</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td>2174</td>
<td>1769</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>610</td>
<td>25264</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>882</td>
<td>784</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td>237</td>
<td>10723</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>8.7%</td>
<td>91.3%</td>
</tr>
</tbody>
</table>

Growing Method: CHAID
Dependent Variable: credit

Table 7: Classification - tree2

<table>
<thead>
<tr>
<th>Sample</th>
<th>Observed</th>
<th>Predicted</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td>2085</td>
<td>1801</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>396</td>
<td>25422</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>929</td>
<td>774</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td>190</td>
<td>10826</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>8.8%</td>
<td>91.2%</td>
</tr>
</tbody>
</table>

Growing Method: CRT
Dependent Variable: credit

Table 8: Classification - tree3

<table>
<thead>
<tr>
<th>Sample</th>
<th>Observed</th>
<th>Predicted</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td>1540</td>
<td>2303</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>336</td>
<td>25593</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>676</td>
<td>1070</td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td>159</td>
<td>10748</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>6.6%</td>
<td>93.4%</td>
</tr>
</tbody>
</table>

Growing Method: QUEST
Dependent Variable: credit
شكل 2: المقارنة بين دقة تصنيف الإحصاء اللوجستي وأشجار القرار

(7) الخلاصة والمقترحات للبحوث المستقبلية

درس البحث البيانات التاريخية من القروض الاستهلاكية التي تصدرها إحدى المؤسسات المالية للأفراد الذين تعتبرهم عملاء مؤهلين للحصول على تلك القروض. وما لذا التصنيف الدقيق يفيد كل من الدائن (زيادة الربح أو تقليل الخسائر) والمدين (تجنب الخسائر). فقد استخدم البحث الإحصاء اللوجستي وطرق استنتاج القاعدة (نمذج القروض) للتنبؤ بما إذا كان طالب القرض الجديد سيذهب قرضه أو سيتعثر فيه؟ وتوقع في تلك الإحصاء اللوجستي على أشجار القروض، ويمكن تفسير هذه القواعد لكل من مدبري الائتمان بالبنك (الذين يحتاجون فهمًا قبل الموافقة على تفتيشها) وطالبي القروض كسبب لحكمهم من القضي.

وينبغي إجراء مزيد من البحوث بشأن تحسين كرب الوقت اختبار نماذج سطح القرارات، إذ أن ضبط النماذج فضلاً عن استخدام مجموعات البيانات المتوازنة وغير المتوازنة يساهم في تحسين أداء التصنيف. كما ينبغي مقارنة نتائج سطح القرارات بنتائج الطرق الأخرى للتغلب في البيانات التي تقابل الائتمان (كالتكتيكات العصبية والأنظمة اللوجستية، وأنظمة الت⏰). للوقوف على أفضلها في تصنيف الصحيح لحاولي القرض، وأخيرًا، ينبغي على البنك تحديد سياسة الإفراط الأكثر ربحية ومن خطر الخسائر استنادًا إلى الأرباح المتوقعة على القروض الجيدة، ونحو الخسائر على القروض السيئة، والتكاليف الثابتة والمتغيرة لعمليات الإفراط أو تسمح للباحثين بحسابها بدلاً من اعتبار تلك البيانات موضوعات في غاية السرية.


   https://www.lendingclub.com/info/download-data.action


