استخدام شجرة القرار للتنبؤ بسداد القروض للحد من المخاطر المالية في البنوك

م.د عمر عادل عبدالوهاب
جامعة ديالى – كلية الادارة والاقتصاد
omaradil.d87@gmail.com

م.م مريم محمد سلمان
جامعة ابن سينا للعلوم الطبية والصيدلانية

Using tree decision to predict loan repayments to reduce bank financial risk

Lec. Dr. Omar Adil Abdalwahab
College of Management and Economics University of diyala

Asst.Lec. Maryam Mohammed Salman
Ibn Sina University of Medical and Pharmaceutical Sciences, Iraq, Bagdad

تاريخ استلام البحث 2017/4/16 تاريخ قبول النشر 2018/1/15

المستلخص:

في السنوات الاخيرة، تزايد اهتمام البنوك بالعلاقة مع العملاء اذ تعتبر عامل مهم للغاية لنجاحها و التحدي الذي يواجه البنك هو طريقة الاحتفاظ بالعملاء ذو الاهمية الربحية الاكثر وكيفية القيام بذلك بأقل تكلفة. في الوقت نفسه، يحتاجون إلى إيجاد هذا الحل وتنفيذه بسرعة واكثر مرونة. أن استخدام الطرائق الكلاسيكية في تحليل البيانات بغية الكشف عن الاحتيال في المجالات المختلفة ولا سيما المالية. أذ إنها تستدعي متطلبات اكثر تعقيدا وتأخذ وقتا اطول في التعامل مع مجالات مختلفة من المعرفة مثل والقانون والاقتصاد والممارسات التجارية والمالية. يمكن أن تكون حالات الاحتيال متشابهة في المحتوى والمظهر ولكنها عادة ما تكون غير متطابقة لذلك سيتم استخدام شجرة القرار بالاضافة الى طرائق اخرى والتي تعتبر طرائق تنقيب عن بيانات والتي تم استخدامها في القطاعات المصرفية و التي توفر المنتج المناسب للعميل المناسب مع مخاطر أقل. مخاطر الائتمان مثل التخلف عن سداد القروض و هي تعتبر المصدر الرئيسي للمخاطر التي يواجهها القطاع المصرفي، اذ يمكن تطبيق طرائق التنقيب عن البيانات مثل التصنيف والتنبؤ للتغلب على هذه المخاطر إلى حد كبير او تقليل اثارها السلبية. اذ يقدم هذا البحث نموذجاً للتنبؤ يخدم العاملين في القطاع المصرفي والبنوك في التنبؤ بالعملاء الذين يقدمون طلبات للحصول على قروض من البنوك والمصارف. تم في هذا البحث استخدام نموذجين في التنبؤ لسداد القروض في البنوك العراقية، الاولى نموذج الانحدار اللوجستي وهي تعتبر طريقة كلاسيكية في التنبؤ والطريقة الثانية هي طريقة شجرة القرار بالاعتماد على برنامج PYTHON، اذ تم حساب دقة التنبؤ للنموذجين و توصل البحث الى ان استخدام طريقة شجرة القرار هي الافضل في التنبؤ بسداد القروض.

الكلمات المفتاحية: *شجرة القرار،الانحدار اللوجستي، تنقيب البيانات، القروض.*

Abstract:

In recent years, the interest of banks in the relationship with customers has increased, as it is a very important factor for their success, and the challenge facing the bank is the way to retain customers with the most profitability and how to do it at the lowest cost. At the same time, they need to find and implement this solution quickly and more flexibly. The usage of classical methods in data mining in order to detect fraud in various fields, especially, financial. It calls for more complex requirements and takes more time in dealing with different areas of knowledge such as law, economics, business and financial practices. Fraud cases can be similar in content and appearance, but they are usually not identical. The decision tree is used in addition to other methods, which are data mining methods that will be used in the banking sectors and that provide the right product for the right customer with fewer risks. Credit risks such as loan defaults are the main source of risks faced by the banking sector, as data mining methods such as classification

and forecasting can be applied to overcome these risks to a large extent or reduce their negative effects. This research presents a forecasting model that serves workers in the banking sector and banks to predict customers who apply for loans from banks and banks. In this research, two models were used to predict the repayment of loans in Iraqi banks, the first is the logistic regression model, which is considered a classic method for forecasting, and the second method is the decision tree method based on the PYTHON program, as the prediction accuracy of the two models was calculated. The tree decision is the best in predicting loan repayments.

***Keywords**: tree decision, logistic regression, data mining, loans..*

المقدمة:

في السنوات الاخيرة، ظهر مصطلح علم البيانات والذي يعتبر من التطبيقات الحديثة في كافة المجالات الطبية والهندسية والانسانية بالإضافة الى علاقته الوثيقة بعلم الاحصاء[4]. في كثير من الاحيان يكون التمييز بين علم البيانات و علم الاحصاء صعبا، أذ يستخدم علم البيانات طرائق و وسائل تنتمي إلى علم الاحصاء وأدبياته والتي تكون ذات قيمة كبيرة عندما يتم العمل على تعلم الآلة. حيث ان التعلم الالي يُمثل الية تحسين و تعزيز برامج الكمبيوتر بشكل ذاتي وفقاً للخبرة، أذ يمكن تعليم الكمبيوتر من خلال إدخال البيانات وتدريبها وبناء نموذج خاص بهذه البيانات . يعتبر موضوع تنقيب البيانات احد ابرز المواضيع المهمة والمرتبطة بعلم البيانات والاحصاء بالإضافة الى تعلم الالة اذ يعتبر هو الجانب العملي لتنقيب البيانات والتنبؤ بها، اذ يمكن تعريف تنقيب البيانات بانه استخراج المعلومات من البيانات. أو هو طريقة لاستخراج كل المعرفة من البيانات قيد الدراسة. أذ ساعدت تكنولوجيا المعلومات بشكل عام في ما يخص الصناعة المصرفية على التعامل مع التحديات التي يفرضها النظام الاقتصاد الجديد .

وفي الوقت الحاضر، تزايد اهتمام البنوك بالعلاقة مع العملاء اذ تعتبر عامل مهم للغاية لنجاحها و التحدي الذي يواجه البنك هو طريقة الاحتفاظ بالعملاء ذو الاهمية الربحية الاكثر وكيفية القيام بذلك بأقل تكلفة. في الوقت نفسه [1]، يحتاجون إلى إيجاد هذا الحل وتنفيذه بسرعة واكثر مرونة[2]. أن استخدام الطرائق الكلاسيكية في تحليل البيانات بغية الكشف عن الاحتيال في المجالات المختلفة ولا سيما المالية. أذ إنها تستدعي متطلبات اكثر تعقيدا وتأخذ وقتا اطول في التعامل مع مجالات مختلفة من المعرفة مثل والقانون والاقتصاد والممارسات التجارية والمالية. يمكن أن تكون حالات الاحتيال متشابهة في المحتوى والمظهر ولكنها عادة ما تكون غير متطابقة لذلك يتم اللجوء استخدام طرائق اخرى والتي تعتبر طرائق تنقيب عن بيانات والتي تم استخدامها في القطاعات المصرفية و التي توفر المنتج المناسب للعميل المناسب مع مخاطر أقل للائتمان .

اذ تكون مخاطر الائتمان مثل التخلف عن سداد القروض و هي تعتبر المصدر الرئيسي للمخاطر التي يواجهها القطاع المصرفي، اذ يمكن تطبيق طرائق التنقيب عن البيانات مثل التصنيف والتنبؤ للتغلب على هذه المخاطر إلى حد كبير او تقليل اثارها السلبية أدركت البنوك إن العلاقات مع العملاء هي عامل مهم للغاية لنجاحها والتحدي الذي يواجه البنك هو كيفية الاحتفاظ بالعملاء ذو الربحية الأكثر والقيام به بأقل تكلفة. في الوقت نفسه، يحتاجون إلى إيجاد هذا الحل وتنفيذه بسرعة وأن يكون الحل مرنًا.

هدف البحث:

الهدف من هذه الدراسة هو استخدام وتطبيق شجرة القرار في التنقيب عن البيانات، وبناء النموذج المناسب لها وكذلك لاكتشاف المخاطر التي قد تهدد البنك، من خلال تحديد الخصائص التي تشير إلى وجود مخاطر أن يكون العملاء غير قادرين على سداد التزاماتهم المالية والذي يؤدي إلى اضطراب أداء البنك بصورة عامة وذلك بسبب التقصير في الوفاء بالتزامات المدينين.

اهمية البحث:

يتطلع هذا البحث إلى التنبؤ بمخاطر الائتمان باستخدام شجرة القرار ومقارنتها مع الانحدار اللوجستي، حيث يمكن تحديد اهمية البحث بما يلي:

١ . التعريف بمدى أهمية التنبؤ بمخاطر الائتمان للحد من المخاطر المالية وتحديد درجة أو مقدار المخاطرة التي يمكن ادارتها و تقبلها.

٢ . تحديد مدى اختلاف وأفضلية شجرة القرار عن نموذج الانحدار اللوجستي الذي يستخدم في التنبؤ بمخاطر الائتمان في البنوك.

المبحث الأول

الجانب النظري

1- المخاطر المالية والتنبؤ بها

تعتبر المخاطر المالية من المواضيع الحساسة التي تواجه البنوك في السنوات الأخيرة، ان هنالك العديد من المخاطر المالية التي تتعرض لها هذه البنوك، حيث يجب على كل بنك تحديد وتحليل المخاطر المالية التي قد تتعرض لها بهدف اتخاذ الاجراءات اللازمة لتجاوز آثارها السلبية، ويمكن تحديد هذه المخاطر المالية بـ(مخاطر ائتمان ، مخاطر السيولة، مخاطر السوق، مخاطر رأس المال)، اذ يعرف التنبؤ بصورة عامة على انه تقدير أو توقع لما سيحدث مستقبلا[8]، لذا فهو يعتبر حلقة الوصل بين البنوك والبيئة المحيطة بها من اجل الربط بني الحاضر والمستقبل، فبيئة البنوك الخارجية تتغير باستمرار نتيجة لتغيرات الاقتصادية و البيئية والتكنولوجية المتلاحقة[1]، والتي تدفع الى زيادة حالة عدم اليقين والتأكد، وبالتي زيادة المخاطر التي قد يتعرض لها البنك.

هنالك اهداف عديدة من التنبؤ بمخاطر الائتمان والتي يندرج ضمنها مخاطر سداد القروض والتي منها تخفيض درجة عدم التأكد التي ستواجه البنك في المستقبل[7]، والتعرف على الأداء المستقبلي للبنك بالاعتماد على البيانات المسجلة سابقا[5]، وتوفير قاعدة معلومات ضرورية لوضع الخطط ورسم السياسات ورقابة تنفيذها[2]. وعلى الرغم من اهداف التنبؤ الى ان هنالك محددات التي تقيد من فاعلية التنبؤ بمخاطر الائتمان، وقد تؤثر على دقة وصحة التنبؤات والتي هي صعوبة توقع بعض الأحداث المستقبلية، وعدم وضوح الافتراضات أو تعقدها، بالإضافة الى التنبؤات غير الرشيدة التي يتم اعدادها على أسس غير موضوعية أو تتجاهل بعض المتغيرات[6] .

2- طرائق التنبؤ بسداد القروض

تعتبر الطرائق الاحصائي احد ابرز الادوات في التنبؤ للمستقبل على ضوء البيانات السابقة للظاهرة قيد الدراسة، وفي حالة كون المتغير المعتمد من النوع ثنائي الاستجابة أي (0 ، 1) كان لابد من استخدام نماذج تلائم هذه البيانات.

2.1 طريقة الانحدار اللوجستي

مفهوم الانحدار اللوجستي:

يعتبر الانحدار اللوجستي الثنائي أحد أهم تقنيات تحليل الإنحدار، بإعتباره أكثر النماذج ملائمة لوصف البيانات التي تكون الاستجابة فيها ثنائية ويكون فيها المتغيرات التوضيحي ذات مستويات متعددة. حيث أن الانحدار اللوجستي يستخدم عدة متغيرات متوقعة والتي يمكن أن تكون رقمية أو فئوية. يمكن توسيع هذا لنمذجة عدة فئات[3]. يعرف الانحدار اللوجستي أيضا بتسميات أخرى كنموذج لوجيت (Logit) أو المصنف العام للأنتروبية. يندرج الانحدار اللوجستي ضمن خوارزميات التعلم الآلي الخاضع للإشراف المخصصة لمهام "التصنيف" . اكتسب على مدى العقدين الماضيين سمعة طيبة خاصة في القطاع المالي بسبب قدرته البارزة في الكشف عن المختلسين. أدناه مخطط للإستخدام العام للانحدار اللوجستي وغيره من المصنفات الخطية الشائعة[11].

يعتمد نموذج الانحدار اللوجستي على فرض أساسي هو أن المتغير التابع (Y) ثنائي الاستجابة يأخذ أحدى القيمتين(0,1) أما النجاح (Success) باحتمال أو الفشل (Failure) باحتمال ($1 - \pi i$) لذلك يكون المتغير (yi) يتوزع حسب توزيع برنولي (πi) Ber .

$yi \sim$Ber (π) $i = 1 , 2 , \dots , n$

أي أن

ومن ثم فأن دالة الكثافة الاحتمالية تكون وفق الصيغة الاتية :

$$P(yi) = \pi iyi\ (1 - \pi i)\ 1 - y \qquad (1)$$

أذ أن :

yi متغير تابع ثنائي الاستجابة (0,1)

πi احتمال حدوث الاستجابة عندما $yi=1$

πi – 1 احتمال عدم حدوث الاستجابة عندما $yi=0$

لذلك فأن توقع المتغير يمثل احتمال حدوث الاستجابة (πi) وكالاتي:

$$E(yi) = pr(y = 1) = \pi i \qquad (2)$$

أما تباين المتغير yi حسب توزيع برنولي كالاتي:

$$V(y_i) = \pi_i(1 - \pi_i) \qquad (3)$$

ليكن X1,X2,...,Xp مجموعة من المتغيرات التوضيحية ولتكن n تمثل عدد المشاهدات لهذه المتغيرات التي تكون المصفوفة X . $X = (Xij)_{nxp}$

اذ ان :

I=1,2,…,n، n تمثل حجم العينة.

J=1,2, …,p (p) تمثل عدد المتغيرات التوضيحية.

فاذا كان$[y1, y2, \dots, yn] = yi$ عينة عشوائية من المتغير ثنائي الاستجابة وأن $(0,1) \in$ وبالتالي فأن أنموذج الانحدار اللوجستي يكتب بالصيغة الاتية:

$$yi = \pi i + £i \qquad (4)$$

اذ أن πi تمثل دالة الانحدار اللوجستي (احتمال الاستجابة)

$$\pi i = \frac{\exp\{\beta o + \beta 1 xi1 + \cdots + \beta p Xip\}}{1 + \exp\{\beta o + \beta 1 xi1 + \cdots + \beta p Xip\}} \qquad (5)$$

او يمكن كتابته بالصيغة الاتية :

$$\pi i = \frac{\exp\{\beta o + \sum_{j=1}^{p} \beta j Xij\}}{1 + \exp\{\beta o + \sum_{j=1}^{p} \beta j Xij\}} \qquad (6)$$

$$(1-\pi i) = \frac{1}{1+\exp\{\beta o+\sum_{j=1}^{p}\beta jXij\}} \qquad (7)$$

حيث ان $\beta o, \beta 1, \dots, \beta p$ هي معلمات النموذج وان £i يمثل الخطأ العشوائي يتوزع توزيع برنولي بمتوسط صفر وتباين $\pi i(1-\pi i)$

2.2 طريقة شجرة القرار

مفهوم شجرة القرارات:

ان العديد من قرارات ً إدارة العمليات غالبا ما تتطلب أدوات لاتخاذ قراراتها والتي تراعي النتائج أو الآثار المالية للبدائل، وان الأداة عادة ما تجد طريقها لعمليات اتخاذ القرار(Processes Making-Decision) في إدارة العمليات وأحدى الأدوات لتحليل هذه القرارات تعرف باسم شجرة القرارات والتي تقارن الأداء المالي للبدائل المتوقعة. يعد أسلوب شجرة القرار المدخل العام لمجموعة واسعة من قرارات العمليات[10]، وأصبحت شجرة القرارات مشهورة في تحليل القرار في عام(1960)، وهي تعد الأداة التي لديها العديد من التطبيقات العملية، وهي تستخدم في تقييم البدائل المختلفة لتوسيع الطاقة الإنتاجية حين يكون الطلب غير مؤكد والقرارات متعاقبة، وغالبا ما يستخدمها متخذ القرار للحصول على وصف أو صورة واضحة لبدائل القرار ونتائجها المحتملة.

يتم استعمال تعلم شجرة القرار Decision Tree كنموذج تنبؤيPredictive model ، والذي يقوم بتخطيط الملاحظات حول عنصر ما (ممثلة بالفروع) لاستنتاج قيمة هدف هذا العنصر (ممثلة بالأوراق)، ان احد اساليب النموذج التنبؤ المستخدمة في الاحصاء واستخراج البيانات والتعلم الالي [3]. في نماذج الشجرة يمكن للمتغير المستهدف ان يأخذ مجموعة منفصلة من القيم تسمى اشجار التصنيف، ففي هذه الهياكل الشجرية ،الورقية العقدة Leaf Node تمثل تسميات الفصل وتمثل الفروع المنطق الاقتران من المميزات التي تؤدي الي تسميات هذه الفئة. تسمى اشجار القرار حيث يمكن للمتغير المستهدف ان يأخذ قيما مستمرة عادة ارقام حقيقية real numbers اشجار الانحدار . في تحليل القرار ، يمكن استخدام شجرة القرارات لتمثيل القرارات بشكل مرئي وصريح وصنع القرار . Decision making في استخراج البيانات ، تصف شجرة القرارات البيانات ، لكن يمكن ان تكون تصنيفات النتائج كدخل لاتخاذ القرارات[11].

المبحث الثاني

الجانب التطبيقي

1. التطبيق

تم اجراء عملية التنقيب على بيانات القروض التابعة للمصرف التجاري العراقي لغرض اعطاء الادارة صورة مستقبلية و مساعدتها على اتخاذ القرار اذ يمكنها من خلاله الحكم على الزبون فيما اذا كان التعامل معه يشكل خطرا على المصرف. وبذلك تستطيع التركيز على الزبائن الذين يكون موثوق منهم واستبعاد غير الموثوق بهم لأنه يكون السبب فيما يتحمله المصرف من خسائر ناجمة عن اقراضه بسب عدم التسديد. حيث تم اخذ متغيرات البحث الخاصة بالزبائن بالاعتماد على قرارات المصرف السابقة في اعطاء القروض.

2. نتائج التحليل

1- المؤشرات الاحصائي:

يعتبر التحليل الاحصائي من اهم الادوات التي يستخدمها الباحثين في تحليل البيانات من اجل الوصول الى قرارات صحيحة عن طريق نتائج تحليل البيانات، لذلك يتم اجراء الاساليب الاحصائية من اجل التعرف على طبيعة البيانات والتعرف عليها. سيتم في هذا المبحث التحليل الاحصائي لبيانات الدراسة المتمثلة بمتغيرات (الجنس، الحالة الاجتماعية، التعليم، الحالة الوظيفية، امتلاك بطاقة سابقا و حالة القرض)، حيث سيتم حساب التكرارات والنسب المئوية كما في جدول رقم (1)

جدول رقم (1)

يبين التكرارات والنسب المئوية لمتغيرات الدراسة

Gender		
	Frequency	Percent
Female	123	20.0
Male	491	80.0
Total	614	100.0
Married		
Not married	215	35.0
Married	399	65.0
Total	614	100.0
Education		
Not Graduate	134	21.8

Graduate	480	78.2
Total	614	100.0
Self Employed		
No	512	83.4
Yes	102	16.6
Total	614	100.0
Credit History		
Not found	106	17.3
Found	508	82.7
Total	614	100.0
Loan Status		
No	192	31.3
Yes	422	68.7
Total	614	100.0

المصدر: الجدول من اعداد الباحث

من خلال جدول رقم (1) نلاحظ أن اغلب عينة الدراسة كانت من الذكور اذ بلغ عددهم (491) بنسبة (80%) في حين كانت نسبة الاناث هي (20%). كما يلاحظ ان الاكثرية في عينة الدراسة من المتزوجين اذ بلغ عددهم (399) بنسبة (65%) في حين بلغت نسبة الغير متزوجين (35%). كما يتبين ايضا ان عدد المتعلمين هو الاكثر اي بلغ (480) بنسبة (78.2%) حيث كانت نسبة الغير متعلمين هي (21.8%). ومن خلال الجدول اعلاه نلاحظ ان لا يمتلكون عمل خاص بهم كانوا هم الاغلبية في عينة الدراسة اذ بلغ عددهم (512) بنسبة (83.4%) وكانت نسبة الذين يمتلكون عمل خاص بهم (16.6%). كما تبين من خلال حساب التكرارات للذين يمتلكون بطاقة بنكية ان عددهم هو (508) بنسبة (82.7%) وكانت نسبة الذين لا يمتلكون بطاقة بنكية (17.3%). ومن خلال جدول (1) يتبين ان عدد الذين تحصلوا على قرض في عينة الدراسة هو (422) بنسبة (68.7%) في حين كانت نسبة الذين لم يتحصلوا على قرض هي (31.3%).

ومن خلال النتائج اعلاه يتضح ان اغلب عينة الدراسة قد حصلوا على قرض من الذين كانوا من الذكور المتزوجين، والذين هم من المتعلمين الغير حاصلين على وظائف خاصة بهم، من الذين يمتلكون بطاقات بنكية.

جدول رقم (2)
يبين المؤشرات الاحصائية الخاصة بمتغيرات الدراسة

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Applicant Income	614	150	81000	5403.459	6109.04167
Co applicant Income	614	0	41667	1621.245	2926.24837
Loan Amount	614	9	700	146.5928	85.62432
Loan Amount Term	614	12	480	341.1401	66.27544

المصدر: الجدول من اعداد الباحث

من خلال جدول رقم (2) والذي يبين حساب المؤشرات الاحصائية المتمثلة بعدد المشاهدات، اقل قيمة ، اعلى قيمة، الوسط الحسابي و الانحراف المعياري. حيث نلاحظ ان متغير (Applicant Income) قد كانت اقل قيمة له هي (150) واعلى قيمة له (81000) في حين بلغت قيمة الوسط الحسابي له (5403.459) بانحراف معياري قدره (6109.04167). ويلاحظ ايضا ان قيمة متغير (Co applicant Income) قد كانت اقل قيمة له هي (0) واعلى قيمة له (41667) في حين بلغت قيمة الوسط الحسابي له (1621.245) بانحراف معياري قدره (2926.24837). ومن خلال نتائج الجدول يتضح ان (Loan Amount) قد كانت اقل قيمة له هي (9) واعلى قيمة له (700) في حين بلغت قيمة الوسط الحسابي له (146.5928) بانحراف معياري قدره (85.62432). كما يتضح من خلال نتائج جدول رقم (2) ان متغير (Loan Amount Term) قد كانت اقل قيمة له هي (12) واعلى قيمة له (480) في حين بلغت قيمة الوسط الحسابي له (341.1401) بانحراف معياري قدره (66.27544).

2- اختبار مشكلة التعدد الخطي:

ان من اهم الاجراءات الواجب عملها قبل الشروع في تحليل البيانات لا سيما في ما يخص تصنيف البيانات هو اختبار وجود مشكلة التعدد الخطي بين المتغيرات المستقلة و المتغير التابع، اذ تعتبر هذه المشكلة من ابرز المشكلات التي تواجه تحليل البيانات عند وجود عدد من المتغيرات المستقلة حيث ان وجود هذه المشكلة يؤدي الى نتائج مضللة وغير دقيقة وبالتالي يؤثر على علمية اتخاذ القرار، لذلك تم اختبار المتغيرات المستقلة والكشف عن وجود مشكلة التعدد الخطي عن طريق اختبار معامل (VIF) حيث تعتبر البيانات تعاني من مشكلة التعدد الخطي اذا كانت قيمة اختبار معامل (VIF) اكبر من 10%، واشارت بعض

الدراسات الى ان المتغيرات تعاني من مشكلة التعدد الخطي اذا كانت قيمة اختبار معامل (VIF) اكبر من 5%.

جدول رقم (3)

Model		Collinearity Statistics	
		Tolerance	VIF%
	Gender	0.878	1.139
	Married	0.866	1.154
	Education	0.948	1.054
	Self Employed	0.977	1.023
	Applicant Income	0.631	1.584
	Co-applicant Income	0.878	1.139
	Loan Amount	0.610	1.638
	Loan Amount Term	0.960	1.042
	Credit History	0.993	1.007

المصدر: الجدول من اعداد الباحث بالاعتماد برنامج SPSS24

من خلال جدول رقم (3) نلاحظ ان جميع المتغيرات المستقلة قد حصلت على معامل (VIF) اقل من 5% ويدل هذا على عدم وجود مشكلة التعدد الخطي بين المتغيرات المستقلة في البحث قيد الدراسة.

3- اساليب تصنيف البيانات:

يعتبر تصنيف البيانات من اهم أدوات اتخاذ القرار لا سيما تلك التي تعتمد على الاستجابة الثنائية (0,1)، اذ ان Statistical classification عبارة عن عملية احصائية يتم فيها توزيع بيانات معينة لمجتمع احصائي على مجموعات مختلفة بناء على معلومات كمية تستند إلى واحدة او اكثر من الخواص الأساسية لهذه البيانات أو أعضاء المجتمع الإحصائي . تستند عملية التصنيف هذه على خاصيات أصيلة في العناصر (التي قد تكون : رموزا أو متغيرات ....) و تستند على مجموعة تدريب من هذه العناصر.

وهنالك عدة اساليب في تصنيف البيانات لا سيما التي تدخل في تنقيب البيانات وهي اسلوب شجرة القرارات، اسلوب الغابة العشوائية والانحدار اللوجستي.

4- نتائج شجرة القرارات: تم استخدام اسلوب شجرة القرارات في تحليل متغيرات الدراسة من اجل الحصول على افضل النتائج، اذ يعتبر اسلوب شجرة القرارات احد اساليب تنقيب البيانات

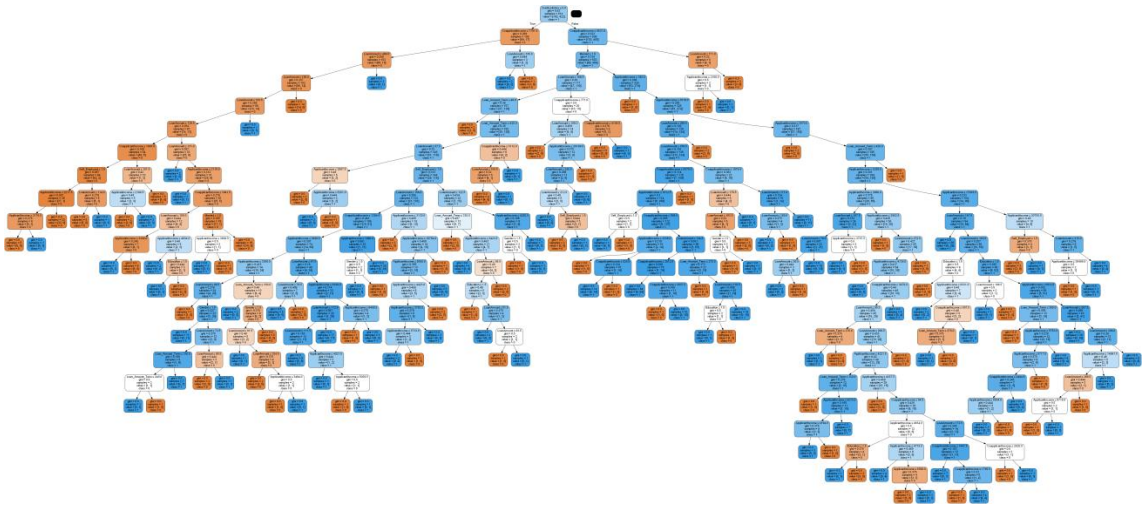
جدول رقم (4)

يبين نتائج مخرجات شجرة القرارات

Variables	Importance	Accuracy
Credit History	0.269558	0.6858
Loan Amount	0.252645	
Applicant Income	0.207474	
Co-applicant Income	0.130674	
Loan Amount Term	0.07103	
Gender	0.029896	
Married	0.018193	
Self Employed	0.012106	
Education	0.008423	

من خلال جدول رقم (4) يبين نتائج مخرجات شجرة القرارات والتي تظهر اهمية المتغيرات المستقلة والتي تؤثر في عملية اتخاذ القرار لمنح القرض من قبل المصرف، اذ نلاحظ أن متغير (***History Credit***) يمتلك اعلى اهمية من بين المتغيرات المستقلة اذ بلغ (0.269558)، في حين جاء المتغير (***Loan Amount***) في المرتبة الثانية اذ بلغت اهميته (0.252645)، وجاء المتغير (***Applicant Income***) في المرتبة الثالثة اذ حصل على اهمية بلغت (0.207474) وجاء المتغير (***Co-applicant Income***) في المرتبة الرابعة بأهمية بلغت (0.130674). أما بقية المتغيرات (***Loan Amount Term*** ،***Gender*** ،***Married*** ،***Self Employed*** ،***Education***) فلم تكن ذو اهمية في عملية التنبؤ بمنح القرض من قبل المصرف.

في حين بلغت قيمة Accuracy من خلال استخدام شجرة القرارات للتنبؤ فيمنح القرض هي (68.58%)



شكل رقم (1)

يبين شجرة القرارات الخاصة بالتنبؤ بمنح القرض

من خلال شكل (1) يبين شكل عملية اتخاذ القرار من خلال التفرعات الخاصة بمتغيرات الدراسة.

5- نتائج الانحدار اللوجستي: تم استخدام اسلوب الانحدار اللوجستي في تحليل متغيرات الدراسة من اجل الحصول على افضل النتائج، اذ يعتبر هذا الاسلوب هو الاساس في تصنيف البيانات.

جدول رقم (5)

يبين نتائج مخرجات الانحدار اللوجستي

Variables	Importance	Accuracy
Credit History	2.69405	0.52577
Applicant Income	0.34182	
Loan Amount	0.21197	
Co applicant Income	0.20053	
Gender	0.18138	
Loan Amount Term	0.00388	
Self Employed	0.00198	
Education	0.00005	
Married	0.00002	

من خلال جدول رقم (5) يبين نتائج مخرجات الانحدار اللوجستي والتي تظهر اهمية المتغيرات المستقلة والتي تؤثر في عملية اتخاذ القرار لمنح القرض من قبل المصرف، اذ نلاحظ أن متغير (***Credit History***) يمتلك اعلى اهمية من بين المتغيرات المستقلة اذ بلغ (2.69405)، في حين جاء المتغير (***Applicant***

(***Income***) في المرتبة الثانية اذ بلغت اهميته (0.34182)، وجاء المتغير (***Loan Amount***) في المرتبة الثالثة اذ حصل على اهمية بلغت (0.21197). وجاء المتغير (***Co-applicant Income***) في المرتبة الرابعة باهمية بلغت (0.20053). وجاء المتغير (***Gender***) في المرتبة الخامسة باهمية بلغت (0.18138) أما بقية المتغيرات (***Loan Amount Term ، Married، Self Employed، Education***) فلم تكن ذو اهمية في عملية التنبؤ بمنح القرض من قبل المصرف. في حين بلغت قيمة Accuracy من خلال استخدام شجرة القرارات للتنبؤ فيمنح القرض هي (52.58%)

الاستنتاجات والتوصيات

1. الاستنتاجات

مما تقدم من نتائج البحث التي تم عرضها تم التوصل الى الاستنتاجات التالية:

1- يتضح ان اغلب الذين حصلوا على قرض من الذين كانوا من الذكور المتزوجين، والذين هم من المتعلمين الغير حاصلين على وظائف خاصة بهم، من الذين يمتلكون بطاقات بنكية.

2- تم ملاحظة ان المتغيرات للذين من الممكن اعطائهم القرض حسب نتائج شجرة القرارات حيث تمحورت حول Credit History و Loan Amount و Applicant Income و Co-applicant Income

3- تم ملاحظة ان المتغيرات للذين من الممكن اعطائهم القرض حسب نتائج الانحدار اللوجستي حيث تمحورت حول Credit History و Applicant Income و Loan Amount و Co applicant Income و Gender

4- من خلال النتائج تم ملاحظة ان طريقة شجرة القرارات كانت اعلى دقة من طريقة الانحدار اللوجستي.

2. التوصيات

بناء على ما جاء من الاستنتاجات تم كتابة التوصيات التالية

1- زيادة نشر الوعي في المصرف وبالأخص الجهة التنفيذية وبيان اهمية تنقيب البيانات من خلال توضيح مدى الحاجة الملحة لاتخاذ القرار المستندة على المعرفة لدورها الكبير في تجنب المصرف مخاطر عدم تسديد المقترضين.

2- تطوير وتدريب متخذين القرارات على التقنيات الحديثة لكي يسهل عليهم التوصل الى المعلومات المطلوبة في اتخاذ القرار دون الحاجة الى المعلومات من جهة اخرى مما يساعد على الحفاظ على سرية وامنية المعلومات الخاصة بالزبائن.

3- استخدام وسائل تقنيات تنقيب بيانات واساليب احصائية جديدة تتلاءم من طبيعة البيانات.

المصادر

[1] Al Janabi, M.A., 2016. Value at risk prediction under illiquid market conditions: A comparison of alternative modeling strategies. *Risk Management in Emerging Markets. Emerald Group Publishing Limited*, pp.253-291.

[2] Aljifri, K. and Ahmad, H.I., 2019. Preferred valuation techniques in the UAE: A comparative study of financial and nonfinancial sectors. *International Journal of Economics and Financial Issues*, *9*(1), p.51.

[3] Bagley, S. C., White, H., & Golomb, B. A. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. Journal of Clinical Epidemiology, (2001). 54(10), 979-985.

[4] Brown M,. 2014 - Data Mining For Dummies, a john wiley & sons inc, Canada

[5] Camp, Lawrence, "Problem Loan Management Loan Workout for the Palestinian Banker" FMI May 2004.

[6] Charles, Tapiero (Risk and financial management: mathematical and computational methods) Hoboken, NJ: Wiley, 2004

[7] Dennies Cox, Michael Cox, The Mathematics of Banking and Finance, John Wiley& Sons Ltd, West Sussex, England. , (2006)

[8] Erik Banks and Richard Dunn, "Practical risk management: an executive guide to avoiding surprises and losses", John Wiley & Sons Ltd, England, 2003, p15.

[9] Han, J. and Kamber,M. 2012- Data Mining: Concepts And Techniques, Morgan Kaufmann Publishers, USA.

[10] Menard, S. W.. Applied logistic regression analysis (quantitative applications in the social sciences) (2nd ed.). Thousand Oaks, CA: Sage Publications. (2001)

[11] Nigam, K., J. D. Lafferty, and A. McCallum. Using maximum entropy for text classification. IJCAI-99 workshop on machine learning for information filtering. (1999)