التنبؤ بتعثر القروض المصرفية باستخدام التحليل التمييزي وخوارزميات التنقيب في البيانات دراسة مقارنة بين الأساليب الاحصائية و أساليب الذكاء الاصطناعي في تقييم القروض المصرفية

الدكتورة: راميا الجبيلي

قسم الإحصاء والبرمجة – كلية الاقتصاد– جامعة تشرين– سوريا

ملخص

يلعب النشاط الاقتراضي دوراً هاماً في نجاح المصرف، فهو يشكل المصدر الرئيسي لايراداتة و بدونة يفقد المصرف وظيفتة الأساسية كوسيط مالي في الاقتصاد. لكنة في ذات الوقت نشاط مفعم بالمخاطر. الأمر الذي يتطلب استخدام أساليب و استراتيجيات تقلل من هذه المخاطر. تتيح بعض الأساليب الاحصائية و أساليب الذكاء الاصطناعي القاعدة التي تمنح متخذ القرار القدرة على التنبؤ بحالة القرض قبل المنح و بالتالي تجنب خطر التعثر.

يتناول هذا البحث مقارنة تصنيف القروض المصرفية لدى المصرف الزراعي السوري باستخدام أحد الأساليب الاحصائية (التحليل التمييزي) وبعض أساليب الذكاء الاصطناعي (شجرة القرارات، الجار الأقرب، الشبكات العصبونية الاصطناعية).

وقد توصلت الدراسة الى أن استخدام كل من الأساليب السابقة يمكّن المصرف من التنبؤ بحالة القرض، لكن تعتبر أساليب الذكاء الاصطناعي و على وجة الخصوص أسلوب الشبكات العصبية الاصطناعية هو الأسلوب الأفضل في تصنيف القروض المصرفية وذلك حسب معايير التقييم المستخدمة (MAE,R2, RMSE, MSE,) وبحسب دقة التنبؤ التي تم حسابها.

الكلمات المفتاحية: القروض المصرفية، التحليل التمييزي، الشبكات العصبية الاصطناعية، شجرة القرارات، خوارزمية الجار الأقرب.

A comparative study between statistical methods and artificial intelligence methods in evaluating bank loans

Abstract

The borrowing activity plays an important role in the success of the bank, as it constitutes the main source of its revenues, and without it the bank loses its primary function as a financial intermediary in the economy. At the same time, it is a risky activity. Which requires the use of methods and strategies that reduce these risks. Some statistical and artificial intelligence methods provide the basis that gives the decision maker the ability to predict the loan situation before granting, thus avoiding the risk activity. This research deals with a comparison of the classification of bank loans at the Syrian Agricultural Bank using one of the statistical methods (discriminant analysis) and some methods of artificial intelligence (decision tree, nearest neighbor, artificial neural networks). The study found that using each of the previous methods enables the bank to predict the loan situation, but artificial intelligence methods, in particular, artificial neural networks, are the best method for classifying bank loans, according to the evaluation criteria used (MSE, RMSE, MAE, R2) and according to the prediction accuracy that was calculated.

Keywords: bank loans, discriminant analysis, artificial neural networks, decision tree, nearest neighbor algorithm.

مقدمة:

يتعرض المصرف ضمن نشاطة لأنواع مختلفة من المخاطر(مخاطر السوق، مخاطر الائتمان، مخاطر التشغيل)، لكن تعتبر مخاطر الائتمان من المخاطر المالية الأكثر حدوثا و التي تهدد سلامة ومتانة المصرف .على الرغم من التقدم في تقنيات قياس المخاطر الائتمانية إلا أنة يبقى موضوعاً مهماً ومدروساً على نطاق واسع في قرارات الإقراض المصرفي وذلك نتيجة التطور المتسارع للمعاملات المصرفية وزيادة نطاق عمل المؤسسات المصرفية و ارتباطها الشديد بعدم التأكد وبالتالي صعوبة بناء نموذج أمثل قادر على قياس مخاطر القروض الممنوحة نتيجة عجز النماذج التقليدية في مواجهة مشكلة القروض المتعثرة المتزايدة باستمرار. مما يستدعي استخدام تقنيات متطورة لإدارة المخاطر الائتمانية التي ترتكز على التقدم الحاصل في مجال الذكاء الاصطناعي المستخدم في مجالات كثيرة في محاولة لبناء أنظمة مالية قوية وسليم. من بينها تقنية الشبكات العصبية الاصطناعية و خوارزمية الجار الأقرب و شجرة القرارات والتي اضحت طرق فعالة في ادارة وتسيير المخاطر في بيئة الأعمال.

1. الجانب النظري:

1.1. الدراسات السابقة:

1. دراسة [21] بعنوان: " Neural Network Approach for Credit Risk Evaluation

هدفت الدراسة الى تطبيق اسلوب الشبكات العصبونية الاصطناعية في تقييم مخاطر الائتمان الممنوح للمؤسسات الايطالية صغيرة الحجم حيث تم اقتراح نهجين:

- استخدام الشبكات العصبونية ذات التغذية الامامية.
- اقراح بنية هيكلية خاصة للشبكة العصبونية.

وقد توصلت الدراسة الى جدارة الشبكات العصبونية في تصنيف المؤسسات المقترضة الى سليمة وعاجزة.

2. دراسة [13] بعنوان: محاولة تقدير خطر عدم تسديد القرض باستعمال طريقة القرض التنقيطي والتقنية العصبية الاصطناعية بالبنوك الجزائرية، دراسة حالة:البنك الجزائري للتنمية الريفية". BADR

هدف البحث الى استخدام تقنيات الشبكات العصبية الاصطناعية وطريقة القرض التنقطي في بنك الفلاحة و التنمية الريفية لتصنيف القروض الى مؤسسات سليمة واخرى عاجزة، وذلك لأجل محاولة تقدير خطر عدم تسديد القرض وقد وصلت دقة التصنيف 96,2% بطريقة القرض التنقييطي و 100% بطريقة الشبكات العصبية الاصطناعية.

3. دراسة[14] بعنوان:"تقنية الشبكات العصبية الاصطناعية كأحد أساليب ذكاء الأعمال لتسيير مخاطر القروض: دراسة حالة البنك الجزئري الخارجي."

هدفت الدراسة الى تطبيق تقنية الشبكات العصبية الاصطناعية وهي أحد الأساليب المتعددة المعايير لمحاولة تقدير خطر القرض لمساعدة متخذ القرار بالمؤسسة البنكية. ومحاولة تطبيقها على مجموعة من مؤسسات استفادت من قرض على الأقل لدى بنك الجزائر الخارجي، وكالة بن مهيدي بين فترتي (2005–2008)، وصلت دقة التصنيف باستخدام الشبكات العصبونية في مرحلة التدريب 100% وفي مرحلة الاختبار 90% بينما وصلت دقة التصنيف باستخدام القرض التنقيطي الى 90.7% في عينة البناء و 90% في عينة الاختبار.

4. دراسة[25] بعنوان A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks model

هدفت الدراسة الى اختبار القدرة التنبؤية لثلاثة نماذج (التحليل التمييزي، التحليل اللوغاريتمي، التحليل الاحتمالي) للتنبؤ بالفشل المالي للشركات العامة الصناعية في تايوان. وخلصت الدراسة إلى أفضلية التنبؤ بالفشل المالي للشركات العامة في تايوان باستخدام التحليل اللوغاريتمي. حيث بلغت نسبة التنبؤ الصحيح باستخدام التحليل التمييزي79.49% ، و باستخدام التحليل الوغاريتمي 84.62%، و باستخدام التحليل الاحتمالي 51.06%.

5. دراسة[22] بعنوان: A Theoretical Approach to Financial Distress Prediction Modeling. Managerial Finance

هدفت الدراسة الى اختبار الأساس النظري لنمذجة التنبؤ بالفشل المالي في ثماني دول، تم استخدام نموذج الانحدار اللوجستي باستخدام معطيات بانل و تقنية الشبكات العصبية الاصطناعية. حيث توصلت الدراسة الى أفضلية التنبؤ باستخدام الشبكة العصبية الاصطناعية حيث حققت نسبة تنبؤ صحيح بمقدار 89.88% بينما حقق نموذج الانحدار اللوجستي نسبة تنبؤ صحيح بمقدار 87.18% .

6. دراسة[9] بعنوان: "التنبؤ بالتعثر المالي لبعض المؤسسات العاملة في القطاع الخاص الصناعي بولاية غرداية باستخدام التحليل التمييزي خلال الفترة 2009-2014

هدفت الدراسة إلى معرفة مدى قدرة المؤشرات المالية على التنبؤ بتعثر المؤسسات الصناعية. توصلت الدراسة الى بناء نموذج باستخدام التحليل التمييزي يتكون من أربعة مؤشرات مالية، حيث بلغت دقة التنبؤ 100%.

من الدراسات السابقة نستخلص استخدام الباحثين للعديد من التقنيات (الشبكات العصبية الاصطناعية، التحليل التمييزي، التحليل اللوغاريتمي و النماذج الاحتمالية) بهدف تصنيف القروض المصرفية. بينما تفتقر الدراسات السابقة الى استخدام خوارزمية الجار الأقرب و شجرة القرارات في تصنيف القروض. تتميز دراستنا الحالية عن الدراسات السابقة في الوقوف على مدى فاعلية هذه الخوارزميات في التنبؤ بتصنيف القروض ومقارنتها مع التحليل التمييزي و الشبكات العصبية الاصطناعية بهدف الوصول الى التصنيف الأفضل.

1.2. مشكلة البحث:

نظرا للارتباط الوثيق بين العمل المصرفي و المخاطر الناتجة عن الائتمان توجب على ادارة البنوك اتخاذ مجموعة من الاجراءات و التدابير بغية الحد من تلك المخاطر وذلك عن طريق تحري الدقة في اختيار العملاء. مما أدى إلى ظهور العديد من النماذج في مجال قياس الخطر منها نموذج (beaver,taffler et tisshow, scoring) وبالرغم مما توفره هذه النماذج من مزايا إلا أنها قائمة على أساس علاقات خطية عاجزة عن

استيعاب كل الحالات الممكنة المعبرة عن طبيعة خطر القرض. وبغية تخطي القصور الذي يعتري هذه النماذج ، يرى البعض أن تبني نماذج أكثر واقعية مثل نماذج الذكاء الاصطناعي الذي يعالج التراكيب الغير خطية قد يعالج المسألة بشكل أفضل ويقلل مخاطر الائتمان ويزيد العائد، ومن بين هذه الأنظمة ،الشبكات العصبية الاصطناعية، خوارزمية الجار الأقرب، شجرة القرارات بشكلها التصنيفي، ولهذا الغرض سنقوم من خلال هذه الدراسة

بالمقارنة بين التحليل التمييزي كأحد الأساليب الاحصائية و الشبكات العصبية الاصطناعية و خوارزمية الجار الأقرب و شجرة القرارات للتنبؤ بحالة القرض في المصرف الزراعي في محافظة طرطوس. ولعل أهم مايميز هذه النماذج قدرتها على التعلم بطريقة التجربة و الخطأ مما جعلها نظاما

يتلاءم و أهداف تقدير خطر القرض وبناء على ما سبق يمكن طرح التساؤلات التالي:

- ما مدى دقة النماذج المقترحة في تصنيف القروض المصرفية في المصرف الزراعي في طرطوس؟
- هل يتمتع أحد هذه النماذج المقترحة بالأفضلية في تصنيف القروض المصرفية؟

1.3. فرضيات الدراسة:

1. لا توجد فروق ذات دلالة احصائية بين نتائج تصنيف القروض وفق نموذج التحليل التمييزي وتصنيف المصرف للقروض الممنوحة والتي لم يتم سدادها بعد.
2. لا توجد فروق ذات دلالة احصائية بين نتائج تصنيف القروض وفق نموذج الشبكات العصبية الاصطناعية وتصنيف المصرف للقروض الممنوحة والتي لم يتم سدادها بعد.
3. لا توجد فروق ذات دلالة احصائية بين نتائج تصنيف القروض وفق خوارزمية الجار الأقرب وتصنيف المصرف للقروض الممنوحة والتي لم يتم سدادها بعد.
4. لا توجد فروق ذات دلالة احصائية بين نتائج تصنيف القروض وفق شجرة القرارات وتصنيف المصرف للقروض الممنوحة والتي لم يتم سدادها بعد.
5. لايوجد أفضلية لأحد النماذج المقترحة في تصنيف القروض المصرفية للمصرف الزراعي في طرطوس.

1.4.أهمية الدراسة : تنبع أهمية البحث من الدور الذي يمكن أن يلعبة نماذج الذكاء الصنعي و الأساليب الاحصائية في اتخاذ القرارات الائتمانية وذلك من خلال الاستفادة منها في مجال المفاضلة بين القروض المقدمة وتصنيفها مما يخفض من خطر خسارة القرض الممنوح وتخفيض العبء في دراسة ملفات الاقراض.

1.5. أهداف الدراسة: تهدف الدراسة الى توضيح دور كل من النماذج المقترحة في التنبؤ بقدرة طالب القرض على سداد القرض، وبالتالي اتخاذ القرار المناسب في منح أو عدم منح هذا الائتمان.

1.6. متغيرات الدراسة:

المتغيرات المستقلة :مبلغ القرض، مدة القرض، طريقة السداد، الغاية من القرض، سعر الفائدة المحسوبة على القرض، الضمانات، سمعة العميل، النشاط الذي يزاوله العميل.

المتغير التابع :تصنيف القروض بحسب معيار الجودة (جيد، متعثر، رديء.)

1.7. زمان ومكان البحث:

الحدود المكانية :المصرف الزراعي السوري في محافظة طرطوس.

الحدود الزمانية :الفترة الممتدة بين 2010-2016

1.8. منهج البحث: لدراسة الموضوع واختبار الفرضيات تم الاعتماد على المنهج الوصفي في وصف النشاط الاقراضي وشرح آلية عمل كل من النماذج المقترحة. كما اعتمدنا على المنهج التحليلي من خلال تطبيق كل من النماذج المقترحة للوصول الى تصنيف القروض المصرفية.

2.1. القروض المصرفية ومخاطرها:

القروض المصرفية :تعرف بأنها" تلك الخدمات المقدمة للعملاء التي يتم بمقتضاها تزويد الأفراد و المؤسسات و المنشات في المجتمع بالأموال اللايمة على أن يتعهد المدين بسداد تلك الأموال و فوائدها و العمولات المستحقة عليها والمصاريف دفعه واحدة او على أقساط في تواريخ متفق محددة .ويتم تدعيم هذه العلاقة بتقديم مجموعة من

الضمانات التي تكفل للمصرف استرداد أمواله في حال توقف العميل عن السداد بدون أية خسائر. [15] "

2.1.1. خصائص القرض المصرفي: تعتمد المصارف على مجموعة من الخصائص لتوصيف القروض الممنوحه من قبلها منها: [8]

- قيمة القرض: يمثل الأموال التي تمنح للعميل حيث يحدد من قبل الجهة المانحة للقرض بحسب نوع القرض و الملاءة المالية للعميل.
- مدة القرض : الفترة الزمنية الممتدة بين منح القرض واستحقاقة.
- معدل سعر الفائدة: وهو ثمن استخدام الأموال، ويختلف سعر الفائدة باختلاف قيمة القرض ومدة القرض ودرجة المخاطرة.
- الضمانات: وهي عبارة عن اصول عينية و شخصية تقدم لمواجهة مخاطر عدم السداد.
- طريقة السداد: هل سيتم تسديد القرض دفعه واحدة أم على أقساط دورية.
- الهدف من القرض : الغرض الذي سوف يستخدم العميل فيه الأموال المقترضة.

2.1.2. أهمية القروض المصرفية: [3]

- تعتبر القروض المصرفية المصدر الرئيسية الذي يعتمد علية المصرف في تأمين ايراداته.
- يعتبر القرض وسيله مناسبة لتحويل الأموال من شخص لآخر.
- تلعب القروض دور في مواجهة الانكماش و التضخم.
- تمويل حاجات القطاعات المختلفة من صناعة وزراعة وتجارة وخدمات.
- زيادة القوة الشرائية للأفراد من خلال تحسين وخلق فرص عمل.

2.1.3. مفهوم القروض المصرفية المتعثرة: قروض عجز المقترضين عن سدادها في تاريخ الاستحقاق إما بسبب عدم الرغبة في ذلك أو نتيجة حدث غير محسوب له وظروف أثرت على قدرتهم على الوفاء باستحقاقاتهم". [30]

تم تحديد مفهوم القروض المتعثرة في سورية وفقا لقرار مجلس النقد و التسليف رقم(597/م ن/ب4)بتاريخ(9 19/12/2009) المتضمن اعتماد تصنيف الديون و تكوين المخصصات والاحتياطات لتنظيم الأعمال المصرفية في سورية على أنة:

"القرض الذي يتأخر صاحبة عن السداد المترتب علية من أقساط لمدة تتراوح من180 إلى 395 يوم، ليتم بعد ذلك جدولته واعتباره دينا رديئ ". [35] الجدول التالي يوضح آلية جدولة القروض المتأخرة في المصارف السورية:

الجدول (1): تصنيف القروض المتأخرة في المصارف السورية.

ر.ت	الفئة	عدد أيام التأخر عن السداد
أ-	ديون دون المستوى	90 يوما الى 179 يوم
ب-	ديون مشكوك بتحصيلها	180 يوما الى 395 يوم
ت-	ديون رديئة	360 يوم فأكثر

المصدر: قرار مجلس النقد و التسليف رقم (597 /م ن /ب4)

2.1.4. الاثار السلبية للقروض المتعثرة [30] : لتعثر القروض أثار سلبية تطال كل من المقترض و المصرف الممول.

بالنسبة للمقترض:

- تزداد الخسائر المتحققة في المشروع نتيجة تعاظم حجم دينة المتعثر من خلال تراكم الأقساط و الفوائد وغرامات التأخير.
- صعوبة الحصول على السيولة اللايمة لاستمرار عجلة التشغيل.
- زيادة حجم الطاقات العاطلة عن العمل في المشروع.

بالنسبة للمصرف:

- سوء توزيع الموارد في المصرف:إن بيادة نسبة القروض المتعثرة ، يؤدي الى اتباع سياسة حذرة في منح التسهيلات الجديدة مما يؤدي الى حرمان مشاريع رابحة من فرصة الحصول على التمويل من المصرف.
- الآثار المترتبه على الأرباح :يترتب على القروض المتعثرة آثار واسعه على عائدات المصرف نتيجة تجميد أمواله في قروض لمشاريع تعثرت مما يحرمه من امكانية استثمارها بمشاريع رابحة.هذا بالإضافة الى تخصيص جزء من الايرادات الكلية في بيادة مخصصات الديون المشكوك في تحصيلها.

- تأثيرها على المكانة السوقية للمصرف :تحد القروض المتعثرة من قدرة المصرف على النمو و التوسع بالإضافة الى انخفاض ثقة المودعين فيه.

2.1.5. واقع الاقراض في المصرف الزراعي السوري: يعتبر المصرف الزراعي السوري من أقدم المصارف العاملة في سورية، حيث تأسس عام 15/8/1888،وقد تخصص المصرف الزراعي السوري منذ تأسيسة بالائتمان الزراعي لتمويل النشاطات الزراعية بأشكالها المتعددة، وقد بلغ رأسمالة 10 مليارات ليرة سورية بموجب المرسوم التشريعي رقم 30/4/2005. يلعب المصرف الزراعي السوري دورا حيويا في التنمية الزراعية في سورية، ومما ساعد المصرف على أداء هذا الدور انتشارة الواسع الذي يغطي جميع المناطق الريفية في سورية. وصل عدد فروعة الى 106 فروع منتشرة بكافة المحافظات والمدن السورية.

يمنح المصرف الزراعي السوري القروض قصيرة الأجل لتمويل نفقات الزراعة من حراثة وحصاد وري ومحروقات وقيمة البذور والأسمدة والمخصبات ومواد المكافحة، كما يمنح القروض متوسطة الآجل لشراء الآليات والآلات والأدوات اللايمة للاستثمار الزراعي، والقروض طويلة الآجل لتصنيع المنتجات الزراعية بشقيها النباتي والحيواني وإنشاء مشاريع الري والصرف. [1]

2.2. التحليل التمييزي(DA):

مفهوم التحليل التمييزي: يعد اسلوب التحليل التمييزي من الأدوات المهمة في التحليل الاحصائي المتعدد المتغيرات المهتمة بالتمييز بين مجموعتين أو أكثر عن طريق دوال تمييزية محددة تعمل على تعظيم الفروق بين متوسطات المجموعات وتقليل التشابة في الوقت ذاتة، فالعملية التصنيفية هي العملية اللاحقة لعملية تكوين الدالة التمييزية .[33]
وهناك عدة دوال تمييزية:[10]

1. دالة التمييز الخطي Linear Discriminant Analysis
2. دالة التمييز التربيعي Quadratic Discriminant Analysis

يهدف التحليل التمييزي إلى:

1. تصميم دوال المتغير التابع، المعروفة بدوال التمايز للمتغيرات المستقلة وتحديد الأفضل منها في التمييز بين فئات المتغير التابع. وتقوم هذه الدوال بالتنبؤ بموقع

المشاهدات الجديدة، وذلك من خلال تعظيم التباين بين المجموعات وتصغير التباين داخل المجموعة.

2. اختبار مدى وجود فروق ذات دلالة إحصائية بين مجموعات المتغير التابع بالنسبة للمتغيرات المستقلة.
3. التعرف على أكثر المتغيرات المستقلة تأثيراً في تصنيف مجموعات المتغير التابع، أي تحديد المتغيرات المستقلة التي تسهم بأكبر قدر من الاختلاف بين فئات المتغير التابع.
4. تقسيم المتغير التابع إلى فئات بناء على قيم المتغيرات المستقلة.
5. تقييم دوال التصنيف (معنوية أم غير معنوية).
6. تصنيف مشاهدات جديدة إلى أحد مجموعات المتغير التابع في حالة عدم المعرفة المسبقة بالمجموعة التي تنتمي إليها تلك المفردات[18] .

يستخدم التحليل التمييزي في الحالات التي تكون فيها المجموعات معروفة مسبقًا. الهدف من التحليل التمييزي هو تصنيف لملاحظة أو عدة ملاحظات في هذه المجموعات المعروفة. على سبيل المثال، في التصنيف الائتماني، يعرف البنك من التجربة السابقة أن هناك عملاء جيدين (يسددون قروضهم دون أي مشاكل) وعملاء سيئون (أظهروا صعوبات في سداد قرضهم). عندما يطلب عميل جديد قرضًا، يتعين على البنك أن يقرر ما إذا كان سيعطي القرض أم لا بالاعتماد على السجلات السابقة للعملاء السابقين لدى البنك. حيث تصنف قاعدة التمييز العميل إلى واحدة من المجموعتين الحاليتين (جيد،سيء)، ويجب أن يقيِّم التحليل التمييزي مخاطر "القرار السيئ" المحتمل[33].

2.2. مراحل التحليل التمييزي:

المرحلة الأولى: تحليل البيانات (وصفي)

1. حساب الإحصائيات النموذجية (الإحصاء الوصفي)
2. الاختبار الإحصائي (إحصاءات ويلكس)

المرحلة الثانية:

- حساب الوظيفة المميزة

1. البحث عن معاملات دالة التمييز المعيارية لمعرفة ما إذا كان للتنبؤ دور مهم في تحديد المجموعة
2. أكبر معامل للتنبؤ في دالة التمييز المعيارية هو المعامل الذي يلعب دورًا مهمًا في دالة التمييز.
3. حساب معامل الارتباط الجزئي.
4. القيود: يكون تفسير أكثر من حدث أكثر صعوبة

المرحلة الثالثة:

- مقارنة الوظائف التمييزية

1. ما هي الوظيفة التمييزية الأقوى في تمييز المجموعات
2. تحقق من قيم eigenvalues ، المعروفة أيضًا باسم القيم المفردة أو الجذور المميزة. كل دالة مميزة لها قيمة واحدة. يعكس قيمة مقدار التباين الموضح في تجميع متغيرات توقع أول في دالة مميزة
3. نسبة القيم الذاتية لتقييم النسبة المهمة لوظيفة التمييز

- الحصول على القيم الذاتية[29]

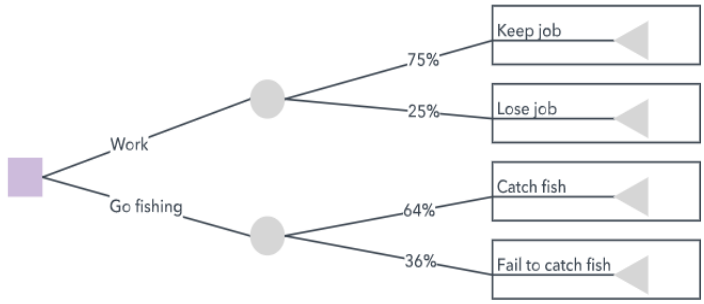
من أجل فهم كيفية استخدام تقنية التحليل التمييزي، يجب معرفة النقاط الأساسية التي تحدد انتماء المتغير إلى مجموعتين أو أكثر؛ حيث يمكن أن تكون هذه البيانات بيانات مالية، مدن، بلدان، بيانات طبية أو التعليمية، البيانات التاريخية أو أيا كان آخر. ويجب أن يكون لدينا على الأقل متغيرين مثل 2X & X1 مع 2n& n1 والتي تنتمي إلى كل مجموعة، ويجب أخذ المتغيرات في كل مجموعة من المجموعات ذات توزيع طبيعي متعدد المتغيرات، ومصفوفات التغاير التي يجب أن تكون المتغيرات في كل مجموعة قريبة تقريبًا من بعضها البعض، بعد أن يكون لدينا تحقق من هذه النقاط، أولاً وقبل كل شيء سنفترض أن لدينا أكثر من متغيرين لكل منهما المجموعة.[19]

2.3. شجرة القرارات decision tree:

شجرة القرار هي نموذج استكشافي يظهر على شكل شجرة يمثل كل فرع من فروعها سؤال تصنيفي. يسمح للفرد أو المنظمة بمقارنة الإجراءات الممكنة بناءً على تكاليفها واحتمالاتها وفوائدها. يمكن استخدامها لرسم خوارزمية تتنبأ بالاختيار الأفضل رياضيًا.

تبدأ شجرة القرار عادةً بعقدة واحدة تتفرع إلى نتائج محتملة. كل من هذه النتائج تؤدي إلى عقد إضافية، والتي تتفرع إلى احتمالات أخرى. هذا يعطيها شكل الشجرة.

هناك ثلاثة أنواع مختلفة من العقد: عقد الفرصة وعقد القرار وعقد النهاية. في الشكل (1) تُظهر عقدة فرصة(ممثلة بدائرة) احتمالات نتائج معينة. تُظهر عقدة القرار(ممثلة بمربع) قرارًا يتعين اتخاذه، وتوضح عقدة النهاية النتيجة النهائية لمسار القرار[28]



الشكل (1) مثال توضيحي لشجرة القرارات [28]

في التعلم الآلي ، تعد شجرة القرار نموذجًا تنبؤيًا ؛ وهي تمثل علاقة تعيين بين سمات الكائن وقيم الكائن. تمثل كل عقدة في الشجرة كائنًا ، ويمثل كل مسار متفرع قيمة سمة محتملة ، وتتوافق كل عقدة طرفية مع الكائن الذي يمثله المسار من عقدة الجذر إلى عقدة الورقة. تحتوي شجرة القرار على مخرجات واحدة فقط. [6]

مزايا شجرة القرار[26]:

1. سهل الفهم والتفسير.
2. يتطلب القليل من إعداد البيانات: غالبًا ما تتطلب التقنيات الأخرى تطبيع البيانات ، ويجب إنشاء متغيرات وهمية وإزالة القيم الفارغة.

3. تكلفة استخدام الشجرة (أي التنبؤ بالبيانات) هي لوغاريتمية في عدد نقاط البيانات المستخدمة لتدريب الشجرة.
4. قادرة على التعامل مع البيانات العددية والفئوية: عادةً ما تكون التقنيات الأخرى متخصصة في تحليل مجموعات البيانات التي تحتوي على نوع واحد فقط من المتغيرات.
5. قادرة على التعامل مع مشاكل المخرجات المتعددة.
6. يمكن التحقق من صحة النموذج باستخدام الاختبارات الإحصائية. هذا يجعل من الممكن حساب موثوقية النموذج.

2.4 خوارزمية الجار الأقرب K–Nearest Neighbor Algorithm

تعتبر خوارزمية الجار الأقرب من خوارزميات التنقيب Data mining و التي صيغت في بداية الخمسينات من القرن الماضي وهي تعتمد في تصنيف البيانات عل أساليب التعلم البطيئة Lazy Learning. وهي طريقة تصنيف لامعلمية تعتمد على تشابة لنقطة ما لمجموعة تدريبية training set حيث تستخدم نقاط أقرب جيران لتتمكن من تصنيف البيانات وذلك عن طريق تجميع القيم المزودة عبر أنماط التدريب في جوار النقطة الحالية. وهي تستخدم في تصنيف البيانات كما ويتم استخدامها للتنبؤ بالسلاسل الزمنية. ويمكن استخدامها في نماذج الانحدار الخطي والغير خطي. حيث تفترض خوارزمية الجار الأقرب K–NN أن البيانات منشرة في الفضاء على شكل نقاط، يتم حساب المسافة بين هذة النقاط من خلال العديد من الطرق منها (المسافة الإقليدية، المسافة التربيعية) والتي تعد المسافة الاقليدية الأكثر استخداماً. بعد حساب المسافة بين النقاط يتم اختيار الجيران الأقرب للمشاهدة المراد تقديرها حيث يتم اختيار عدد الجيران المراد التقدير بة بالتجريب ويتم اعتماد العدد الذي يعطي أقل خطأ للتقدير. ومن ثم يتم أخذ الوسط الحسابي للجيران الأقرب كتقدير للقيمة المطلوبة[34] .

قبل الخوض في خطوات الخوارزمية سوف نتطرق لطرق حساب المسافة بين النقاط [34]:

1. البعد الاقليدي (Euclidean distance):وهو البعد بين نقطتين مرسومتان بالرسم الانتشاري ويتم حسابة بالصيغة التالية:

$$d(y_i, y_j) = \sqrt{\sum_{k=1}^{n} (y_{ik} - y_{jk})^2}$$

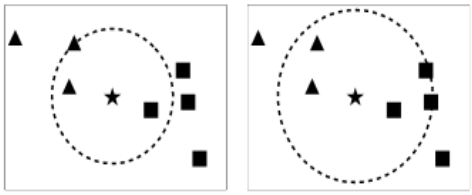
2. البعد التربيعي (square distance): هو عبارة عن مجموع مربعات الفروق بين مشاهدتين ويحسب البعد التربيعي بالصيغة التالية:

$$d(y_i, y_j) = \sum_{k=1}^{n} (y_{ik} - y_{jk})^2$$

خطوات خوارزمية الجار الأقرب في التصنيف [20]:

1. تحدد العامل المتغير(K الباراميتر) K وهو رقم أقرب جيران (لا توجد قيمة مثالية ل k حقيقية تعمل مع كل مجموعة البيانات).
2. تحديد المسافة بين جميع النقاط مع النقطة الغير مصنفة.
3. فرز المسافات لجميع عينات التدريب وحدد الجار الأقرب معتمداً على أقل مسافة.
4. احصل على جميع فئات بيانات التدريب الخاصة بالقيم المفرزة والتي تتدرج تحت K
5. نجد التصنيف الأكثر شيوعًا للادخالات k في قاعدة البيانات الأقرب إلى النقطة الغير مصنفة.

يظهر الشكل (2) مثال تخطيطي لكيفية عمل هذه الخوارزمية. يُظهر الشكل نقطة بيانات غير مصنفة ، ممثلة بنجمة ، والتي يجب تصنيفها باستخدام خوارزمية kNN في مجموعة بيانات ثنائية الأبعاد. تتم الإشارة إلى فئات مجموعة البيانات إما بمثلث أو مربع. لتصور مشاركة البيانات في تصويت الأغلبية ، يتم رسم دائرة باستخدام مسافة أبعد جار مشمول كنصف قطر. يوضح الشكل (a،1) 3. = k أقرب جيران ، حيث يجب تصنيف نقطة البيانات غير المصنفة كمثلث .ومع ذلك ، يوضح الشكل (b،1) عندما 5 =kسيشترك مربعان آخران في تصويت الأغلبية ، مما يتسبب في تعيين فئة النجم على مربع.



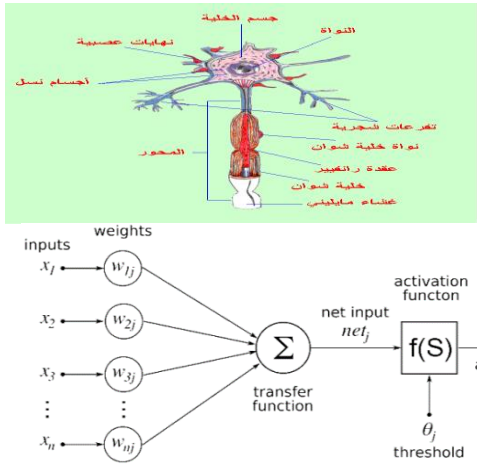
الشكل (2) كيفية عمل خوارزمية الجار الأقرب (المصدر: [20])

تتميز خوارزمية الجار الأقرب بدقة عالية وعدم تأثرها بالقيم المتطرفة كما أنها لاتحتاج الى افتراضات عند ادخال البيانات. و يمكن تطبيقها على البيانات الاسمية و الكمية. ويمكن استخدامها في مشاكل التصنيف و الانحدار. لكن من عيوبها أنها تحتاج الى دقة حساب عالية و الى ذاكرة عالية حيث تخزن جميع بيانات التدريب (فهي حساسة لحجم البيانات فكلما زاد حجم البيانات زادت حاجتها للذاكرة واستغرقت وقت أكبر في التصنيف)

2.5. (الشبكات العصبية الصنعية) Artificial Neural Networks, ANN :

2.5.1. المفاهيم الأساسية لهيكلة الشبكات العصبية الصنعية: تعرف الشبكات العصبية الصنعية Artificial Neural Networks, ANN على أنها تقنيات حسابية مصممة لمحاكاة الطريقة التي يؤدي بها الدماغ البشري مهمة معينه، وذلك عن طريق معالجة ضخمة موزعة على التوازي، ومكونة من وحدات معالجة بسيطة، هذه الوحدات تمثل عناصر حسابية تسمى عصبونات أو عقد(Neurons, Nodes) والتي لها خاصية عصبية حيث أنها تقوم بتخزين المعلومات والمعارف، حيث أن للإنسان وحدات إدخال توصله بالعالم الخارجي وهي حواسه الخمس، كذلك تحتاج الشبكات العصبية الصنعية لوحدات إدخال، ووحدات معالجة يتم فيها عمليات حسابية لضبط الأوزان فيها، والتي نحصل من خلالها على ردة الفعل المناسبة لكل مدخل من مدخلات الشبكة. كما تتشابه الشبكة العصبونية الصنعية مع الدماغ البشري في أنها تكتسب المعرفة بالتدريب وتخزن هذه المعارف باستخدام قوى وصل داخل العصبونات تسمى الأوزان التشابكية، لذا تتكون الشبكات العصبونية من مجموعة من وحدات المعالجة تدعى بالعصبونات.**[2]** :

و الشكل (3) يبين مقارنة بين العصبون البيولوجي والعصبون الصنعي كالآتي:



العصبون البيولوجي العصبون الصنعي

الشكل (3) مقارنة بين العصبون البيولوجي والعصبون الصنعي (المصدر [2])

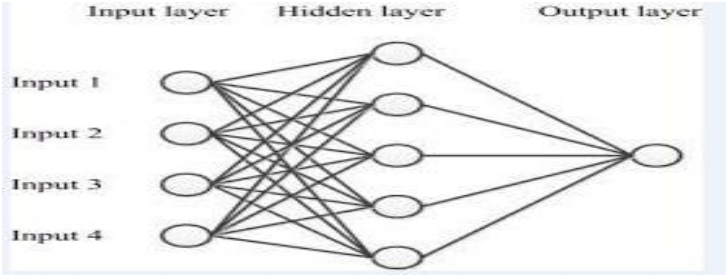
2.5.2. بنية الشبكة العصبية الصنعية: [24]

1. تتم المعالجة في عناصر معالجة بسيطة تدعى العصبونات.
2. تمر الإشارات بين العصبونات عبر خطوط ربط.
3. يرفق كل خط بوزن معين (قيمة عددية) والتي تضرب مع الإشارات الداخلة إلى العصبون.
4. يطبق على كل عصبون تابع تنشيط (غير خطي عادة) إلى دخل الشبكة ليحدد إشارة الخرج الناجمة عنه.

وتوصف الشبكات العصبية بالآتي: شكل الترابط بين العصبونات، الطريقة التي تحدد أوزان هذه الترابطات (التدريب أو التعليم)، نوع تابع التنشيط المستخدم.

إن لكل عصبون حالة داخلية خاصة به تدعى الفعالية والتي يحصل عليها من تطبيق تابع رياضي معين على معطيات الدخل ونتيجته تستقبل من قبل العصبون. كما يرسل كل عصبون فعالية كإشارة دخل للعديد من العصبونات الأخرى المرتبطة معه ويجب أن نلاحظ أن العصبون يرسل إشارة واحدة فقط في نفس اللحظة ولكن بإمكانه إرسال هذه الإشارة إلى العديد من العصبونات الأخرى. [7]

ويوضح الشكل (4) التالي نموذجاً بسيطاً لشبكة عصبية صنعية مؤلفة من ثلاث طبقات.



الشكل (4) شبكة عصبية صنعية مؤلفة من ثلاث طبقات [7]

ويتضح من الشكل السابق أن الشبكة العصبية تتكون من ثلاث طبقات على النحو التالي:

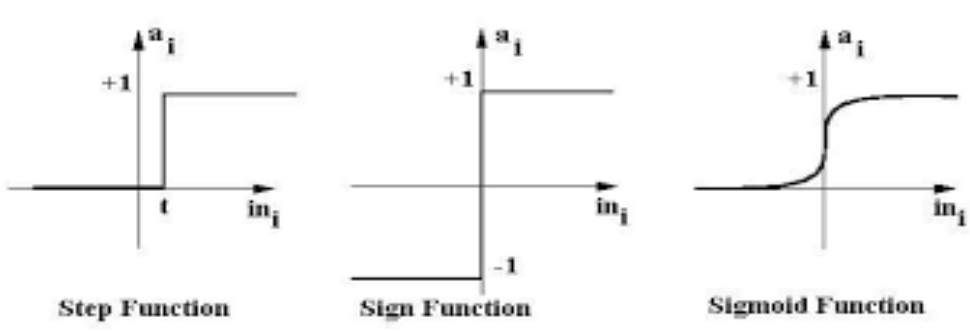
a.طبقة المدخلات (Input Layer).(البيانات)

b.طبقة مخفية (Hidden Layer).(الطبقة التي تتم بها معالج البيانات للتنبؤ للمستقبل وهي مخفية)

c.طبقة المخرجات (Output Layer).(قيم البيانات المتنبأ بها)

2.5.3. مفاهيم تقنيات الشبكات العصبية الصنعية: [5]

1. الطبقات (Layers): وهي عبارة عن العدد الكلي لطبقات الشبكة العصبية، وهي عبارة عن طبقة المدخلات، وطبقة المخرجات، بالإضافة إلى طبقة مخفية واحدة أو عدد من الطبقات المخفية.
2. حجم الطبقة (Layer size): يرتبط بعدد العقد (nods) الموجودة في كل طبقة وهو يحدد بحجم الذاكرة المحددة بالبرنامج الخاص بتشغيل الشبكة العصبية الصنعية.
3. الدالة العصبية المستخدمة (Neural function): وتدعى بدالة التفعيل أو دالة التنشيط (Activation function) والتي يمكن أن تأخذ عدة أشكال، حيث يوضح الشكل(5) التالي بعض أشكال دالة التفعيل:



الشكل (5) بعض أشكال دالة التفعيل(التنشيط) [5]

2.5.4. خصائص الشبكات العصبية الصنعية: أهم خصائص الشبكات العصبية [4]

1. تعتمد على أساس رياضي قوي في تحليل بيانات الدراسة الصحية.
2. تمثل إحدى تطبيقات تكنولوجيا التشغيل الذكي للمعلومات التي تقوم على محاكاة العقل البشري.
3. تقبل أي نوع من البيانات الكمية أو النوعية.
4. لها القدرة على تخزين المعرفة المكتسبة من خلال الحالات التي يتم تشغيلها على الشبكة.
5. يمكن تطبيقها في العديد من المجالات العلمية المختلفة، ومن أهمها المجالات الاقتصادية والتنبؤ بالطلب.

2.5.5. أنواع ونماذج الشبكات العصبية الصنعية: [11]

1. **الشبكات العصبية ذات التغذية الأمامية:** (Feed Forward Neural Networks) وهي الشبكات التي يخلو تركيبها من وجود حلقة مغلقة من الترابطات بين الوحدات المكونة لها. حيث تتكون الشبكة من هذا النوع من طبقتين على الأقل، كما تتواجد في كثير من الأحيان طبقات مخفية بين طبقة المدخلات وطبقة المخرجات، وتنتقل العمليات الحسابية في اتجاه واحد إلى الأمام من طبقة المدخلات إلى طبقة المخرجات عبر الطبقة أو الطبقات المخفية.
2. **الشبكات العصبية ذات التغذية الراجعة:** (Feed Back Neural Networks) في هذا النوع من الشبكات تتعلق المخرجات لبعض العصبونات بنتائج العصبونات ذاتها

في لحظة زمنية سابقة أو بمخرجات عصبونات أخرى، وبالتالي فان الإشارة في هذا النمط من الشبكات العصبية تتجه أماماً أي نحو طبقة الخرج وأيضاً بالاتجاه العكسي لتكون دخلاً لعصبونات أخرى في الوقت ذاته ويطلق على هذا النوع من الشبكات بالشبكات الديناميكية لأن الخرج في لحظة ما يتعلق بالدخل باللحظة ذاتها وبلحظات سابقة أي تمتلك ذاكرة ومن بعض الأمثلة عليها شبكات Hopfield & Elman & Jordan & NARX.

3. **الشبكات العصبية ذات الترابط الذاتي:** (Auto Associative Neural Networks)

وهي الشبكات التي تلعب كافة العناصر المكونة لها دوراً نموذجياً، يتمثل في استقبال المدخلات وبث المخرجات في نفس الوقت.

2.5.6. خطوات تصميم الشبكة العصبية الصنعية: [23]

1. يتم تجميع البيانات التي تستخدم في تدريب أو اختيار الشبكة.
2. يتم تعريف البيانات الخاصة بالتدريب للشبكة، ووضع خطة التدريب والتعلم.
3. بناء هيكل الشبكة وتحديد تركيب الشبكة في عدد المدخلات والطبقات ونوع الشبكة.
4. اختيار طريقة التعلم حسب ما هو متاح من أدوات أو تطوير الشبكة أو القرارات.
5. وضع قيم للأوزان والمتغيرات، ثم بعد ذلك يتم تعديل القيم في الأوزان عن طريق التغذية الأمامية و الراجعة.
6. تحويل البيانات إلى النوع المناسب للشبكة ويتم ذلك عن طريق كتابة معادلة لتجهيز البيانات والبرامج الجاهزة.
7. وفيهما تتم عمليتي التدريب والاختبار من خلال تكرار عرض المدخلات والمخرجات المرغوبة إلى الشبكة، ومنها مقارنة القيم الفعلية (القيم المرغوبة) مع القيم المحسوبة، ثم حساب الفرق بين القيمتين السابقتين (الخطأ)، ثم تعديل الأوزان لتقليل الفرق حتى يصبح فرقاً مقبولاً

8. وهنا تستطيع الشبكة الوصول إلى النتائج المرغوبة من خلال استخدام مدخلات التدريب، وبهذا يمكن الاعتماد على الشبكة في الاستخدام كنظام مستقل قائم بذاته أو كجزء من النظام.

2.5.7. مزايا استخدام نموذج الشبكات العصبية الصنعية: [12]

1. لا تتأثر كثيراً عند إدخال أي مجموعة إضافية من البيانات الجديدة والتي لم تكن موجودة في البداية.
2. لا تتطلب وجود افتراضات قوية بين نوعية البيانات والعلاقة بين المتغيرات كما في الأساليب الإحصائية الأخرى.
3. ليس هناك حاجة لترتيب المتغيرات حسب أهميتها.
4. تعطى قرارات واضحة في النماذج المعروفة، كما يمكنها معالجة مشكلات اتخاذ القرارات والتي تتصف بمدخلات غير واضحة.

3. النتائج والمناقشة:

تهدف الدراسة الى اجراء مقارنة بين عدد من الخوارزميات (شجرة القرارات، الجار الأقرب و الشبكات العصبية الاصطناعية والتحليل التمييزي) للوصول الى الخوارزمية التي تمتلك القدرة الأكبر في تصنيف القروض المصرفية (متعثر،رديء، جيد). للوصول الى الهدف المنشود سنقوم بتطبيق كل من الخوارزميات على عينة الدراسة ومن ثم المقارنة بينها من خلال معايير التقييم التالية (MAE ,R2, RMSE, MSE) للحكم على الخوارمية الأفضل.

يمثل مجتمع الدراسة في مجموعة ملفات القروض الموجودة في المصرف الزراعي في مدينة طرطوس و البالغة3074 حيث قمنا بسحب عينة عشوائية بدون إعادة وتم حساب حجمها وفقا للقانون التالي:

$$n=\frac{N.Z^2.[P(1-P)]}{N.d^2+Z^2.[P(1-P)]}=\frac{3074.(1.96)^2.(0.3\times 0.7)}{3074\times(0.05)^2+(1.96)^2.(0.3\times 0.7)}=292.03\cong 292$$

حيث أن:

N: حجم مجتمع البحث.

n: حجم العينة المدروسة وهو 292

Z: معامل التوبيع الطبيعي المعياري عند مستوى دلالة 0.05 ويساوي 1.96

P: نسبة الظاهرة المدروسة في المجتمع، وهي نسبة التعثر المصرح بها في المصرف وتساوي 30%

D: وهو مستوى الدقة النسبية من تقدير

وصف متغيرات الدراسة:

المتغير التابع : المتعلق بجودة القرض المقدم و تقييم المصرف له ويشمل على المستويات الثلاثة التالية

- القرض الجيد: يلتزم المقترض بسداد التزاماته في مواعيد استحقاقها أو خلال فترة لا تتجاوي2 شهور.
- القرض المتعثر: يتأخر المقترض عن سداد التزاماته لمدة تتراوح من 2 أشهر إلى سنة.
- القرض الرديء: يتأخر فيه المقترض عن السداد لمدة تتجاوي السنة.

المتغيرات المستقلة: بعضها يرتبط بخصائص القرض وبعضها بطالب القرض:

- مبلغ القرض.
- السمعه الاقراضية.
- مدة القرض.
- طريقة السداد.
- سعر الفائدة المحسوبة على القرض الممنوح.
- الغاية من القرض.
- الضمانات المقدمة.
- دخل الزبون الطالب للقرض.

3.1 تصنيف القروض المصرفية باستخدام خوارزمية الجار الأقرب: لتصنيف القروض المصرفية باستخدام الجار الأقرب تم ادخال البيانات الى البرنامج orange وهو من البرامج الأكاديمية المتخصصة في التنقيب في البيانات والذكاء الاصطناعي. حيث تم استخدام أعداد مختلفة ل k (عدد الجيران الأقرب) وقد تم التوصل إلى أنسب قيمة ل k

هي 2 و التي تعطي أفضل تقييم باستخدام معايير التقييم المعتمدة كما يظهرة الجدول رقم (2).

الجدول (2): معايير المقارنة بحسب قيمة k

MSE	RMSE	MAE	R2	قيمة k
0.126	0.355	0.163	0.719	2
0.131	0 .426	0.231	0.596	3
0.205	0.453	0.259	0.543	4
0.218	0.467	0.264	0.513	5
0.224	0.474	0.273	0.5	6
0.229	0.479	0.277	0.489	7

المصدر: الجدول من إعداد الباحثة بالاعتماد على مخرجات برنامج orange.

الجدول (3) التالي يوضح نتائج تصنيف القروض بحسب خوارزمية الجار الأقرب.

الجدول رقم (3): نتائج تصنيف القروض بحسب خوارزمية الجار الأقرب عند k=2

		تقييم المصرف	Predicted group membership			total
			رديء	متعثر	جيد	
original	count	رديء	5	23	1	29
		متعثر	0	87	0	87
		جيد	0	28	148	176
	%	رديء	17.25	79.31	3.44	100.0
		متعثر	0	100	0	100.0
		جيد	0	15.9	84.1	100.0

a.82.19 % of original grouped cases correctly classified.

المصدر: من إعداد الباحثة بالاعتمادعلى مخرجات البرنامج orange

من الجدول السابق نجد أن:

- مجموع حالات التوافق بين التصنيف حسب المصرف والتصنيف الناتج عن خوارزمية الجار الأقرب يساوي مجموع عناصر القطر الرئيسي 5+87+148=240 أما مجموع حالات الاختلاف بين التصنيفين يساوي مجموع العناصر الغير قطرية: 28+1+23=52 و بالتالي فان درجة دقة التنبؤ: $p=\frac{(292-52)}{292}=0.8219$

- تمكن النموذج من تصنيف 148 من أصل 176 قرض جيد تصنيفاً صحيحاً أي بمعدل دقة 84.1%. في حين تم تصنيف 28 قرض جيد ضمن فئة القروض المتعثرة ، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الجيدة 15.9% لصالح مجموعة القروض المتعثرة.
- أما بالنسبة لمجموعة القروض المتعثرة تم تصنيف 87 من 87 قرض تصنيفاً صحيحاً أي بمعدل 100%، ليبلغ خطأ التصنيف في القروض المتعثرة 0%.
- بلغ التصنيف الصحيح ضمن مجموعة القروض الرديئة 5 قرضاً من أصل 29 بمعدل17.25% و بلغ عدد القروض الرديئة المصنفة ضمن مجموعة القروض المتعثرة 23 قروض، و ضمن مجموعة القروض الجيدة 1 قروض، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الرديئة79.31% و 3.44% لصالح مجموعتي القروض المتعثرة و الجيدة على التوالي (بمعدل عام 82.75%).

لتقييم دقة التصنيف يمكن استخدام عدد من الاختبارات مثل (kappa, press's Q)، سنقتصر على استخدام الاختبار press's Q. لاجراء الاختبار فان الفرضيات:

الفرضية الابتدائية : التصنيف الناتج من خوارزمية الجار الأقرب ليس له دلالة احصائية.

الفرضية البديلة: التصنيف الناتج من خوارزمية الجار الأقرب له دلالة احصائية.

قيمة مؤشر الاختبار: $$pressQ = \frac{[n-(n\,correct \times c)]^2}{n(c-1)}$$

c: عدد مجموعات التصنيف

n:حجم العينة

n correct : عدد المشاهدات المصنفة بشكل صحيح.

$$pressQ = \frac{[292-(240\times 3)]^2}{292(3-1)} = 313.67$$

بمقارنة القيمة السابقة مع القيمة الجدولية و التي تبلغ 6.63 نرفض فرضية العدم و نقر بأن التصنيف الناتج من النموذج دال احصائياً.

3.2 تصنيف القروض المصرفية باستخدام خوارزمية شجرة القرار:

لتصنيف القروض المصرفية باستخدام خوارزمية شجرة القرارات تم ادخال البيانات الى البرنامج orange. الجدول التالي (4) يظهر نتائج تصنيف القروض بحسب خوارزمية شجرة القرار.

الجدول رقم (4): نتائج تصنيف القروض بحسب خوارزمية شجرة القرار

		تقييم المصرف	Predicted group membership			total
			رديء	متعثر	جيد	
original	count	رديء	8	20	1	29
		متعثر	1	82	4	87
		جيد	0	26	150	176
	%	رديء	27.59	68.97	3.44	100.0
		متعثر	1.15	94.25	4.6	100.0
		جيد	0	14.77	85.23	100.0

a.82.19 % of original grouped cases correctly classified.

المصدر: من إعداد الباحثة بالاعتمادعلى مخرجات البرنامج orange

من الجدول السابق نجد أن:

- مجموع حالات التوافق بين التصنيف حسب المصرف والتصنيف الناتج عن خوارزمية شجرة القرار يساوي مجموع عناصر القطر الرئيسي 8+82+150=240 أما مجموع حالات الاختلاف بين التصنيفين يساوي مجموع العناصر الغير قطرية: 26+4+1+1+20=52 و بالتالي فان درجة دقة التنبؤ:

$$p = \frac{(292-52)}{292} = 0.8219$$

- تمكن النموذج من تصنيف 150 من أصل 176 قرض جيد تصنيفاً صحيحاً أي بمعدل دقة 85.23%. في حين تم تصنيف 26 قرض جيد ضمن فئة القروض المتعثرة ، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الجيدة 14.77% لصالح مجموعة القروض المتعثرة.

- أما بالنسبة لمجموعة القروض المتعثرة تم تصنيف 82 من 87 قرض تصنيفاً صحيحاً أي بمعدل 94.25%، ليبلغ خطأ التصنيف في القروض المتعثرة 5.75% لصالح مجموعة القروض الجيدة و الرديئة.
- بلغ التصنيف الصحيح ضمن مجموعة القروض الرديئة 8 قرضاً من أصل 29 بمعدل27.59% و بلغ عدد القروض الرديئة المصنفة ضمن مجموعة القروض المتعثرة 20 قروض، و ضمن مجموعة القروض الجيدة 1 قروض، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الرديئة68.97% و 3.44% لصالح مجموعتي القروض المتعثرة و الجيدة على التوالي (بمعدل عام 72.41%).

لتقييم دقة التصنيف سنستخدم الاختبار press's Q، لاجراء الاختبار فان الفرضيات:

الفرضية الابتدائية : التصنيف الناتج من خوارزمية شجرة القرارات ليس له دلالة احصائية.

الفرضية البديلة: التصنيف الناتج من من خوارزمية شجرة القرارات له دلالة احصائية.

$$pressQ = \frac{[292-(240\times 3)]^2}{292(3-1)} = 313.67$$

بمقارنة القيمة السابقة مع القيمة الجدولية و التي تبلغ 6.63 نرفض فرضية العدم و نقر بأن التصنيف الناتج من النموذج دال احصائياً.

الجدول (5) يظهر معايير التقييم للتصنيف بحسب خوارزمية شجرة القرار:

الجدول (5): معايير المقارنة بحسب خوارزمية شجرة القرارات

MSE	RMSE	MAE	R2
0.138	0.371	0.2	0.693

المصدر: الجدول من إعداد الباحثة بالاعتماد على مخرجات برنامج orange.

3.3 تصنيف القروض المصرفية باستخدام الشبكات العصبونية:

لتصنيف القروض المصرفية باستخدام الشبكات العصبية الاصطناعية تم ادخال البيانات الى البرنامج orange. الجدول التالي (6) يظهر نتائج تصنيف القروض بحسب الشبكات العصبونية.

الجدول رقم (6): نتائج تصنيف القروض بحسب الشبكات العصبية الاصطناعية

		تقييم المصرف	Predicted group membership			total
			رديء	متعثر	جيد	
original	count	رديء	24	4	1	29
		متعثر	1	85	1	87
		جيد	5	3	168	176
	%	رديء	82.76	13.79	3.45	100.0
		متعثر	1.15	97.7	1.15	100.0
		جيد	2.85	1.7	95.45	100.0

a.94.86 % of original grouped cases correctly classified.

المصدر: من إعداد الباحثة بالاعتمادعلى مخرجات البرنامج orange

من الجدول السابق نجد أن:

- مجموع حالات التوافق بين التصنيف حسب المصرف والتصنيف الناتج عن التحليل التمييزي يساوي مجموع عناصر القطر الرئيسي 168+85+24=277 أما مجموع حالات الاختلاف بين التصنيفين يساوي مجموع العناصر الغير قطرية: 5+3+1+1+4=15 و بالتالي فان درجة دقة التنبؤ:

$$p = \frac{(292-15)}{292} = 0.9486$$

- تمكن النموذج من تصنيف 168 من أصل 176 قرض جيد تصنيفاً صحيحاً أي بمعدل دقة 95.45%. في حين تم تصنيف 3 قرض جيد ضمن فئة القروض المتعثرة و 5 قرض جيد ضمن فئة القروض الرديئة، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الجيدة 1.7% و 2.85% لصالح مجموعتي القروض المتعثرة و الرديئة على التوالي (بمعدل عام 4.55%).

- أما بالنسبة لمجموعة القروض المتعثرة تم تصنيف 85 من 87 قرض تصنيفاً صحيحاً أي بمعدل 97.7%، بينما تم تصنيف 1 قرضاً متعثراً ضمن مجموعة القروض الرديئة و 1 قروض متعثرة ضمن القروض الجيدة بمعدل 1.15% و 1.15 % على التوالي (بمعدل عام 2.30%).
- بلغ التصنيف الصحيح ضمن مجموعة القروض الرديئة 24 قرضاً من أصل 29 بمعدل82.76% و بلغ عدد القروض الرديئة المصنفة ضمن مجموعة القروض المتعثرة 4 قروض، و ضمن مجموعة القروض الجيدة 1 قروض (بمعدل عام 17.24%).

لتقييم دقة التصنيف سنستخدم الاختبار press's Q:

الفرضية الابتدائية : التصنيف الناتج من النموذج المقترح ليس له دلالة احصائية.

الفرضية البديلة: التصنيف الناتج من النموذج المقترح ليس له دلالة احصائية.

$$pressQ = \frac{[292-(277\times 3)]^2}{292(3-1)} = 497.47$$

بمقارنة القيمة السابقة مع القيمة الجدولية و التي تبلغ 6.63 نرفض فرضية العدم و نقر بأن التصنيف الناتج من النموذج دال احصائياً.

الجدول (7) يظهر معايير التقييم للتصنيف بحسب خوارزمية الشبكات العصبية الاصطناعية:

الجدول (7): معايير المقارنة بحسب الشبكات العصبية الاصطناعية

MSE	RMSE	MAE	R2
0.02	0.143	0.04	0.954

المصدر: الجدول من إعداد الباحثة بالاعتماد على مخرجات برنامج orange.

3.4 تصنيف القروض المصرفية باستخدام التحليل التمييزي: لتصنيف القروض المصرفية باستخدام التحليل التمييزي تم ادخال البيانات الى برنامج spss. و قد أظهرت النتائج على معايير التقييم في الجدول رقم (8).

الجدول رقم (8): نتائج تصنيف القروض بحسب التحليل التمييزي

		تقييم المصرف	Predicted group membership			total
			رديء	متعثر	جيد	
original	count	رديء	16	6	7	29
		متعثر	23	58	6	87
		جيد	17	13	146	176
	%	رديء	55.2	20.7	24.1	100.0
		متعثر	26.4	66.7	6.9	100.0
		جيد	9.7	7.4	83.0	100.0

a.75.3% of original grouped cases correctly classified.

المصدر: من إعداد الباحثة بالاعتمادعلى مخرجات البرنامج orange

من الجدول السابق نجد أن:

- مجموع حالات التوافق بين التصنيف حسب المصرف والتصنيف الناتج عن التحليل التمييزي يساوي مجموع عناصر القطر الرئيسي 16+58+146=220 أما مجموع حالات الاختلاف بين التصنيفين يساوي مجموع العناصر الغير قطرية: 6+7+23+6+17+13=72 و بالتالي فان درجة دقة التنبؤ:

$$p = \frac{(292-72)}{292} = 0.7534$$

- تمكن النموذج من تصنيف 146 من أصل 176 قرض جيد تصنيفاً صحيحاً أي بمعدل دقة 83%. في حين تم تصنيف 13 قرض جيد ضمن فئة القروض المتعثرة و 17 قرض جيد ضمن فئة القروض الرديئة، لتبلغ نسبة خطا التصنيف ضمن مجموع القروض الجيدة 7.4% و 9.7% لصالح مجموعتي القروض المتعثرة و الرديئة على التوالي (بمعدل عام 17.1%).

- أما بالنسبة لمجموعة القروض المتعثرة تم تصنيف 58 من 87 قرض تصنيفاً صحيحاً أي بمعدل 66.7%، بينما تم تصنيف 23 قرضاً متعثراً ضمن مجموعة القروض الرديئة و 6 قروض متعثرة ضمن القروض الجيدة بمعدل 26.4% و 6ز9 % على التوالي (بمعدل عام 33.3%).
- بلغ التصنيف الصحيح ضمن مجموعة القروض الرديئة 16 قرضاً من أصل 29 بمعدل55.2% و بلغ عدد القروض الرديئة المصنفة ضمن مجموعة القروض المتعثرة 6 قروض، و ضمن مجموعة القروض الجيدة 7 قروض (بمعدل عام 44.8%).

لتقييم دقة التصنيف سنستخدم الاختبار press's Q:

الفرضية الابتدائية : التصنيف الناتج من التحليل التمييزي ليس له دلالة احصائية.

الفرضية البديلة: التصنيف الناتج من التحليل التمييزي له دلالة احصائية.

$$pressQ = \frac{[292-(220\times 3)]^2}{292(3-1)} = 231.89$$

بمقارنة القيمة السابقة مع القيمة الجدولية و التي تبلغ 6.63 نرفض فرضية العدم و نقر بأن التصنيف الناتج من النموذج دال احصائياً.

الجدول (9) يظهر معايير التقييم للتصنيف بحسب التحليل التمييزي:

الجدول (9): معايير المقارنة بحسب التحليل التمييزي

MSE	RMSE	MAE
0.428	0.655	0.298

المصدر: الجدول من إعداد الباحثة بالاعتماد على مخرجات برنامج spss.

3.5 مقارنة بين الأساليب الاحصائية و أساليب الذكاء الاصطناعي المقترحة لتصنيف القروض المصرفية

الجدول (10) يوضح معايير التقييم لكل من الاساليب المقترحة :

الجدول (10) معايير التقييم لكل من الاساليب المقترحة

الأساليب	MAE	RMSE	MSE
الجار الأقرب	0.163	0.355	0.126
شجرة القرارات	0.2	0.371	0.138
الشبكات العصبية الاصطناعية	0.04	0.143	0.02
التحليل التمييزي	0.298	0.655	0.428

المصدر: الجدول من إعداد الباحثة بالاعتماد على الجداول (2،5،7،9).

من الجدول (10) السابق نجد أن معايير التقييم لجميع أساليب الذكاء الاصطناعي أقل من معايير التقييم للتحليل التمييزي. و أن قيمة معايير التقييم بالنسبة للشبكات العصبية الاصطناعية أقل من باقي أساليب الذكاء الاصطناعي المقترحة. الملحق (1) يوضح تصنيف القروض المصرفية للمصرف الزراعي في طرطوس بحسب الشبكات العصبية الاصطناعية ل 76 قرض (لم يتم وضع 292 قرض نظراً لمحدودية عدد الصفحات) .

4. الاستنتاجات:

يعتبر قرار منح القروض من القرارات المفعمة بالخطورة بالنسبة للبنوك المانحة للقرض. حيث أن تعثر القروض وعدم سدادها قد ينتج عنة افلاس المصرف المانح. وهذا مادفع الكثير من المصارف الى اعتماد الأساليب الكمية و الاحصائية للتنبؤ بحالة القرض، وبالتالي اتخاذ القرار المناسب في منح أو عدم منح هذا الائتمان.

تناول البحث تصنيف القروض المصرفية باستخدام بعض أساليب الذكاء الاصطناعي و أحد الأساليب الاحصائية. ومن خلال الدراسة العملية، تمكن البحث من الاجابة على كل من التساؤلات المطروحة حيث تم تحديد دقة التنبؤ في تصنيف القروض المصرفية في المصرف الزراعي في طرطوس باتباع كل من الأساليب المقترحة و قد حصلنا على دقة تنبؤ عالية (أكبر من 70%) باتباع كل من الاساليب المقترحة. و بالمقارنة بين الأساليب المقترحة فقد توصلنا الى أفضلية أساليب الذكاء الاصطناعي وخصوصاً الشبكة العصبية الاصطناعية حيث بلغت دقة التنبؤ بحسب خوارزمية الجار الأقرب وشجرة القرارات (82.19%) بينما بحسب التحليل التمييزي فقد بلغت دقة التنبؤ الى (75.3%)، بحسب الشبكات العصبية الاصطناعية فقد بلغت دقة التنبؤ الى (94.86%) . كما أن المقارنة بين الأساليب المقترحة باستخدام معايير التقييم (MSE, MAE,R2, RMSE,) أعطت نفس النتيجة (أفضلية أساليب الذكاء الاصطناعي في التنبؤ بتصنيف القروض المصرفية). كما تم من خلال اختبار الفرضيات عدم وجود فرق ذو دلالة احصائية بين التصنيف بحسب الطرق المقترحة و تصنيف القروض في المصرف المدروس.

التوصيات

- نوصي العاملين في اتخاذ القرار بمنح او عدم منح القروض المصرفية بالاعتماد على الشبكات العصبونية الاصطناعية التي تعطي دقة وسرعة في اتخاذ القرار.
- نوصي بدراسة أعمق لاستخدام اسلوب التحليل التمييزي وأسلوب الشبكات العصبونية الاصطناعية كأسلوب هجين لتصنيف القروض.

5. المصادر:

1. البساط، هشام. (2001). تحديث وتطوير القطاع المالي والمصرفي في سورية، ورقة مقدمة من صندوق النقد العربي، دمشق، سورية.
2. الحسين، قصي حبيب.(2009). مقدمة في الشبكات العصبية الاصطناعية. جامعة الامام جعفر الصادق علية السلام، لبنان،الصفحة 3-4.
3. الزبيدي،حمزة(2002). أساسيات الادارة المالية. مؤسسة الوراق للنشر و التوبيع، الأردن.
4. العتيبي، فوزي. (2003). استخدام السلاسل الزمنية و الشبكات العصبية في التنبؤ بالأرقام القياسية: دراسة تطبيقية على الأرقام القياسية لأسعار المستهلك بدولة الكويت. ماجستير في الاحصاء التطبيقي، جامعة بور سعيد، مصر. الصفحة 78.
5. الظفيري، عوض. (2004). استخدام الشبكات العصبية الاصطناعية و الأساليب الاحصائية التقليدية للتنبؤ بأعداد الركاب على الخطوط الجوية الكويتية. ماجستير في الاحصاء التطبيقي، جامعة قناة السويس، مصر. الصفحة 54.
6. المبرمج العربي. (2021). تاريخ الاسترداد 13 augast, https://arabicprogrammer.com/article/6268503061،
7. اسماعيل، فريال. (2011). تطبيقات خوارزمية Boosting في التنبؤ لمعالجة البيانات التسلسلية باستخدام الشبكات العصبونية التكرارية. مجلة جامعة تشرين للبحوث و الدراسات العلمية، سلسلة العلوم الأساسية المجلد 33، العدد 1، 2011.الصفحة 159.
8. انجرو،ايمان. الأمين، ماهر. (2006) . التحليل الائتماني ودوره في ترشيد عمليات الإقراض المصرفي بالتطبيق على المصرف الصناعي السوري. سلسلة العلوم الاقتصادية والقانونية،جامعة تشرين، المجلد 28 العدد 3.
9. بن سانية، عبد الرحمان وآخرون. (2017).التنبؤ بالتعثر المالي لبعض المؤسسات العاملة في القطاع الخاص الصناعي بولاية غرداية باستخدام التحليل التمييزي خلال الفترة 2009-2014. مجلة رؤى اقتصادية،العدد 7، ص277-

10.جاسم، سكينة. (2020). دراسة مقارنة بين اسموب التحميل التمييزي الخطي و اسموب التحميل التمييزي المبي. المجلة العراقية للعلوم الادارية.

11.جمعة ،أحمد. (2010). استخدام الشبكات العصبية الإصطناعية في إكتشاف الأخطاء الجوهرية في البيانات المالية: دراسة تطبيقية. المجلة المصرية للدراسات التجارية، جامعة المنصورة، مجلد 36، العدد 1، 2012، الصفحة 139-178. https://www.asjp.cerist.dz/en/article/30780299.

12.صالح، فاتن. (2009). أثر تطبيق الذكاء الاصطناعي والذكاء العاطفي على جودة اتخاذ القرارات. رسالة ماجستير، جامعة الشرق الأوسط للدراسات العليا، الاردن. الصفحة 43.

13.صوار ، يوسف،(2008). محاولة تقدير خطر عدم تسديد القرض باستعمال طريقة القرض التنقيطي والتقنية العصبية الاصطناعية بالبنوك الجزائرية ، دراسة حالة: البنك الجزائري للتنمية الريفية .الجزائر: جامعة ابي بكر بلقايد. تلمسان

14.طاوش قندوسي، دياب يقاي ، صوار يوسف (2012). تقنية الشبكات العصبية الاصطناعية كأحد أساليب ذكاء الأعمال لتسيير مخاطر القروض : دارسة حالة البنك الجزائري. المؤتمر العلمي السنوي الحادي عشر ذكاء الأعمال واقتصاد المعرفة.

15.عبد الحميد, عبد الطيف .البنوك الشاملة عملياتها و إدارتها، الدار الجامعية، الإسكندرية،مصر 2000. ، ص103

16.علي بسام ،محمد الدين خير أنهار،داؤد عبد سهير ،2010، استخدام تنقيب البيانات للتنبؤ بظاهرة تسرب طلبة الكلية بالتطبيق على كلية الإدارة و الاقتصاد، المجلة العراقية للعلوم الاحصائية (18)،209-236.

17.مبروك ،عادل.(2007) . الحلول البديلة لمواجهة ظاهرة الائتمان المصرفي المتعثر.مصر، جامعة القاهرة .

18.نده، كنده عبد الحميد. (2019). استخدام التحليل التمييزي لتصنيف المستوى المعيشي للأسر في محافظة اللاذقية. المجلة الجزائرية للأبحاث الاقتصادية والمالية.

19. A. H. Blasi. Applications of Discriminant Analysis. TECHNICAL REPORT. (2020).
20. D.J.KOK, "Building a quantum KNN classifier with Qiskit: theoretical gains put to practice", Master's thesis, Radboud University Nijmegen, February, 2021.
21. E. Angelini, G.Tollo, A. Roli, "Neural Network Approach for Credit Risk Evaluation، "The Quarterly Review of Economics and Finance،Vol. 48, no. 4, PP733-755, 2008.
22. I. Onur and Y. Tezer, "A Theoretical Approach to Financial Distress Prediction Modeling". Managerial Finance, vol. 43, no. 2, pp 212-230, 2017.
23. J. CHANG. Multi-Step-a Head neural networks for flood forcasting. Hydrological sciences, Hournal, 52 (1), 2007, P121-126.
24. J. HAIR, W. BLACK, B. BABIN, R. ANDERSON, Multivariate Data Analysis, 7th Edition, Upper prentice Hall, 2010, p28.
25. L. Tzong-Huei, "A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models". Neurocomputing, vol.72, no.16, pp 3507-3516, (2009).
26. Learn. (2021). Learn scikit. augast, 2021من ، https://scikitlearn.org/stable/modules/tree.html
27. L. Simar ،W. Härdle. Applied Multivariate Statistical Analysis , Springer, fourth edition, London 2015.
28. Lucidchart. Online Diagram Software & Visual Solution |. تاريخ الاسترداد 13 augast, 2021. https://www.lucidchart.com/pages/decision-tree
29. M. kuliah. Program Studi Matematika. Universitas Islam Negeri Makassar،2018.
30. M .mosses. " the causes and impacts of loan default to microfinance institutions (MFIs) activities: the case of Pride

Tanzania Ltd Pampa Branch-Mwanz".TANZANIA, Mzumbe university،2015،

31. N. Mohammad, A. Nahar Onni, "Credit Risk Grading Model and Loan Performance of Commercial Banks in Bangladesh", European Journal of Business and Management, Vol.7, No.13, 2015.

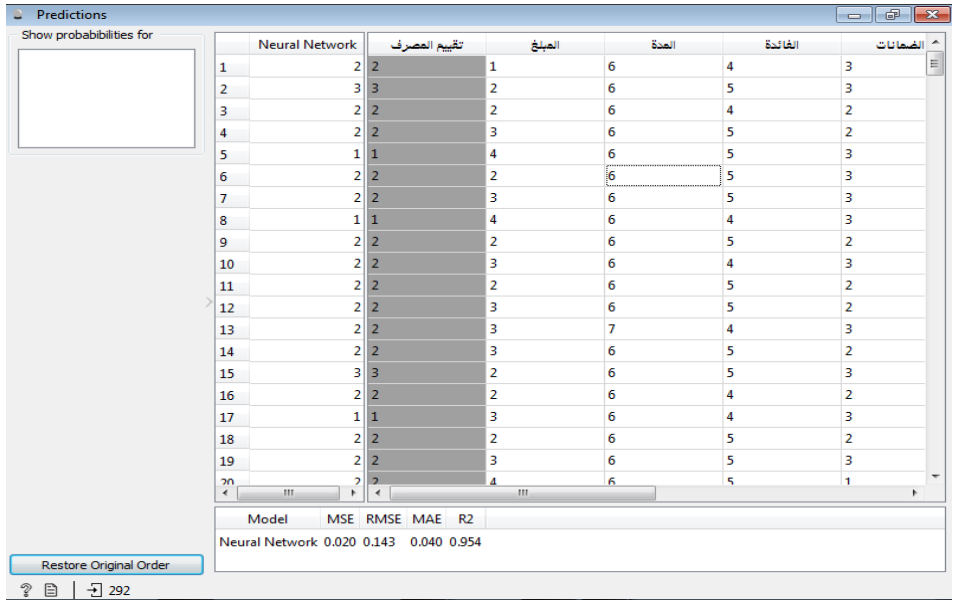
32. R. Kumar and R. Verma, "Classification algorithms for data mining: A survey", International Journal of Innovations in Engineering and Technology (IJIET), vol.1, no.2, pp.7–14, 2012.

33. S. Jain ،M. Kuriakose. Discriminant analysis – simplified. International Journal of Contemporary Dental and Medical Reviews, September , 2019.

34. Y. Kim, B. Jin Kim and Gwan-Su Yi, "Reuse of imputed data in microarray analysis increases imputation efficiency", BMC Bioinformatics,Vol.5, 2004.

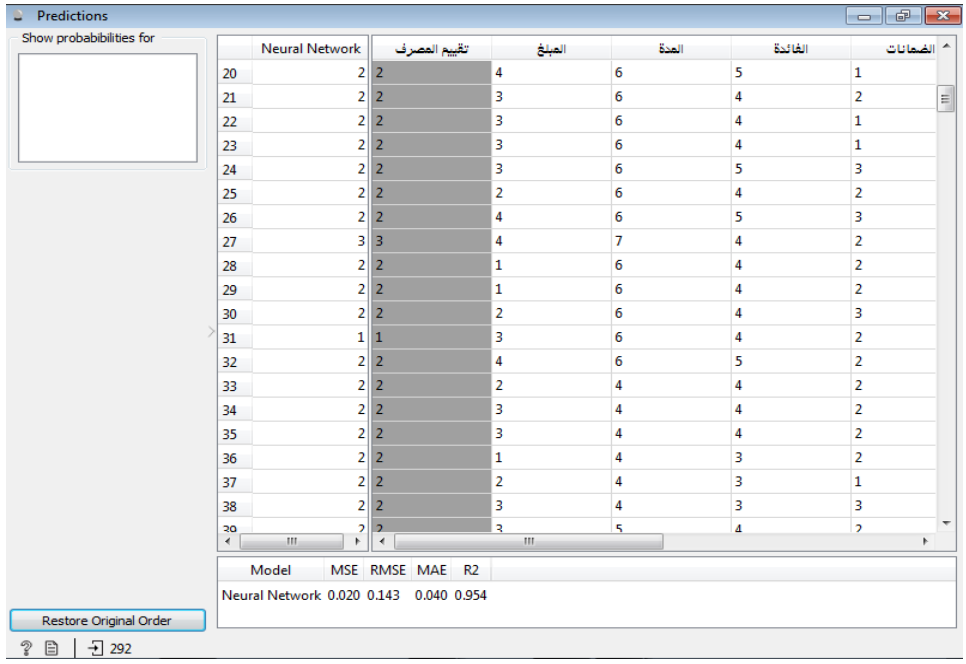
35. http://f.cb.gov.sy/91afb30f686a186aa64db5c288d9a7242437555525412512.pdf

الملحق (1) تصنيف القروض المصرفية في المصرف الزراعي في طرطوس بالاعتماد على الشبكات العصبية الاصطناعية



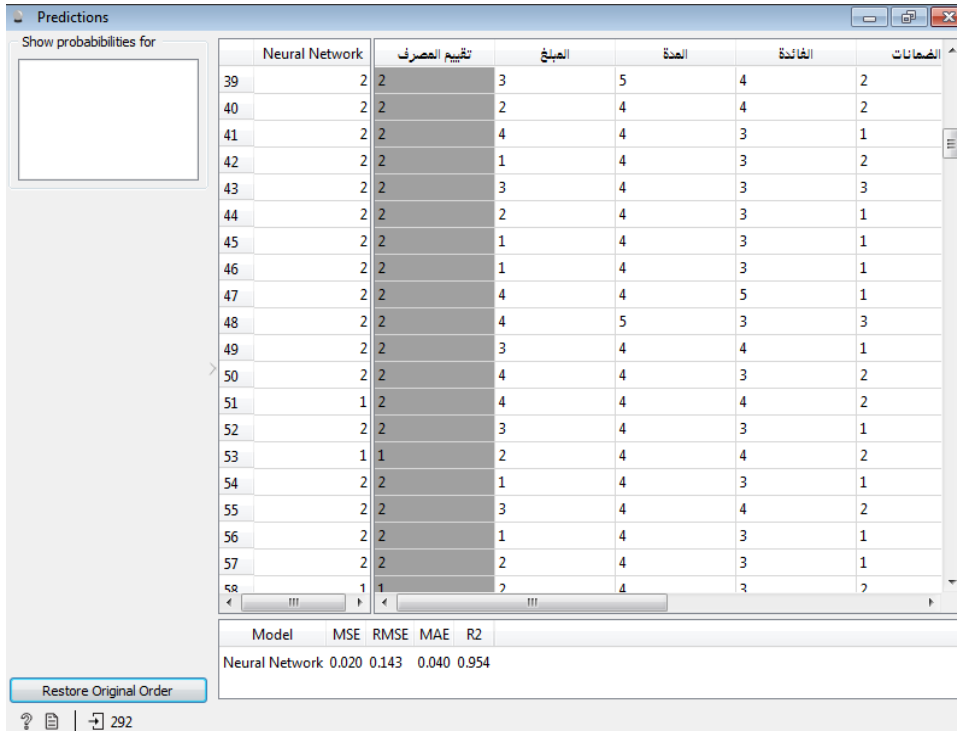
	Neural Network	تقييم المصرف	المبلغ	المدة	الفائدة	الضمانات
1	2	2	1	6	4	3
2	3	3	2	6	5	3
3	2	2	2	6	4	2
4	2	2	3	6	5	2
5	1	1	4	6	5	3
6	2	2	2	6	5	3
7	2	2	3	6	5	3
8	1	1	4	6	4	3
9	2	2	2	6	5	2
10	2	2	3	6	4	3
11	2	2	2	6	5	2
12	2	2	3	6	5	2
13	2	2	3	7	4	3
14	2	2	3	6	5	2
15	3	3	2	6	5	3
16	2	2	2	6	4	2
17	1	1	3	6	4	3
18	2	2	2	6	5	2
19	2	2	3	6	5	3
20	2	2	4	6	5	1

Model	MSE	RMSE	MAE	R2
Neural Network	0.020	0.143	0.040	0.954



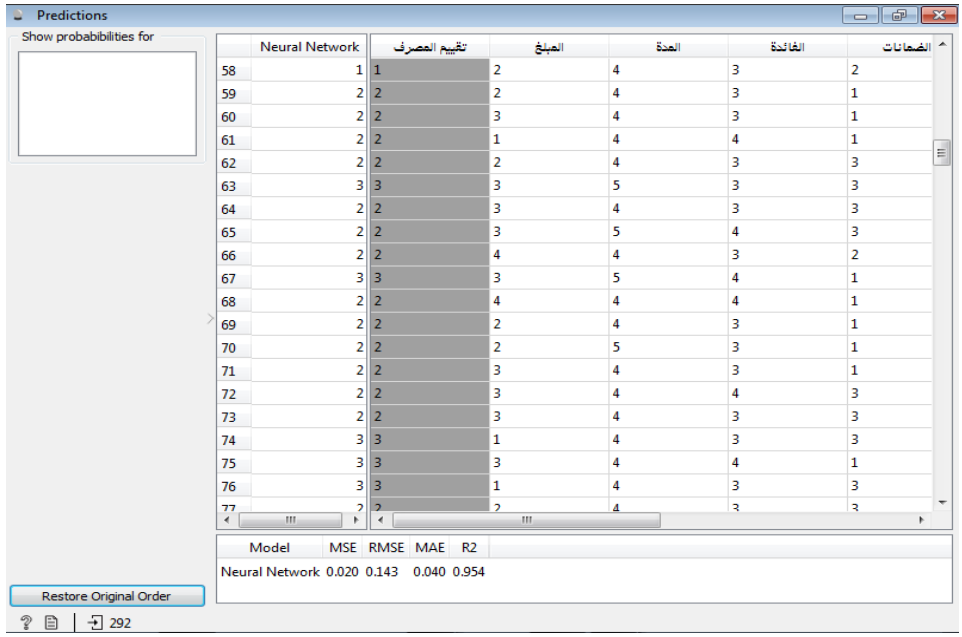
	Neural Network	تقييم المصرف	المبلغ	المدة	الفائدة	الضمانات
20	2	2	4	6	5	1
21	2	2	3	6	4	2
22	2	2	3	6	4	1
23	2	2	3	6	4	1
24	2	2	3	6	5	3
25	2	2	2	6	4	2
26	2	2	4	6	5	3
27	3	3	4	7	4	2
28	2	2	1	6	4	2
29	2	2	1	6	4	2
30	2	2	2	6	4	3
31	1	1	3	6	4	2
32	2	2	4	6	5	2
33	2	2	2	4	4	2
34	2	2	3	4	4	2
35	2	2	3	4	4	2
36	2	2	1	4	3	2
37	2	2	2	4	3	1
38	2	2	3	4	3	3
39	2	2	3	5	4	2

Model	MSE	RMSE	MAE	R2
Neural Network	0.020	0.143	0.040	0.954



	Neural Network	تقييم المصرف	المبلغ	المدة	الفائدة	الضمانات
39	2	2	3	5	4	2
40	2	2	2	4	4	2
41	2	2	4	4	3	1
42	2	2	1	4	3	2
43	2	2	3	4	3	3
44	2	2	2	4	3	1
45	2	2	1	4	3	1
46	2	2	1	4	3	1
47	2	2	4	4	5	1
48	2	2	4	5	3	3
49	2	2	3	4	4	1
50	2	2	4	4	3	2
51	1	2	4	4	4	2
52	2	2	3	4	3	1
53	1	1	2	4	4	2
54	2	2	1	4	3	1
55	2	2	3	4	4	2
56	2	2	1	4	3	1
57	2	2	2	4	3	1

Model	MSE	RMSE	MAE	R2
Neural Network	0.020	0.143	0.040	0.954



	Neural Network	تقييم المصرف	المبلغ	المدة	الفائدة	الضمانات
58	1	1	2	4	3	2
59	2	2	2	4	3	1
60	2	2	3	4	3	1
61	2	2	1	4	4	1
62	2	2	2	4	3	3
63	3	3	3	5	3	3
64	2	2	3	4	3	3
65	2	2	3	5	4	3
66	2	2	4	4	3	2
67	3	3	3	5	4	1
68	2	2	4	4	4	1
69	2	2	2	4	3	1
70	2	2	2	5	3	1
71	2	2	3	4	3	1
72	2	2	3	4	4	3
73	2	2	3	4	3	3
74	3	3	1	4	3	3
75	3	3	3	4	4	1
76	3	3	1	4	3	3

Model	MSE	RMSE	MAE	R2
Neural Network	0.020	0.143	0.040	0.954