أثر استخدام المراجع الخارجي لأساليب التنقيب في البيانات على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية
"دراسة تطبيقية"

ا.د. محمد على وهدان*
د. رجب محمد عمران **
شيماء حمدى مصطفى على ***

(*) **أ.د. محمد على وهدان** : أستاذ المراجعة ووكيل كلية التجارة للدراسات العليا والبحوث - جامعة المنوفية له العديد من الأبحاث المنشورة محلياً وعالمياً، واهتماماته البحثية تتمثل فى المراجعة،نظم المعلومات المحاسبية،المحاسبة الإدارية، الضرائب

Email : wahdan@msn.nl

(**) **د.رجب محمد عمران**: كلية التجارة، جامعة مدينة السادات

Email ragab.omran@com.usc.edu.eg

(***)**شيماء حمدى مصطفى على** : باحثة ماجستير كلية التجارة جامعة المنوفية

Email :shimaabadr718@gmail.com

ملخص البحث:

استهدفت الدراسة قياس أثر استخدام المراجع الخارجي لأساليب التنقيب في البيانات على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية. وذلك من خلال إجراء دراسة تطبيقية على عينة مكونة من (٥٠) شركة من الشركات المقيدة في البورصة المصرية في مؤشر EGX100، بإجمالي مشاهدات (٢٣٥) مشاهدة على مدار الفترة من ٢٠١٨م حتى ٢٠٢٢م. وتناولت الدراسة كل من خوارزمية الشبكة العصبية الإصطناعية، وخوارزمية الغابة العشوائية، وخوارزمية الجار الأقرب كأحد أساليب التنقيب في البيانات، وخوارزمية الإنحدار اللوجستي كأحد الأساليب التقليدية، وذلك لقياس دقة التنبؤ بالتعثر المالي. وتم تقييم هذه الخوارزميات باستخدام ثلاثة معايير لتقييم الأداء وهم (معيار الدقة Accuracy، ومعيار Kappa، ومعيار F1.Score). وتم الاعتماد على (٢٣) متغير مستقل للتنبؤ بالتعثر المالي، ثم تم استخلاص أهم (٩) مؤشرات من بينهم تساعد المراجع الخارجي بدرجة كبيرة في التنبؤ بدقة بالتعثر المالي للشركات. بينما تم قياس التعثر المالي للشركات من خلال تحليل محتوى تقرير المراجع الخارجي للشركات محل الدراسة، حيث تم إعطاء القيمة (١) في حالة وجود تعثر مالي (عدم وجود استقرار مالي)، والقيمة (صفر) في حالة عدم وجود تعثر مالي (وجود استقرار مالي).

وتوصلت الدراسة إلى أن أساليب التنقيب في البيانات (خوارزمية الشبكة العصبية الإصطناعية، وخوارزمية الغابة العشوائية، وخوارزمية الجار الأقرب) كانت جميعها أكثر دقة في عملية التنبؤ بالتعثر المالي للشركات مقارنة بالأساليب التقليدية (خوارزمية الإنحدار اللوجستي). وأن أكثر الأساليب دقة في عملية التنبؤ كانت خوارزمية الشبكة العصبية الإصطناعية، يليها في المرتبة الثانية خوارزمية الغابة العشوائية، والمرتبة الثالثة خوارزمية الجار الأقرب، وأخيراً خوارزمية الإنحدار اللوجستي.

الكلمات الافتتاحية:

أساليب التنقيب في البيانات – التعثر المالي – خوارزمية الشبكة العصبية الإصطناعية – خوارزمية الغابة العشوائية – خوارزمية الجار الأقرب – خوارزمية الإنحدار اللوجستي.

Abstract:

The study aimed to measure the effect of the external auditor's use of data mining techniques on the accuracy of predicting financial distress for companies listed on the Egyptian Stock Exchange compared to traditional methods. This is done by conducting an empirical study on a sample of (50) companies listed on the Egyptian Stock Exchange in the EGX100 index, with 235 views over the period from 2018 to 2022. The study examined the artificial neural network algorithm, the random forest algorithm, and the k-nearest neighbor algorithm as one of the data mining techniques, and the logistic regression algorithm as one of the traditional methods, in order to measure the accuracy of predicting financial distress. These algorithms were evaluated using three performance evaluation criteria: (Accuracy criterion, Kappa criterion, and F1.Score criterion). (23) Independent variables were relied upon to predict financial distress, and then the most important (9) indicators were extracted from among them, which greatly assist the external auditor in accurately predicting companies' financial distress. While the financial distress of companies was measured by the content analysis of the external auditor's report for the companies under study, the value was given (1) in the event of financial distress (lack of financial stability), and the value (zero) in the absence of financial distress (presence of financial stability).

The study found that data mining techniques (artificial neural network algorithm, random forest algorithm, and k-nearest neighbor algorithm) were all more accurate in the process of predicting companies' financial distress compared to traditional techniques (logistic regression algorithm). The most accurate technique in the prediction process was the artificial neural network algorithm, followed in second place by the random forest algorithm, third place by the k-nearest neighbor algorithm, and finally the logistic regression algorithm.

Key words:

Data mining techniques – financial distress – artificial neural network algorithm – random forest algorithm – k-nearest neighbor algorithm – logistic regression algorithm.

القسم الأول

الإطار العام للبحث

١- مقدمة البحث:

تزايدت أهمية التنبؤ بالتعثر المالي والإفلاس بعد الأزمة المالية العالمية في عام ٢٠٠٨م، والتي أدت إلى زيادة حالات الإفلاس بدول مختلفة (الجبالي، ٢٠٢٠). لذا أصبح موضوع التعثر المالي من الموضوعات التي نالت اهتماماً كبيراً من الباحثين، بسبب الآثار السلبية التي يمكن أن تنعكس على الاقتصاد الوطني وعلى العديد من الشركات العاملة فيه، كما تهدد عملية التعثر المالي الشركات وتهدد استمرارها من خلال تأثيرها على رأسمالها وسيولتها، الأمر الذي ينعكس سلباً على ربحيتها (الشيخ، ٢٠٢٠).

الأمر الذي دفع الكثير من المهتمين بمهنة المراجعة في البحث عن وسائل وأساليب حديثة لدعم الرأي المهني لمراجعي الحسابات ومساعدتهم في التأكد من قدرة المنشأة على الاستمرارية في المستقبل (الجبلي، ٢٠٢٠). ونظراً لأهمية فرض الاستمرارية، فقد خصص المعيار المصري رقم (٥٧٠) الاستمرارية، طبقاً للمعايير المحاسبية التي تساعد المراجع للوقوف على تشخيص قدرة المنشأة على الاستمرار من خلال مراجعته للقوائم المالية (العطار، ٢٠٢٠).

ومن هنا ظهرت حاجة المراجع الخارجي إلى استخدام مجموعة من الأساليب الإحصائية للتنبؤ بالتعثر المالي للشركات، وقد اعتمدت العديد من الدراسات على النماذج والأساليب الإحصائية التقليدية مثل تحليل التمايز الخطي، وتحليل التمايز المتعدد المتغيرات، وتحليل الإنحدار الخطي، وتحليل الإنحدار المتعدد، ونموذج لوجيت، إلا أن هذه الأساليب قد لا تساعد على التنبؤ الدقيق بالتعثر المالي، لأنها لا تستطيع التعامل إلا مع عدد محدود من المتغيرات، وبالتالي فإنها تقوم باستبعاد بعض البيانات والمعلومات المرتبطة بالظاهرة محل الدراسة (خطاب، ٢٠١٩).

ولذلك تزايد الحاجة إلى أساليب حديثة يمكن أن تساعد المراجع الخارجي في التنبؤ بالتعثر المالي للشركات، وبالتالي التقرير عن مدى قدرة المنشأة على الاستمرارية في المستقبل، ومن هذه الأساليب الحديثة أساليب التنقيب في البيانات، والتي تعد أحد الأدوات الهامة والناجحة في تحليل محتوى البيانات في الماضي والمستقبل، حيث تعتبر امتداداً للأساليب التقليدية، ولكن تتميز عنها بقدرتها على تحليل كميات هائلة من البيانات بالإضافة إلى قدرتها على التنبؤ بالمستقبل (علي، ٢٠٢٣).

حيث ظهر التنقيب في البيانات Data Mining في أواخر الثمانينات، كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات، وذلك بتحويلها من مجرد بيانات متراكمة وغير مفهومة إلى معلومات قيّمة يمكن استغلالها والاستفادة منها، والتي تقوم على (التصنيف، التجميع، التنبؤ، والتصوير) من خلال استخدام بعض الأساليب كالشبكات العصبية الاصطناعية، ونموذج الانحدار اللوجستي، وشجرة القرارات، والخوارزميات الجينية (النقودي، ٢٠٢٢).

وانطلاقاً من أهمية استخدام أساليب التنقيب في البيانات في المجالات المحاسبية المختلفة وباعتبارها أحد الأدوات الحديثة في مجال المحاسبة والمراجعة، تسعى هذه الدراسة إلى تحليل وتحديد أثر استخدام المراجع الخارجي لأسلوب التنقيب في البيانات على عملية التنبؤ بالتعثر المالي للشركات.

٢- مشكلة البحث:

تحظى مشكلة التنبؤ بالتعثر المالي للشركات واحتمال تعرضها للإفلاس اهتمام كبير لدى فئات مختلفة تشمل المحللين الماليين، والمحاسبين، والمراجعين، والمستثمرين، والموظفين، والمديرين، والدائنين. وتبرز أهمية هذه المشكلة في أن عدم حصول هؤلاء المستخدمين على تحذيرات مبكرة نسبياً حول احتمال التعثر المالي أو الإفلاس يمكن أن يؤدى إلى اتخاذهم قرارت غير سليمة ومن ثم تحملهم لتكلفة مرتفعة نتيجة القرارات الخاطئة، كما أن الافتقار لتلك التحذيرات قد يترتب عليه ضياع الفرص المحتملة لتفادى التعثر أو الإفلاس المالي (Elewa, 2022).

ويرجع اهتمام أصحاب المصالح بوضع نظام إنذار مبكر للتنبؤ بالتعثر المالي، لكي يمكّنهم من فهم وضع ربحية الشركات وتعديل استراتيجياتها الاستثمارية للحد من الاستثمارات المتوقعة ذات الصلة بالخسائر التي حدثت، ومع ذلك فإن التطور السريع للأسواق وتكامل الاقتصاد العالمي، أدى إلى زيادة عدد الشركات التي تعاني من حدوث التعثر المالي على مر السنوات (خطاب وعلي، ٢٠١٩).

وتزداد أهمية المراجعة في وقت الأزمات المالية بسبب زيادة حاجة مستخدمي القوائم المالية إلى وجود إنذار مبكر للتعثر المالي، مما يمكنهم من اتخاذ القرار الصحيح في الوقت المناسب، وفي الغالب يوجه مستخدمي القوائم المالية اللوم لمراجعي الحسابات في حالة عدم قيامهم بهذا الإنذار خاصة عندما تتعرض الشركات لمشكلات مالية حادة قد تجعلها غير قادرة على الاستمرار في المستقبل، ومن هذا المنطلق أصبح مراقب الحسابات في مصر مسؤولاً عن إبداء رأيه بشأن قدرة

الشركة على الاستمرارية، وتتبع هذه المسئولية من معيار المراجعة المصري رقم (٥٧٠) بشأن الاستمرارية (مشابط، ٢٠٢٠).

إن عملية المراجعة بأساليبها التقليدية لا يمكنها من التعرف بشكل دقيق على مدى قدرة الشركة على الاستمرار من عدمه، ويرجع ذلك إلى التشغيل الالكتروني للبيانات وكبر حجم المنشآت وكثرة البيانات، الناتجة عن التطور التكنولوجي ونمو منشآت الأعمال، مما دفع المراجع إلى التحول من أسلوب المراجعة الشاملة إلى المراجعة بالعينات، وهو الأمر الذي يظهر مدى الحاجة إلى إيجاد أساليب حديثة تساعد المراجع في أداء عملية المراجعة بالفعالية والكفاءة المناسبتين، والتي تدعم المراجع في مواجهة الحجم الكبير من البيانات ذات التشغيل الالكتروني (الشورى والغندور، ٢٠٢٠).

وعلى ذلك تتجسد مشكلة الدراسة في قصور الأساليب التقليدية الحالية المستخدمة في التنبؤ بالتعثر المالي للشركات، وتزايد الحاجة إلى الاعتماد على أساليب حديثة يمكنها التعامل مع الحجم الكبير من البيانات، حتى يمكن للمراجع الخارجي التنبؤ بدقة بالتعثر المالي للشركات محل المراجعة. ومن ثم يمكن صياغة المشكلة في ضوء التساؤل الرئيسي التالي:

" هل يؤدي استخدام المراجع الخارجي لأساليب التنقيب في البيانات (الشبكات العصبية الاصطناعية – الغابات العشوائية – الجار الأقرب) إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي)؟"

وينبثق من هذا التساؤل مجموعة من التساؤلات الفرعية التالية:

١) هل يؤدي استخدام المراجع الخارجي لأسلوب الشبكات العصبية الإصطناعية إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي)؟

٢) هل يؤدي استخدام المراجع الخارجي لأسلوب الغابات العشوائية إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي)؟

٣) هل يؤدي استخدام المراجع الخارجي لأسلوب الجار الأقرب إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي)؟

٣- تحليل الدراسات السابقة:

في ضوء استقراء وتحليل الدراسات السابقة التي ترتبط بمتغيرات الدراسة، يمكن تناول أهم الدراسات التي توضح مشكلة الدراسة، من خلال المحاور الرئيسية التالية:

- دراسات تناولت التنبؤ بالتعثر المالي.
- دراسات تناولت أساليب التنقيب في البيانات.
- دراسات تناولت العلاقة بين استخدام المراجع الخارجي لأساليب التنقيب في البيانات والتنبؤ بالتعثر المالي.

١/٣- دراسات تناولت التنبؤ بالتعثر المالي

استهدفت دراسة (معلم وطيار، ٢٠١٩) التعرف على قدرة نموذج ألتمان على التنبؤ بالتعثر المالي لشركات قطاع التأمين في الجزائر، بالتطبيق على عينة مكونة من (١٢) شركة تأمين مسجلة على موقع مركز السجل التجاري الوطني الجزائري خلال الفترة من ٢٠١٣م حتى ٢٠١٥م. وذلك من خلال تناولها نموذج ألتمان المُعدّل الخاص بالمؤسسات الصناعية وغير الصناعية، والمكون من أربعة مؤشرات متمثلة في (رأس المال العامل إلى مجموع الأصول، الأرباح المحتجزة إلى مجموع الأصول، الأرباح قبل الفوائد والضرائب إلى مجموع الأصول، القيمة الدفترية لحقوق المساهمين إلى إجمالي الديون). وقد توصلت الدارسة إلى أن نصف الشركات المكونة لعينة الدراسة متعثرة، وأن نموذج ألتمان أظهر قدرة تنبؤية عالية في كشف هذا التعثر المالي.

وفي نفس السياق سعت دراسة (Elewa, 2022) إلى تحديد أثر استخدام نماذج ألتمان على التنبؤ بالتعثر المالي في مصر. وذلك بالتطبيق على عينة مكونة من (٤٤) وحدة من مؤشر (EGX 70) خلال الفترة من ٢٠١٦م حتى ٢٠٢٠م (أي ما يعادل ٢٢٠ مشاهدة). وتمثّلت نماذج ألتمان المستخدمة للتنبؤ بالتعثر المالي في نموذج (Altman Z-Score, 1968) ونموذج (Altman Z-Score,1993) المُعدّل. كما تم تطبيق تحليل الانحدار اللوجستي لفحص تأثير النسب المستخدمة في نماذج ألتمان على التنبؤ بالتعثر المالي للشركات. وتوصلت الدراسة إلى أن استخدام نماذج ألتمان لها تأثير هام على جودة التنبؤ بالتعثر المالي، وأن نموذج (Altman Z-Score,1993) المُعدّل يقدم نتائج أفضل من نموذج (Altman Z-Score, 1968) للتنبؤ بالتعثر المالي للشركات.

بينما تناولت دراسة (Hertina & Dari, 2023) تحديد ما إذا كان هناك فرق بين نماذج التعثر المالي في التنبؤ بالإفلاس، بالإضافة إلى تحديد نماذج التعثر المالي الأكثر دقة في التنبؤ بالإفلاس، وذلك بالتطبيق على عينة مكونة من (١٦) شركة من شركات البناء والتشييد المدرجة في بورصة إندونيسيا خلال الفترة من ٢٠٢٠م حتى ٢٠٢١م. واستخدمت الدراسة ثلاثة من نماذج التعثر المالي وهم Springate و Taffler و Zmijewski. وتوصلت الدراسة إلى وجود اختلافات جوهرية بين نماذج التعثر المالي في التنبؤ بإفلاس شركات التشييد والبناء المدرجة في بورصة إندونيسيا، وأن النموذج الأكثر دقة في التنبؤ بالإفلاس هو نموذج Taffler بمعدل دقة (١٠٠ %)، يليه نموذج Zmijewski بمعدل دقة (٨١.٢٥ %)، وأخيراً نموذج Springate بمعدل دقة تبلغ (١٢.٥٠ %).

في حين هدفت دراسة (Kušter, 2023) إلى تحليل دقة نماذج مختارة للتنبؤ بالإفلاس، وذلك باستخدام القوائم المالية لـ (٢٣) شركة أفلست خلال الفترة من ٢٠١٣م حتى ٢٠١٩م. واستخدمت الدراسة ثلاث مجموعات مختلفة من النماذج. المجموعة الأولى تضم نماذج كلاسيكية تم تطويرها باستخدام التحليل التمييزي وهم (Altman, modified Altman, Springate, Taffler and Tishaw, Grover)، والمجموعة الثانية تضم نماذج تم تطويرها باستخدام الانحدار اللوجستي وهم (Ohlson, Zmijewski, Grigaravičius)، والمجموعة الثالثة تعتمد على الذكاء الاصطناعي وتم استخدام (شجرة القرارات Decision tree). وتوصلت الدراسة إلى أن نماذج الانحدار اللوجستي، مثل (Zmijewski و Ohlson) أظهرت أفضل النتائج في مجموعة نماذج التنبؤ الكلاسيكية. وأن شجرة القرار كانت النموذج الأكثر دقة حيث بلغ معدل دقتها التنبؤية ١٠٠% لأنها تنبأت بإفلاس جميع الشركات التي تم تحليلها قبل عام واحد من الإفلاس الفعلي.

٢/٣- دراسات تناولت أساليب التنقيب في البيانات:

تناولت دراسة (Mohammadi et al., 2020) أداء خمسة نماذج إحصائية وتعلم آلي شائعة في اكتشاف الاحتيال في التقارير المالية، حيث تمثلت هذه النماذج في كل من: (شبكة بايز Bayesian Network، التحليل التمييزي Discriminant Analysis، الإنحدار اللوجستي Logistic Regression، الشبكات العصبية الإصطناعية artificial neural network، آلة ناقل الدعم Support Vector Machine). وتناولت الدراسة (١٩) مؤشر كمتغير مستقل بشكل مبدئي، وتم تصفيتها إلى (٩) مؤشرات فقط تمثلت في (إنتاجية الموظف، والحسابات المدينة إلى

صافي المبيعات، والديون إلى حقوق الملكية، والمخزون إلى صافي المبيعات، والمبيعات إلى إجمالي الأصول، والعائد على حقوق الملكية، والعائد على المبيعات، والالتزامات إلى مصاريف الفوائد، والأصول إلى الالتزامات)، وذلك لاستخدام هذه المؤشرات بواسطة خوارزميات التصنيف للكشف عن وجود احتيال مالي أو عدم وجود احتيال مالي (المتغير التابع). وتوصلت الدراسة إلى أن الشبكة العصبية الإصطناعية تؤدي أداءً جيدًا مقارنةً بالشبكة البايزية والتحليل التمييزي والإنحدار اللوجستي وآلة ناقل الدعم.

كما تناولت دراسة (عثمان وإسماعيل، ٢٠٢٢) اختبار قدرة نماذج الشبكات العصبية متعددة الطبقات في تحسين دقة التنبؤ بمخاطر الائتمان التي تواجه منشآت الأعمال مقارنة بالنماذج الإحصائية التقليدية (الإنحدار اللوجستي)، وذلك بالتطبيق على عينة من منشآت الأعمال المسجلة بسوق الأوراق المالية المصرية وعددها (٨٣) منشأة خلال الفترة من ٢٠١٠م حتى عام ٢٠١٩م بإجمالي (٨٢٤) مشاهدة. وتضمنت الدراسة (٣٠) مؤشر مالي بصورة مبدئية كمتغيرات مستقلة، والتنبؤ بمخاطر الائتمان كمتغير تابع وتم تصنيفه إلى (منشآت معرضة لمخاطر الائتمان، ومنشآت غير معرضة لمخاطر الائتمان). وتوصلت الدراسة إلى العديد من النتائج أهمها الحصول على (١٠) مؤشرات كانت الأكثر تأثيراً على دقة التنبؤ بمخاطر الائتمان (معدل العائد على إجمالي الأصول، معدل دوران الدائنون، معدل دوران الأصول، نسبة الديون إلى إجمالي الأصول، مضاعف حقوق الملكية، القيمة السوقية للسهم إلى القيمة الدفترية، نسبة صافي رأس المال العامل، نسبة صافي الربح، معدل العائد على الاستثمار، نسبة إجمالي الديون إلى إجمالي حقوق الملكية)، كما توصلت الدراسة إلى أن نماذج الشبكات العصبية متعددة الطبقات أكثر دقة تنبؤية مقارنة بالإنحدار اللوجستي، حيث بلغ معدل الدقة في الشبكات العصبية (٨٦.٦٧ %)، بينما كان معدل الدقة في الإنحدار اللوجستي (٨٥.١٩ %).

وفي نفس السياق قامت دراسة (النقودي، ٢٠٢٢) بتقييم دقة الأساليب المختلفة للتنقيب في البيانات باعتبارها أحد التقنيات الحديثة في مجال الذكاء الاصطناعي في التنبؤ بمنح القروض للحد من المخاطر الائتمانية وانعكاس ذلك على جودة القوائم المالية، وذلك بالتطبيق على عينة مكونة من (٦١٤) مفردة من المتقدمين للحصول على قرض من البنك. وتناولت الدراسة أساليب التنقيب في البيانات المتمثلة في كل من (شجرة القرارات، الغابات العشوائية، الإنحدار اللوجستي) لتقييم دقة أساليب التنقيب في البيانات في الحد من المخاطر الائتمانية. وتوصلت الدراسة إلى أن أساليب

التنقيب في البيانات المختلفة تعمل على تحسين دقة التنبؤ وسرعة منح القروض المصرفية، وفقاً لتصنيفاتها القائمة على التصنيف والتجميع والتنبؤ والتصور، كما توصلت الدراسة إلى تباين مستوى دقة التنبؤ بين أساليب التنقيب في البيانات، حيث أن نموذج الغابات العشوائية كان أكثر دقة في التنبؤ بمعدل دقة (٩٤ %)، يليه أسلوب شجرة القرارات بمعدل دقة (٨٢ %)، ويأتي في المرتبة الثالثة أسلوب الإنحدار اللوجستي (٦٧ %).

أيضاً هدفت دراسة (Pandimurugan et al., 2022) إلى تحليل استخدام خوارزمية الغابات العشوائية باعتبارها أحد نماذج التعلم الآلي كأداة للتنبؤ بالموافقة على منح القروض للعملاء من عدمه، وبالتالي تقليل المخاطر التي ينطوي عليها الأمر بشكل فعّال في المؤسسات المالية أو البنوك. وتوصلت الدراسة إلى أن خوارزمية الغابات العشوائية كان لها أداء ممتاز في التنبؤ بمنح القروض بدرجة أفضل من الإنحدار اللوجستي.

وتناولت دراسة (Ledhem, 2022) تطبيق تقنيات التنقيب عن البيانات المختلفة للتنبؤ بالأداء المالي للخدمات المصرفية الإسلامية في إندونيسيا في الفترة من ٢٠١١م حتى ٢٠٢٠م، حيث تناولت الدراسة كل من (معدل كفاية رأس المال، وجودة الأصول، وإدارة السيولة) كمتغيرات مستقلة، والعائد على الأصول كمتغير تابع. كما اعتمدت الدراسة على مجموعة من تقنيات التنقيب في البيانات تمثلت في (Lasso Regression، الغابات العشوائية Random Forest، الشبكات العصبية الإصطناعية Artificial Neural Networks، الجار الأقرب K–nearest neighbor). كما استخدمت مجموعة من المعايير لتقييم أداء التقنيات المستخدمة وهم (متوسط الخطأ التربيعي MSE، وجذر متوسط الخطأ التربيعي RMSE، ومتوسط الخطأ المُطلق MAE). وتوصلت الدراسة إلى أن كلاً من تقنية الجار الأقرب (KNN) والغابات العشوائية (RF) هما أفضل التقنيات للتنبؤ بالأداء المالي للخدمات المصرفية الإسلامية في إندونيسيا، يليهم الشبكات العصبية الإصطناعية (ANN)، وجاء في المرتبة الأخيرة من حيث دقة التنبؤ (LR).

بينما استهدفت دراسة (الجبيلي، ٢٠٢٣) إجراء مقارنة بين عدد من الخوارزميات (شجرة القرار، الجار الأقرب، الشبكات العصبية، التحليل التمييزي) للوصول إلى الخوارزمية الأكثر دقة في تصنيف القروض المصرفية، وتم المقارنة بين هذه الخوارزميات باستخدام معايير لتقييم الأداء (MSE، RMSE، R^2، MAE) للحكم على الخوارزمية الأفضل، وتمثلت متغيرات الدراسة في جودة القرض المقدم (كمتغير تابع)، وتم صنيفه إلى (متعثر، رديء، جيد)، ومجموعة من المتغيرات

المستقلة تضمنت (مبلغ القرض، السمعة الاقراضية، مدة القرض، طريقة السداد، سعر الفائدة المحسوبة على القرض الممنوح، الغاية من القرض، الضمانات المقدمة، دخل العميل طالب القرض). وتوصلت الدراسة إلى أن أساليب الذكاء الإصطناعي كانت أكثر دقة في التنبؤ من الأساليب الإحصائية، حيث بلغت دقة التنبؤ للشبكات العصبية الإصطناعية (٩٤.٨ %)، يليها أسلوبي الجار الأقرب وشجرة القرارات بمعدل دقة (٨٢.٢ %)، بينما بلغ معدل الدقة لأسلوب التحليل التمييزي (٧٥.٣ %).

٣/٣- دراسات تناولت العلاقة بين استخدام المراجع الخارجي لأساليب التنقيب في البيانات والتنبؤ بالتعثر المالي:

هدفت دراسة (Yang & Nazareth, 2022) إلى قياس فعالية استخدام المراجع الخارجي لخوارزميات التعلم الآلي عند التقرير عن استمرارية المنشأة. واستخدمت الدراسة (٣٥) مؤشراً كمتغيرات مستقلة للتنبؤ بالاستمرارية كمتغير تابع، وتم اختيار (٨) مؤشرات فقط أكثر تأثيراً في التنبؤ بالاستمرارية، وكانت هذه المؤشرات كالتالي: (العائد على الأصول، رأي مراقب الحسابات الداخلي، الربح قبل الضريبة إلى إجمالي الأصول، نسبة المخزون إلى الأصول المتداولة، الالتزامات المتداولة إلى إجمالي الالتزامات، التدفق النقدي التشغيلي إلى الالتزامات المتداولة، رأي مراجع الحسابات في التقارير المالية، نسبة التداول). واعتمدت الدراسة على مجموعة من الخوارزميات تضمنت كل من: (نموذج Adaptive Boost، شجرة القرار، نموذج Gradient boost، الجار الأقرب، تحليل التمايز، الإنحدار اللوجستي، نموذج Naive Bayes، الغابات العشوائية)، كما استخدمت معيارين لتقييم أداء هذه الخوارزميات، وهما: (معيار الدقة Accuracy، ومعيار F2-measure). وتوصلت الدراسة إلى أن الخوارزميات الجينية حققت معدل دقة (٩٨.٨٨ %)، يليها الغابات العشوائية بمعدل دقة (٩٨.٧٧ %)، ثم Gradient boost بمعدل دقة (٩٨.٧٦ %)، ثم الجار الأقرب بمعدل دقة (٩٨.٦١ %)، ثم Adaptive Boost بمعدل دقة (٩٨.٥٨ %)، ثم الإنحدار اللوجستي بمعدل دقة (٩٨.٤١ %)، يليها تحليل التمايز بمعدل دقة (٩٨.٣٣ %)، ثم شجرة القرار بمعدل دقة (٩٨.٠٧ %)، وأخيراً نموذج Naive Bayes بمعدل دقة (٨٢.٢٦ %).

وفي نفس السياق هدفت دراسة (أمين، ٢٠٢٣) اختبار مدى تأثير اعتماد المراجع الخارجي على نموذج هجين من أدوات الذكاء الإصطناعي على تحسين جودة حكمه بشأن الاستمرارية، وذلك بالتطبيق على عينة مكونة من (١٢١) شركة بإجمالي مشاهدات (٤٦١) مشاهدة خلال الفترة من

٢٠١٨م حتى ٢٠٢١م. واعتمدت الدراسة على مجموعة من أدوات الذكاء الإصطناعي تمثلت في (الغابات العشوائية، الشبكات العصبية الإصطناعية، تقنية الدعم الآلي، الجار الأقرب)، والأساليب الإحصائية التقليدية (الإنحدار اللوجستي). كما اعتمدت الدراسة على (٢٤) مؤشر مالي وغير مالي (كمتغيرات مستقلة)، وتم استخلاص (١٣) مؤشر مالي وغير مالي أكثر دلالة في الحكم على الاستمرارية، وذلك باستخدام خوارزمية البحث التطوري. كما تم استخدام مجموعة من معدلات التقييم وفقاً لـ (معامل الدقة، معامل Kappa الإحصائي، قيمة الجذر التربيعي RMSE). وتوصلت الدراسة إلى أن خوارزمية الغابات العشوائية كانت أكثر دقة في التنبؤ بمعدل (٩٣.٤٠ %)، يليها خوارزمية الجار الأقرب بمعدل (٨٨ %)، ثم الشبكات العصبية الإصطناعية بمعدل (٨٦.٨٠ %)، وأخيراً الإنحدار اللوجستي بمعدل (٧٨.٦٨ %)، وبالتالي حققت نماذج أدوات الذكاء الإصطناعي دقة تنبؤية عالية مقارنة بنموذج الإنحدار اللوجستي.

أيضاً هدفت دراسة (Máté et al., 2023) إلى تقييم أداء خوارزميات التعلم الآلي المختلفة في التنبؤ بالتعثر المالي للشركات. وذلك بالاعتماد على مجموعة من المؤشرات المالية الخاصة بالشركات الباكستانية خلال الفترة من عام ٢٠١٦م حتى عام ٢٠٢١م، وتضمنت هذه المؤشرات كل من (العائد على الأصول، والعائد التشغيلي على الأصول، ونسبة تغطية الديون، ومعدل دوران الأصول، وربحية السهم، ونسبة الدين إلى الأصول، والعائد النقدي على الأصول، ونسبة السيولة السريعة). كما تضمنت الدراسة مجموعة من نماذج التعلم الآلي والتي تمثّلت في كل من: (الإنحدار اللوجستي، شجرة القرار، الغابة العشوائية، نموذج Naive Bayes، نموذج Gradient boost، نموذج Adaptive Boost، نموذج Support Vector Machines). وتوصلت الدراسة إلى أن (شجرة القرار، نموذج AdaBoost، نموذج Gradient boost) أكثر دقة حيث بلغ معدل الدقة في كل منهم (١٠٠ %). وأن (نموذج SVM، الإنحدار اللوجستي) تراوحت دقتها بين (٨٩ % و٩٩ %) اعتمادًا على عدد الميزات المستخدمة، حيث أنه مع زيادة عدد الميزات، مالت دقة هذه النماذج إلى الاقتراب من الكمال. بينما جاء أداء نموذج Naive Bayes ضعيفًا، حيث بلغت دقته من (٥٨ % إلى ٧٠ %).

❖ **التعليق على الدراسات السابقة:**

- تناولت الدراسات السابقة مداخل مختلفة في التنبؤ بالتعثر المالي، فهناك دراسات اعتمدت على نماذج تم تطويرها باستخدام التحليل التمييزي ومن هذه النماذج: (Altman, modified

Altman, Springate, Taffler and Tishaw, Grover)، ودراسات أخرى اعتمدت على نماذج تم تطويرها باستخدام الإنحدار اللوجستي مثل نماذج (Ohlson, Zmijewski, Grigaravičius)، ومجموعة ثالثة استخدمت أساليب الذكاء الإصطناعي مثل: (خوارزمية شجرة القرارات، خوارزمية الغابات العشوائية، خوارزمية Naïve Bayes، خوارزمية Support Vectore Machines (SVM)، الشبكات العصبية الإصطناعية ANN، الجار الأقرب KNN). ولكن معظم الدراسات السابقة اعتمدت بدرجة أكبر على نماذج التحليل التمييزي والإنحدار اللوجستي، في حين أن هناك دراسات محدودة اعتمدت على أساليب الذكاء الإصطناعي؛ لذا فإن هذه الدراسة سوف تعتمد على التنبؤ بالتعثر المالي باستخدام أدوات الذكاء الإصطناعي خاصة في ظل التطور التكنولوجي الهائل، وما قد يوفره من دقة وسهولة أكثر في عملية التنبؤ بالتعثر المالي مقارنة بالأساليب التقليدية.

- اختلفت الدراسات السابقة التي تناولت أدوات الذكاء الإصطناعي حول أكثر هذه الأدوات دقة في عملية التنبؤ، فهناك دراسات توصلت إلى أن الشبكات العصبية الإصطناعية كانت أكثر دقة في التنبؤ، وهناك دراسات توصلت إلى أن خوازرمية الغابات العشوائية كانت أكثر دقة، وأخرى توصلت إلى أن خوارزمية الجار الأقرب أكثر دقة في التنبؤ، وغيرها توصلت أن الخوارزميات الجينية أكثر دقة في التنبؤ. ويرجع الاختلاف في دقة التنبؤ بين أدوات الذكاء الإصطناعي إلى المتغيرات التفسيرية المستخدمة في عملية التنبؤ، ودرجة أهميتها النسبية في عملية التنبؤ. وفي هذه الدراسة سوف يتم الاعتماد على أدوات الذكاء الإصطناعي المتمثلة في (الشبكات العصبية الإصطناعية، الغابات العشوائية، الجار الأقرب) على اعتبار أن هذه الأدوات لها قدرة تنبؤية عالية، بالإضافة إلى أنها أكثر الأدوات استخداماً في عملية التنبؤ.
- أن المراجع الخارجي لكي يبدي حكماً دقيقاً بشأن الاستمرارية، فإنه بحاجة إلى استخدام أدوات أكثر دقة للتنبؤ بالتعثر المالي للشركات، وفي ظل التطور التكنولوجي والتحول نحو الرقمنة، فإن هذه الدراسة سوف تسعى إلى دراسة وتحليل مدى استخدام المراجع الخارجي لأدوات التنقيب في البيانات باعتبارها أحد أدوات الذكاء الإصطناعي في زيادة دقة التنبؤ بالتعثر المالي للشركات.

٤- أهداف البحث:

يتمثل الهدف العام للدراسة في "دراسة أثر استخدام المراجع الخارجي لأساليب التنقيب في البيانات (الشبكات العصبية الاصطناعية – الغابات العشوائية – الجار الأقرب) على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الانحدار اللوجستي)". ويمكن تحقيق الهدف العام من خلال تحقيق الأهداف الفرعية التالية:

١) قياس أثر استخدام المراجع الخارجي لأسلوب الشبكات العصبية الاصطناعية على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الانحدار اللوجستي).

٢) قياس أثر استخدام المراجع الخارجي لأسلوب الغابات العشوائية على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الانحدار اللوجستي).

٣) قياس أثر استخدام المراجع الخارجي لأسلوب الجار الأقرب على دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الانحدار اللوجستي).

٥- أهمية البحث:

تستمد هذه الدراسة أهميتها في أنها تتطرق إلى قضية هامة وحديثة في مجال المراجعة بهدف تحديد دور المراجع الخارجي في التنبؤ بالتعثر المالي باستخدام أساليب التنقيب في البيانات في ظل بيئة تتسم بالتطور التقني والتكنولوجي في مجال نظم المعلومات.

١/٥- الأهمية العلمية:

- حداثة مفهوم التنقيب على البيانات حيث يعد محل ارتكاز واهتمام كبير من قبل المراجعين في الآونة الأخيرة، خاصة بعد التوسع الهائل للشركات في استخدام أساليب حديثة لتقنية المعلومات.
- نتيجة لتزايد حاجة الشركات إلى اتخاذ الإجراءات العلاجية للحد من التدهور والخسائر المتوقعة قبل حدوث الأزمات، فإن هذا البحث سوف يساهم في تحسين دقة التنبؤ بالتعثر المالي للشركات من خلال استخدام نموذج يمكن تطبيقه عملياً كمحاولة لمنع حدوث الفشل المالي أو إفلاس الشركات.
- تساعد هذه الدراسة شرائح هامة وكثيرة من المجتمع مثل المستثمرين الحاليين والمرتقبين، المحللين الماليين، الدائنين، الإدارة، وحملة الأسهم وغيرهم من اللذين يفضلون أن تكون لديهم معلومات دقيقة حول سلامة الموقف المالي للشركة.

٢/٥- الأهمية العملية:

- زيادة قدرة المراجع الخارجي في التنبؤ بالتعثر المالي للشركات من خلال التشغيل الموازي لأساليب التنبؤ بالتعثر المالي عن طريق الأسلوبين التقليدي والمتطور باستخدام أساليب التنقيب في البيانات.

- قد تكون نتائج هذه الدراسة محلاً لاهتمام المراجعين الخارجيين في تقييم أداء الشركات المقيدة بسوق الأوراق المالية المصرية، فضلاً عن توجيه الاهتمام لتقرير مراقب الحسابات بشأن استمرارية الشركات خاصة الشركات المتعثرة مالياً.

٦- فروض البحث:

في ضوء مشكلة وتساؤلات وأهداف البحث يمكن صياغة الفرض الرئيسي للبحث على النحو التالي: "**يؤدي استخدام المراجع الخارجي لأساليب التنقيب في البيانات (الشبكات العصبية الاصطناعية – الغابات العشوائية – الجار الأقرب) إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الانحدار اللوجستي)**".

وينبثق من هذا الفرض الرئيسي مجموعة من الفروض الفرعية التالية:

١) يؤدي استخدام المراجع الخارجي لأسلوب الشبكات العصبية الإصطناعية إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي).

٢) يؤدي استخدام المراجع الخارجي لأسلوب الغابات العشوائية إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي).

٣) يؤدي استخدام المراجع الخارجي لأسلوب الجار الأقرب إلى زيادة دقة التنبؤ بالتعثر المالي للشركات المقيدة بالبورصة المصرية مقارنة بالأساليب التقليدية (الإنحدار اللوجستي).

٧- حدود البحث:

تتمثل حدود البحث في كل من:

١/٧- حدود منهجية: سوف تقتصر الدراسة على كل من أسلوب الشبكات العصبية الإصطناعية، والغابات العشوائية، والجار الأقرب كأحد أساليب التنقيب في البيانات، وأسلوب الإنحدار اللوجستي كأحد الأساليب التقليدية، وذلك لقياس دقة التنبؤ بالتعثر المالي.

٢/٧- حدود مكانية: سوف يتم تطبيق هذه الدراسة على الشركات المقيدة في البورصة المصرية عدا المؤسسات المالية نظراً لاختلاف طبيعة نشاطها، وعدا الشركات التي تقوم بنشر قوائمها المالية بعملة أجنبية.

٣/٧- حدود زمنية: سوف يتم تطبيق الدراسة على الشركات المقيدة في البورصة المصرية خلال الفترة من ٢٠١٨م حتى ٢٠٢٢م.

القسم الثاني

الإطار النظري للبحث

١- أساليب التنقيب في البيانات واستخداماتها في مهنة المراجعة:

١/١- مفهوم التنقيب في البيانات:

أوضح (Dawalekar et al., 2019) أن التنقيب في البيانات هو "عملية بحث داخل كميات كبيرة من البيانات للكشف عن العلاقات التي لم يتم كشفها سابقاً بين عناصر البيانات، والمعروفة أيضًا باسم اكتشاف المعرفة في قواعد البيانات Knowledge Discovery Database (KDD)".

كما أشار (الجبلي، ٢٠٢٠) بأن التنقيب في البيانات هو "عملية تحمل في طياتها الكثير من التكنولوجيا الحديثة والتي تستخدم للحصول على معلومات ذات فائدة ودلالة على درجة كبيرة من الأهمية من خلال أحجام هائلة من البيانات الضخمة المخزنة في مخازن ومستودعات البيانات، والتي تعمل على إيجاد واستخلاص أنماط جديدة من العلاقات بينها من أجل الوصول إلى نتائج هادفة تخدم القرارات المستقبلية لمتخذي القرار".

أيضاً ينظر (Papik & Papikova, 2022) إلى تقنية التنقيب في البيانات على أنها عملية بحث داخل كميات هائلة من البيانات، حيث يتم فرزها لإيجاد معلومات وإيجاد علاقات بينها غير معروفة مسبقاً ثم بناء نماذج للتنبؤ بالسلوك ومن ثم الوصول لمعلومات مفيدة لتحسين عملية صنع القرار.

ويرى (Awad & Wathik, 2022) أن التنقيب في البيانات يشير إلى استخلاص أو استخراج المعرفة من كميات كبيرة من البيانات، باستخدام تقنيات تحليل البيانات الآلية لاكتشاف العلاقات التي لم يتم اكتشافها مسبقًا بين عناصر البيانات، ويمكن النظر إلى التنقيب في البيانات كنتيجة للتطور الطبيعي لتكنولوجيا المعلومات.

ويرى (علي، ٢٠٢٣) أن تقنية التنقيب في البيانات هي "العملية التي يتم فيها استخدام الذكاء الاصطناعي والرياضي والاحصائي بالإضافة إلى تقنيات التعلم الآلي لاستخراج المعلومات المفيدة واكتساب المعرفة من قواعد البيانات الكبيرة".

ومن خلال ما سبق يمكن القول أن التنقيب في البيانات هو عملية تجمع بين الإحصاء وتكنولوجيا المعلومات، بغرض اكتشاف علاقات غير معروفة مسبقًا وذات فائدة من داخل أحجام

هائلة من البيانات الضخمة، وذلك من خلال استخدام طرق مختلفة مثل التلخيص والتصنيف والتحليل والتنبؤ والعنقدة وقواعد الارتباط، بما يفيد في عملية اتخاذ القرار.

٢/١- أساليب التنقيب في البيانات:

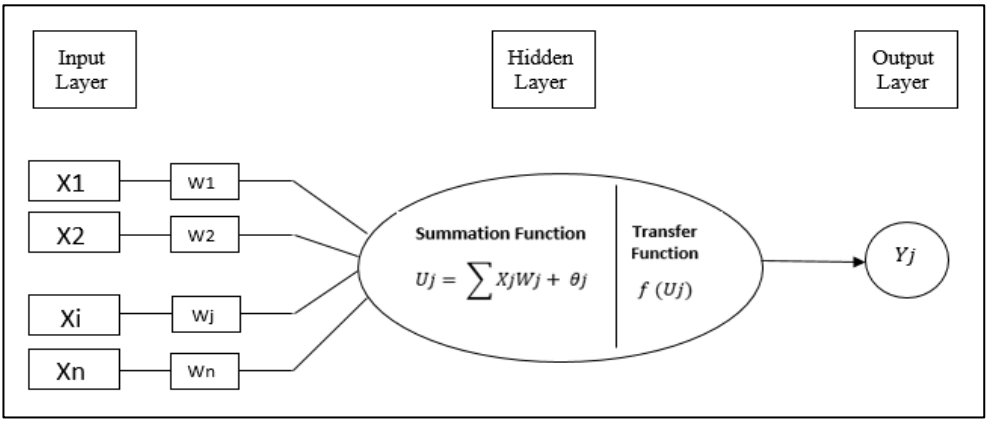
١/٢/١- الشبكات العصبية الإصطناعية Artificial Neural Networks:

تعد الشبكات العصبية الإصطناعية (ANN) أحد أهم تقنيات الذكاء الاصطناعي التي أثبتت قدرتها على التنبؤ، وحل المشكلات بدقة وسهولة مقارنة بالأساليب الإحصائية التقليدية، حيث يطبق أسلوب الشبكات العصبية الاصطناعية بدون الحاجة إلى افتراضات مسبقة، ودون النظر إلى علاقة المتغيرات مع بعضها البعض، على عكس الطرق التقليدية (Chen, 2019). وتعرف بأنها نماذج أو أنظمة حسابية تحاكي الخلايا العصبية في العقل البشري والروابط بينهما (Singh & Banerjee, 2019).

حيث تمثل الشبكات العصبية الإصطناعية نظاماً متكاملاً يتكون من مجموعة مركبة من عدة عناصر مترابطة معاً يطلق عليها نيورونات (Neurons) وهي تعمل في انسجام تام لحل مشكلة معينة من خلال معالجة البيانات في عناصر معالجة بسيطة، وتمر الإشارات (Signals) بين الأعصاب عبر خطوط ربط تسمى روابط اتصال (Connection Links)، ويرفق بكل خط ربط وزن ترجيحي (Weight) معين، والذي يُضرب مع الإشارات الداخلة إلى نيورون، للحصول على مجموع الإشارات الداخلة الموزونة لكل نيورون، ليحدد إشارة المخرجات الناجمة عنه، ويطبق كل نيورون عملية معالجة من خلال دالة تجميع والتي تقوم بتجميع الإشارات الداخلة للنيورون والمرجحة بالأوزان الترجيحية، ودالة تحويل أو تحفيز معينة (عادة تكون غير خطية) تقوم بإجراء معادلات رياضية على القيم الخارجة من دالة التجميع، حتى يتم إنتاج المخرجات النهائية للشبكة العصبية (هاشم، ٢٠٢٠). ويوضح الشكل رقم (١) التالي مكونات الشبكات العصبية على النحو التالي:

شكل رقم (١)

مكونات الشبكات العصبية



المصدر: (هاشم، ٢٠٢٠؛ عثمان وإسماعيل، ٢٠٢٢؛ الجبيلي، ٢٠٢٣).

مـن خـلال الشـكل السـابق يتضـح أن مكونـات الشـبكة العصـبية تتمثـل فيمـا يـأتي (عثمـان وإسماعيل، ٢٠٢٢):

أ- **طبقة المدخلات (Input Layer)**: وتمثل الطبقة الأولى في الشبكات العصبية، وهي عبارة عن جميع عناصر البيانات التي تأتي من خلية عصبية أخرى أو من البيئة الخارجية. وتقوم هذه الطبقة باسـتقبال بيانـات المتغيـرات المسـتقلة مـن المصـادر المختلفـة والمعبـر عنهـا فـي الشـكل السـابق بالمتغيرات (X1, X2, ... Xi, Xn).

ب- الأوزان النسـبية Weights: تقـوم هـذه الأوزان بتحديـد الأهميـة النسـبية لكـل مـدخل مـن المدخلات، وبالتالي تحدد قوة العلاقة بين عناصر ووحدات المعالجة ومدى فعالية عقد الاتصال، ويمكن تعديل الأوزان النسبية من خلال التدريب والتعلم، ويعبر عنها في الشكل السابق بالأوزان (W1, W2, ... Wi, Wn).

ج- الطبقـة الخفيـة (Hidden Layer): وهـي الطبقـة التـي تلـي طبقـة المـدخلات، وقـد تحتـوي الشبكة العصبية على طبقة خفية واحدة أو أكثر من طبقة خفية، وتقوم هذه الطبقة بتصنيف وتمييز وتحليـل المـدخلات بإعطـاء وزن نسـبي لكـل منهـا، ثـم اسـتخدام الـدوال التحليليـة (دالتـي التجميـع والتحويل) لتعديل هذه الأوزان.

د- دالة التجميع (Summation Functions): يطلق على هذه الدالة دالة التنشيط، والتي تقوم بحساب الأوزان النسبية للمدخلات، وذلك من خلال ضرب كل مُدخل من المدخلات في وزنه النسبي للحصول على المجموع.

هـ- دالة التحويل (Transfer Functions): يطلق عليها أيضا دالة التنشيط، والتي تقوم بإجراء المعادلات الرياضية على القيم الخارجة من دالة التجميع، وتعديل الأوزان النسبية باستمرار طوال فترة تدريب الشبكة. وهذه الدالة تحد من مخرجات النيورون، لذا قد تسمى أيضاً بدالة التخميد Squashing حيث تجعل المخرجات ضمن المجال (٠ ، ١) أو (-١ ، ١). وتتمثل أكثر دوال التنشيط المستخدمة في الشبكات العصبية في كل من: (دالة التنشيط الخطية Linear Transfer Function، دالة التنشيط ذات الحد الثابت Hard-limit Transfer Function، دالة التنشيط سيجمويد (الأسية) اللوجستية Log-Sigmoid Transfer Function و دالة التنشيط Tan-sigmoid Function).

و- طبقة المخرجات (Output Layer): هي الطبقة النهائية للشبكات العصبية، وتقوم هذه الطبقة بعرض النتائج التي وصلت إليها من الطبقة السابقة إلى المستخدم النهائي، وبالتالي عرض النتائج النهائية للشبكة العصبية.

٢/٢/١- خوارزمية الغابات العشوائية Random Forest Algorithm:

تُعد الغابة العشوائية أحد خوارزميات التعلم الآلي الخاضع للإشراف التي يمكن استخدامها لحل مشاكل التصنيف والإنحدار، وتم تقديمها لأول مرة بواسطة Breiman في عام ٢٠٠١م، وهي عبارة عن مجموعة من تنبؤات لأشجار القرار، حيث يتم تقسيم البيانات إلى مجموعات جزئية عشوائية متعددة، ثم يتم تدريب شجرة قرار لكل مجموعة جزئية، وبعدها يتم الحصول على نتيجة توقع من كل شجرة قرار، ثم يتم إجراء تصويت لكل نتيجة توقع، حتى يتم تحديد التنبؤ النهائي للنموذج عن طريق اختيار نتيجة التوقع التي حصلت على أكبر عدد من الأصوات. لذلك غالبًا ما توفر الغابة العشوائية دقة أعلى مقارنة بنموذج شجرة القرار. كما تقدم الغابات العشوائية أعلى دقة تنبؤية مقارنة بالنماذج الأخرى في وضع التصنيف (Speiser et al., 2019).

وتحظى الغابات العشوائية بأهمية كبيرة نظراً لدقتها التنبؤية العالية بسبب عدد الأشجار المشاركة في العملية التنبؤية، كما أنها لا تعاني من خطر التجهيز الزائد بسبب أنها تأخذ متوسط جميع التوقعات مما يلغي وجود أي تحيز، هذا بالإضافة إلى قدرتها على معالجة القيم المفقودة، مما

يضمن إمكانية إجراء تنبؤات دقيقة على الرغم من الفجوات الكبيرة في البيانات. وعلى الرغم من هذه المزايا إلا أن الغابة العشوائية تعد بطيئة في توليد التنبؤات لأنها تعتمد على عدد كبير من أشجار القرار في عملية التنبؤ (Máté et al., 2023).

٣/٢/١- خوارزمية الجار الأقرب K-Nearest Neighbor Algorithm:

يعتبر خوارزمية الجار الأقرب من خوارزميات التنقيب والتي ظهرت في بداية الخمسينات من القرن الماضي، وهي مدخل تصنيفي يعتمد على التشابه، وتُعد خوارزمية الجار الأقرب شائعة الاستخدام لتحديد التنظيم المكاني، حيث تقيس المسافة الفاصلة بين كل نقطة وأقرب جار لها، وعلى أساس معدل المسافة الفاصلة بين كل نقطة وأقرب جار لها ومعدل المسافة الفاصلة بين جميع النقاط (معدل التباعد) تتم المقارنة باستخدام توزيعات بواسون العشوائية (أبو الخير، ٢٠١٩).

إن الفكرة الأساسية لخوارزمية الجار الأقرب تكمن في تصنيف الحالات غير المرئية (أو غير المصنفة) إلى الحالات الأقرب لها ضمن حجم معين، حيث تفترض أن البيانات منتشرة في الفضاء على شكل نقاط، يتم حساب المسافة بين هذه النقاط من خلال العديد من الطرق منها المسافة الإقليدية Euclidean distance والمسافة التربيعية Square distance، وبعدها يتم اختيار الجيران الأقرب للمشاهدة المراد تقديرها، حيث يتم اختيار عدد الجيران المراد التقدير به من خلال التجريب، ويتم اعتماد العدد الذي يعطي أقل خطأ للتقدير، ومن ثم يتم أخذ الوسط الحسابي للجيران الأقرب كتقدير للقيمة المطلوبة (الجبيلي، ٢٠٢٣).

حيث تعتبر خوارزمية الجار الأقرب أحد الأدوات الأكثر دقة في عملية التصنيف أو التنبؤ، وتعمل على النحو التالي (Kok et al., 2021):

١) تحديد مسافة جميع النقاط (N) ذات البُعد (D) إلى نقطة البيانات غير المصنفة. ويتطلب تحديد هذه المسافة أداة من أدوات قياس المسافة (مثل المسافة الإقليدية أو التربيعية أو مانهاتن Manhattan). ويتم ذلك عادة عن طريق التعامل مع البيانات على أنها منتشرة في الفضاء على شكل نقاط.

٢) يتم الحصول على نقاط (K) الأقرب إلى نقطة البيانات غير المصنفة، حيث تختار الخوارزمية نقاط البيانات (K) ذات المسافة (D) الأقل من نقطة البيانات غير المصنفة.

٣) وأخيراً يتم التصويت بالأغلبية لتحديد فئة نقطة البيانات غير المصنفة، حيث يشارك أقرب جيران (K) إلى نقطة البيانات غير المصنفة في جلسة تصويت لتحديد فئة النقطة غير المصنفة.

وتتميز خوارزمية الجار الأقرب بدقة عالية وعدم تأثرها بالقيم المتطرفة، كما أنها لا تحتاج إلى افتراضات عند إدخال البيانات، ويمكن تطبيقها على البيانات الإسمية والكمية، كما يمكن استخدامها في مشاكل التصنيف والإنحدار. ولكن من عيوبها أنها تحتاج إلى دقة حساب عالية وإلى ذاكرة كبيرة، حيث تخزن جميع بيانات التدريب، فهي حساسة لحجم البيانات حيث أنه كلما زاد حجم البيانات كلما زاد حاجتها للذاكرة واستغرقت وقت أكبر في التصنيف (الجبيلي، ٢٠٢٣).

٤/٢/١- الانحدار اللوجستي Logistic Regression:

الإنحدار اللوجستي هو طريقة نمذجة رياضية يمكن استخدامها لدراسة العلاقة بين متغير تابع نوعي ومتغير أو أكثر من المتغيرات المستقلة أياً كان نوعها كمية أو نوعية، وتعتبر نماذج الانحدار اللوجستي حالة خاصة من حالات النماذج الخطية المعممة Generalized Linear Models، كما يطلق على هذه النماذج أحياناً نماذج اللوجيت، وتستخدم للتنبؤ بوجود صفة أو خاصية معينة بالاعتماد على قيم متغير أو مدموعة من المتغيرات المستقلة الأخرى التي لها علاقة بالمتغير التابع تماماً كما هو الحال في نماذج الانحدار العامة، كما يمتاز بأنه أكثر مرونة من نماذج الانحدار التقليدية إذ بالإمكان افتراض علاقة معينة تربط بين المتغير التابع والمتغيرات المستقلة، كما يمكن تحويله إلى الشكل الخطي باستخدام تحويلة Logit Transformation، إلا أنه يعاب عليه أنه لا يأخذ الزمن الذي يسبق حدوث الحدث بعين الاعتبار، وعدم مقدرته على التعامل مع بيانات الاختفاء وخاصة البيانات المقطعية (بسيوني، ٢٠٢٢).

ومن أهم استخدامات الانحدار اللوجستي ما يأتي (الجبلي، ٢٠٢٠):

١) يستخدم عندما يكون المتغير التابع وصفياً بحيث تشير القيمة (واحد) إلى وجود خاصية معينة، بينما تشير القيمة (صفر) إلى عدم وجود تلك الخاصية، وتكون المتغيرات المستقلة كمية أو نوعية.

٢) يستخدم للتمييز بين مجموعتين أو أكثر عندما تكون هناك بعض المتغيرات المستقلة وصفية أو نوعية أو عندما تكون المتغيرات لا تتبع التوزيع الطبيعي.

وفي الأخير تتوقف عملية تقييم أدوات التنقيب في البيانات واختيار أفضل هذه الأدوات على مدى توافر المعايير التالية: (Çığşar & Ünal, 2019)

١) **دقة التصنيف Classification Accuracy:** قدرة النموذج على التنبؤ بشكل صحيح، والتي يتم التعبير عنها كنسبة مئوية

٢) **السرعة Speed:** تشير إلى الوقت المستغرق لإعداد النموذج.

٣) **المتانة Robustness:** تشير إلى القدرة على التنبؤ بالنموذج بشكل صحيح على الرغم من أن البيانات تحتوي على بيانات شاذة وقيم مفقودة.

٤) **قابلية التوسع Scalability:** تشير إلى قدرة النموذج على أن يكون دقيقًا ومنتجًا أثناء التعامل مع كمية كبيرة من البيانات.

٥) **القابلية للتفسير Interpretability:** تشير إلى مستوى الفهم الذي يوفره النموذج.

٦) **هيكل القاعدة Rule Structure:** تشير إلى مدى سهولة فهم هيكل قواعد الخوارزميات.

حيث يفضل اختيار خوارزمية التنقيب في البيانات التي تعطي دقة تنبؤ أو تصنيف أكثر، والتي لا تحتاج إلى وقت طويل في إعداد نموذج التنبؤ، والتي لا تتأثر بأية قيم شاذة أو مفقودة، والتي لديها قدرة على التعامل مع كم هائل من البيانات، بالإضافة إلى سهولة فهم النموذج وسهولة فهم هيكل قواعد الخوارزمية المختارة.

٣/١- دور أساليب التنقيب في البيانات في مجال المراجعة:

ظل استخدام مراقب الحسابات لأساليب المراجعة التقليدية لعقود من الزمن، إلا أن ثورة التقدم التكنولوجي وتطلعات أصحاب المصالح ألزمت مراقب الحسابات على تطوير طرق المراجعة التقليدية تماشياً مع متطلبات التكنولوجيا وأصحاب المصالح، حيث أرجعت بعض الدراسات خطأ التقدير والتنبؤ باستمرارية الشركة نتيجة اعتماد مراقب الحسابات على الأساليب التقليدية عند الحكم والتقرير عن استمرارية المنشأة، والتي يعاب عليها بارتفاع نسبة الخطأ في التنبؤ (أمين، ٢٠٢٣).

ومن ثم تزايدت أهمية استخدام الأساليب الحديثة في عملية المراجعة نتيجة الحاجة المستمرة لتطوير أدوات المراجع الخارجي لاكتشاف الأخطاء الجوهرية بالقوائم المالية والتقرير عنها، وكذلك لتطوير أساليب للتنبؤ للتقرير عن مدى استمرارية المنشأة، وتجدر الإشارة إلى أن المراجع غير مسئول عن ضمان مستقبل المنشأة واستمراريتها وعدم تصفيتها، بل مسئوليته تتمثل في أن يدق

ناقوس الخطر إذا ظهرت دلائل ومؤشرات تشير إلى وجود مشاكل تتعلق بمدى قدرة المنشأة على الاستمرارية (الشورى والغندور، ٢٠٢٠).

ومن منظور مراقب الحسابات فإن أساليب التنقيب في البيانات تساعد في العديد من المهام منها زيادة كمية البيانات التي من الممكن تحليلها، وتقليل الوقت المطلوب للتحليل، حيث أنه وفقاً للطرق التقليدية للمراجعة يقوم مراقب الحسابات بتحليل جزء فرعي من البيانات بدلاً من تحليل البيانات جميعها، وهذا الإجراء قد يتضمن احتمالية وجود مخاطر جوهرية لا يمكن اكتشافها، وبالتالي تساعد أساليب التنقيب في البيانات من تحليل البيانات جميعها في وقت قصير جداً لاكتشاف التحريفات الجوهرية بدلاً من الطرق التقليدية للمراجعة (Alareeni & Hamdan, 2022)، كما تساهم أيضاً في تحسين جودة عملية المراجعة من خلال دعم أحكام مراقب الحسابات بصفة عامة والتقرير عن استمرارية الشركة بصفة خاصة (Chu & Yong, 2021).

أشارت العديد من الدراسات أن استخدام أساليب التنقيب في البيانات في مجال المراجعة يساهم بدرجة كبيرة في تحسين دقة التنبؤ بمدى وجود أخطاء جوهرية في القوائم المالية، وكذلك التنبؤ بالغش في القوائم المالية. حيث قامت دراسة (Sharma et al., 2019) بتقييم تقنيات التنقيب والتعلم الآلي في الكشف عن عمليات الاحتيال في القوائم المالية، وتوصلت إلى أن أفضل تقنيات التنقيب وفقاً لمعدل الدقة كانت الغابة العشوائية، يليها شجرة القرار. كما توصلت دراسة (Mohammadi et al., 2020) أن الشبكة العصبية الإصطناعية تؤدي أداءً جيدًا في الكشف عن عمليات الاحتيال في القوائم المالية مقارنةً بالشبكة البايزية والتحليل التمييزي والإنحدار اللوجستي وآلة ناقل الدعم. وفي نفس السياق قامت دراسة (علي، ٢٠٢٣) بتقييم انعكاسات استخدامات تقنيات التنقيب في البيانات في التنبؤ برأي المراجع الخارجي وأثرها على عدالة القوائم المالية، وتوصلت إلى أن أساليب التنقيب في البيانات (الشبكات العصبية، شجرة القرارات، نموذج تعلم الآلة) كانت أكثر دقة في التنبؤ برأي المراجع الخارجي عن عدالة القوائم المالية مقارنة بالأساليب التقليدية.

ومن ناحية أخرى أشارت العديد من الدراسات أن أساليب التنقيب في البيانات تساهم بدرجة أكبر في تدعيم الرأي المهني لمراقب الحسابات حول تقييم قدرة المنشأة على الاستمرارية. حيث قامت دراسة (Yang & Nazareth, 2022) بتقييم فعالية استخدام المراجع الخارجي لخوارزميات التعلم الآلي عند التقرير عن استمرارية المنشأة، وتوصلت الدراسة إلى أن الخوارزميات الجينية كانت

أكثر دقة، يليها خوارزمية الغابات العشوائية، ثم خوارزمية Gradient boost، ثم خوارزمية الجار الأقرب، يليها خوارزمية Adaptive Boost. كما تناولت دراسة (أمين، ٢٠٢٣) أثر اعتماد مراقب الحسابات على نموذج هجين من أدوات الذكاء الإصطناعي على تحسين جودة حكمه بشأن الاستمرارية، وتوصلت إلى أن خوارزمية الغابات العشوائية كانت أكثر دقة في التنبؤ، يليها خوارزمية الجار الأقرب، ثم الشبكات العصبية الإصطناعية، وأخيراً الإنحدار اللوجستي، وبالتالي حققت نماذج أدوات الذكاء الإصطناعي دقة تنبؤية عالية مقارنة بالنماذج التقليدية.

وفي هذا الصدد أيضاً نجد أن أساليب التنقيب في البيانات تساهم بدرجة كبيرة في التنبؤ بالتعثر المالي للشركات، وبالتالي من الممكن للمراجع الخارجي أن يستخدمها كأداة تكنولوجية متطورة في عملية التنبؤ بالتعثر المالي، ومن ثم التقرير عن الاستمرارية بشكل أكثر دقة. حيث قامت دراسة (Shahwan & Fadel, 2020) بتطوير نموذج موثوق به للتنبؤ بالتعثر المالي للشركات المصرية الصغيرة والمتوسطة الحجم، من خلال الاعتماد على الشبكات العصبية الاصطناعية والانحدار اللوجستي والتحليل التمييزي، وتوصلت الدراسة إلى أن نموذج الشبكات العصبية الإصطناعية تفوّق بدرجة أكبر على الأساليب التقليدية الأخرى مثل الانحدار اللوجستي والتحليل التمييزي فيما يتعلق بالدقة التنبؤية. أيضاً قامت دراسة (Máté et al., 2023) بتقييم أداء خوارزميات التعلم الآلي المختلفة في التنبؤ بالتعثر المالي للشركات، وتوصلت الدراسة إلى أن أساليب التنقيب في البيانات (شجرة القرار، نموذج AdaBoost، نموذج Gradient boost) أكثر دقة حيث بلغ معدل الدقة في كل منهم (١٠٠ %) مقارنة بالأساليب التقليدية (الإنحدار اللوجستي).

ومن خلال ما سبق تتضح أهمية استخدام أساليب التنقيب في البيانات في مجال المراجعة، حيث أنها تُعد أكثر دقة وفعالية مقارنة بالأساليب التقليدية من حيث:

- تدعيم الرأي المهني لمراقب الحسابات حول تقييم قدرة المنشأة على الاستمرارية، حيث أنها تعد أكثر دقة في عملية التنبؤ بالتعثر المالي وبالتالي يمكن للمراجع استخدام هذه الأساليب كأداة للتنبؤ بالتعثر المالي بدقة عالية، ومن ثم التقرير عن الاستمرارية بشكل أكثر دقة.
- تحسين كفاءة عملية المراجعة الخارجية بشكل ملحوظ، وذلك من خلال المرونة والسرعة في أداء المهام، وتقديم النتائج في الوقت المناسب، وزيادة درجة الثقة لدى المستخدمين في صدق تقارير المراجعة الخارجية.

- تحسين دقة التنبؤ بمدى وجود أخطاء جوهرية في القوائم المالية، وكذلك التنبؤ بالغش في القوائم المالية.
- تحسين دقة التنبؤ بفشل الشركات.
- تدعيم إجراءات عملية التخطيط للمراجعة وتقييم المخاطر.

٢- ماهية وطبيعة التعثر المالي للشركات (المفهوم – الأسباب والمحددات – نماذج القياس):

٢/١- مفهوم التعثر المالي:

تناولت العديد من الدراسات محاولة تعريف التعثر المالي في الأدب المحاسبي، حيث أشارت دراسة (Fredrick & Osazemen, 2018) بأن التعثر المالي هو حالة من عدم التوازن التي تصيب الشركة، وينتج التعثر المالي نتيجة للعديد من الأسباب الداخلية والخارجية والتي تؤثر على قدرة الشركة على سداد التزاماتها المطلوبة، وبالتالي فهو الحالة التي تكون فيها الشركة غير قادرة على توليد أموال كافية للوفاء بالتزاماتها المالية وقت استحقاقها.

وفي نفس السياق أوضحت دراسة (Widhiadnyana & Ratnadi, 2019) أن التعثر المالي عملية تنتج عن تفاعل العديد من الأسباب والعوامل عبر مراحل زمنية طويلة وصولاً إلى حالة عدم قدرة الشركة على تسديد قروضها ومحاولة الحصول على قروض جديدة وفقدان التوازن المالي والنقدي والتشغيلي، وبالتالي فهو مرحلة من التدهور المالي تسبق الإفلاس أو التصفية.

لذا يمكن القول أن الشركات المتعثرة مالياً قد تفلت من الإفلاس في حالة إعادة البناء المبكر لعملياتها التشغيلية (John & Ogechukwu, 2018)، وأن التعثر المالي قد يكون محفزاً لزيادة الإنتاجية والدخل للمساهمة في حل المشاكل المالية التي قد تتعرض لها الشركة المتعثرة مالياً (محمد، ٢٠٢١).

٢/٢- أسباب التعثر المالي:

يمكن تقسيم أسباب التعثر المالي إلى أسباب داخلية وأسباب خارجية، وتتمثل الأسباب الداخلية فيما يأتي (Michalkova et al., 2018; Kuloba & Ombaba, 2019):

- عدم الكفاءة الإدارية الناتجة عن ضعف الإدارة وأخطائها.
- تزايد أعباء الديون.
- حجم المصنع غير الاقتصادي (القيام بمشاريع كبيرة جداً بالنسبة للشركة).
- الإفراط في الاستثمار في الأصول الثابتة.

- حجم الإنتاج المفرط مقارنة بهيكل التمويل.
- الإفراط في الاقتراض.
- عدم كفاية الرقابة المالية.
- عدم كفاية رأس المال.
- سوء إدارة رأس المال العامل.
- زيادة النفقات.
- التسويق غير الكافي.
- الآثار السلبية لعمليات الإندماج والاستحواذ.
- عدم ملائمة السياسات المالية للشركة.

بينما تتمثل الأسباب الخارجية للتعثر المالي فيما يأتي (Michalkova et al., 2018; Kuloba & Ombaba, 2019):

- السياسات الحكومية المتعلقة بالضرائب ورسوم الاستيراد.
- تزايد حدة المنافسة.
- الارتفاع في أسعار الصرف.
- ارتفاع مستويات التضخم.
- التغيرات السلبية في حجم الطلب على منتجات الشركات.
- التغير في أسعار المدخلات بشكل غير مواتٍ.

بالإضافة إلى ما سبق قد يكون هناك أسباب بيئية أو طبيعية للتعثر المالي مثل الحرائق والزلازل والبراكين والكوارث الطبيعية الأخرى، وهذه الأسباب تكون خارج نطاق تحكم الشركة إلا أنه قد تسبب التعثر المالي لبعض الشركات (محمد، ٢٠٢١).

٣/٢- محددات التعثر المالي:

أشارت دراسة (إبراهيم وأبو السعود، ٢٠١٨؛ Paramartha & Wiagustini, 2021) إلى أن أهم محددات التعثر المالي في الشركات تتمثل في الآتي:

١/٣/٢- السيولة: حيث يعتبر مدى توافر سيولة كافية بالشركة أحد العوامل الرئيسية الهامة لتجنب تعرض الشركات للتعثر المالي، حيث أنه كلما ارتفعت نسب السيولة كلما انخفضت احتمالية تعرض

المنشأة لمخاطر التعثر المالي، ومن النسب التي يمكن استخدامها لقياس السيولة نسبة التداول ونسبة السيولة السريعة.

٢/٣/٢- الربحية: تعتبر الربحية مقياساً لقدرة الشركة على تحقيق أرباح، وكذلك مقياساً لمستوى كفاءتها التشغيلية في إدارة أصولها، ويعد مستوى الربحية أحد عوامل الجذب الرئيسية للدائنين والمستثمرين، حيث أنه كلما انخفضت ربحية المنشأة، كلما زادت درجة الصعوبة في الحصول على رأس مال من أطراف خارجية. والنسب التي يمكن أن تقيس الربحية هي نسبة هامش الربح إلى المبيعات، والقوة الإيرادية، ومعدل العائد على الاستثمار، والعائد على حقوق الملكية. ومن المعروف أن الربحية لها تأثير سلبي كبير على التعثر المالي للمنشأة، مما يعني أنه كلما ارتفعت الربحية كلما انخفضت احتمالية تعرض المنشأة لمخاطر التعثر المالي.

٣/٣/٢- الرافعة المالية: تعتبر الرافعة المالية أحد المحددات الهامة لحدوث التعثر المالي، وتعني الرافعة المالية مدى اعتماد الشركة على الديون في تمويل أصولها، ومن النسب التي يمكن أن تقيس الرافعة المالية نسبة الديون إلى إجمالي الأصول، ومن المعروف أن الرافعة المالية المقاسة بنسبة الديون لها تأثير إيجابي كبير على التعثر المالي للمنشأة، مما يعني أنه كلما ارتفعت نسبة الدين كلما زادت احتمالية تعرض المنشأة لمخاطر التعثر المالي.

٤/٣/٢- معدل النمو: يعتبر معدل النمو في المبيعات أحد العوامل المحددة لاحتمالية تعرض الشركات لمخاطر التعثر المالي، حيث أنه كلما زاد معدل النمو في المبيعات كلما انخفض احتمال تعرض الشركة لمخاطر التعثر المالي.

٥/٣/٢- القيمة السوقية للأسهم: تعتبر القيمة السوقية للأسهم أحدد محددات تعرض الشركات للتعثر المالي، حيث أنه كلما زادت القيمة السوقية للأسهم كلما انخفضت احتمالية تعرض الشركة لمخاطر التعثر المالي.

٦/٣/٢- التدفقات النقدية: تعتبر التدفقات النقدية أحد الأدوات الرئيسية للتنبؤ بالتعثر المالي للشركات وخاصة التدفقات النقدية التشغيلية، حيث أنه كلما زادت التدفقات النقدية للشركة (خاصة التدفقات النقدية التشغيلية) كلما انخفض احتمال تعرض الشركة لمخاطر التعثر المالي.

٤/٢- نماذج التنبؤ بالتعثر المالي:

توجد العديد من النماذج التي اهتمت بالتنبؤ بسلامة المركز المالي للشركات ومدى ابتعادها أو قربها من مرحلة التعثر المالي، وتعتمد هذه النماذج على بعض المؤشرات التي تستخرج من التقارير

المالية، ثم تُعالج باستخدام أساليب إحصائية معينة، ويوضح الجدول رقم (١) التالي العديد من النماذج الكمية المستخدمة في التنبؤ بالتعثر المالي:

جدول رقم (١)

نماذج التنبؤ بالتعثر المالي

م	Model	Analysis Techniques	Formula
١	Altman (1968), Z-Score	التحليل التمييزي Discriminant Analysis	$Z = 1.2\,X1 + 1.4\,X2 + 3.3\,X3 + 0.6\,X4 + 1.0\,X5$ **حيث أن:** X1: رأس المال العامل إلى إجمالي الأصول. X2: الأرباح المحتجزة إلى إجمالي الأصول. X3: الأرباح قبل الفائدة والضريبة إلى إجمالي الأصول. X4: القيمة السوقية لحقوق الملكية إلى إجمالي الالتزامات. X5: المبيعات إلى إجمالي الأصول. **الدرجة:** Z < (١.٨) : شركات متعثرة Z تتراوح بين (١.٩ – ٢.٩) : منطقة رمادية Z > (٢.٩) : شركات غير متعثرة
٢	Modified Altman	التحليل التمييزي Discriminant Analysis	$Z'' = 6.56\,X1 + 3.26\,X2 + 6.72\,X3 + 1.05\,X4$ **حيث أن:** X1: رأس المال العامل إلى إجمالي الأصول. X2: الأرباح المحتجزة إلى إجمالي الأصول. X3: الأرباح قبل الفائدة والضريبة إلى إجمالي الأصول. X4: القيمة الدفترية لحقوق الملكية إلى إجمالي الالتزامات. **الدرجة:** Z < (١.١) : شركات متعثرة Z تتراوح بين (١.١ – ٢.٦) : منطقة رمادية Z > (٢.٦) : شركات غير متعثرة
٣	Springate, 1978	التحليل التمييزي Discriminant Analysis	$S = 1.03\,X1 + 3.07\,X2 + 0.66\,X3 + 0.4\,X4$ **حيث أن:** X1: رأس المال العامل إلى إجمالي الأصول. X2: الأرباح قبل الفائدة والضريبة إلى إجمالي الأصول. X3: صافي الأرباح قبل الضرائب إلى الإلتزامات المتداولة. X4: المبيعات إلى إجمالي الأصول. **الدرجة:** S > (٠.٨٦٢) : شركات متعثرة. S ≤ (٠.٨٦٢) : شركات غير متعثرة.
٤	Taffler and Tishaw	التحليل التمييزي Discriminant Analysis	$Z = 0.53X_1 + 0.13X_2 + 0.18X_3 + 0.16X_4$ **حيث أن:** X1: الأرباح قبل الضرائب إلى الالتزامات المتداولة. X2: الأصول المتداولة إلى إجمالي الالتزامات. X3: الالتزامات المتداولة إلى إجمالي الأصول. X4: (الأصول المتداولة – الالتزامات المتداولة) / تكاليف التشغيل. **الدرجة:** Z > (٠.٢) : شركات متعثرة Z تتراوح بين (٠.٢ – ٠.٣) : منطقة رمادية Z > (٠.٣) : شركات غير متعثرة

م	Model	Analysis Techniques	Formula
٥	**Blums, 2003** **D-Score Model**	**الإنحدار اللوجستي** **Logit Model**	$\mathbf{D = -4.907 - 2.11\,X1 + 0.0006\,X2 - 1.734\,X3 - 0.016\,\Delta P + 0.005\,\Delta S + 5.885\,X4}$ **حيث أن:** X1: صافي الدخل إلى إجمالي الأصول X2: إجمالي الديون إلى إجمالي القيمة السوقية لحقوق المليكة. X3: إجمالي القيمة السوقية لحقوق المليكة إلى إجمالي الأصول. ΔP: تغير سعر السهم لمدة ٦ أشهر ΔS: نمو المبيعات لمدة ٣ سنوات X4: الإلتزامات المتداولة إلى إجمالي الأصول **الدرجة:** إذا كانت قيمة (D-Score) أكبر من أو تساوي (٠.٥) فإن المنشأة تكون مهددة بخطر التعثر المالي، أما إذا كانت أقل من (٠.٥) تكون المنشأة غير متعثرة.
٦	**Zmijewski, 1984**	**الإنحدار الاحتمالي** **Probit Model**	$\mathbf{P = \Phi(-4.336 - 4.513\,X1 + 5.679\,X2 + 0.004\,X3}$ **حيث أن:** X1: صافي الدخل إلى إجمالي الأصول. X2: إجمالي الالتزامات/ إجمالي الأصول. X3: الأصول المتداولة إلى الالتزامات المتداولة. **Φ** : دالة الكثافة التراكمية لمتغير عادي قياسي. **الدرجة:** P < (٠.٣) : شركات غير متعثرة P تتراوح بين (٠.٣ – ٠.٦) : منطقة رمادية P > (٠.٦) : شركات متعثرة
٧	**Ohlson, 1980** **O-Score Model**	**الإنحدار اللوجستي** **Logit Model**	$O = -1.32 - 0.407 log X_1 + 6.03X_2 - 1.43X_3 + 0.0757X_4 - 1.72X_5 - 2.37X_6 - 1.83X_7 + 0.285X_8 - 0.521X_9$ **حيث أن:** X1: إجمالي الأصول/ مؤشر مستوى سعر الناتج القومي الإجمالي GDP. X2: إجمالي الالتزامات/ إجمالي الأصول. X3: رأس المال العامل/ إجمالي الأصول. X4: الخصوم المتداولة/ الأصول المتداولة. X5: تأخذ القيمة (١) إذا تجاوز أجمالي الالتزامات إجمالي الأصول، والقيمة (صفر) خلاف ذلك. X6: صافي الدخل/ إجمالي الأصول. X7: التدفق النقدي/ إجمالي الالتزامات. X8: تأخذ القيمة (١) إذا كان صافي الدخل سالباً خلال آخر عامين، والقيمة (صفر) خلاف ذلك. X9: نسبة التغير في صافي الدخل، ويتم احتسابها من خلال المعادلة $\frac{NI_t - NI_{t-1}}{\lvert NI_t\rvert + \lvert NI_{t-1}\rvert}$ ، حيث NI_t و NI_{t-1} هما صافي الدخل لآخر سنة وما قبلها على التوالي. **الدرجة:** O > (٠.٣) : شركات غير متعثرة O تتراوح بين (٠.٣ – ٠.٦) : منطقة رمادية O < (٠.٦) : شركات متعثرة

م	Model	Analysis Techniques	Formula
٨	Shumway, 2001	Hazared model	$H = \{1 + \exp\left(-\begin{bmatrix}13.303 - 1.982\,X1 + 3.593\,X2 - 0.467\,X3 \\ -1.809\,X4 + 5.791\,X5\end{bmatrix}\right)\}^{-1}$ **حيث أن:** X1: صافي الدخل إلى إجمالي الإلتزمات. X2: إجمالي الإلتزمات إلى إجمالى الأصول. X3: لوغاريتم (عدد الأسهم القائمة في السوق مضروبة في سعر السهم نهاية العام مقسومة علي إجمالي القيمة السوقية). X4: العائد التراكمي للشركة للفترة t-1 مطروحاً منه العائد التراكمي للسندات للفترة t-1 X5: الإنحراف المعياري للبواقي المشتق من تراجع عوائد الأسهم الشهرية للشركة عن عائد السوق في العام t-1 **الدرجة:** إذا كانت قيمة (H) أكبر من أو تساوي (٠.٥) فإن المنشأة تكون مهددة بخطر التعثر المالي، أما إذا كانت أقل من (٠.٥) تكون المنشأة غير متعثرة.

المصدر: (Ashraf et al., 2019; Stankevičienė & Prazdeckaitė, 2021)

وعلى الرغم من أهمية هذه النماذج في التنبؤ بالتعثر المالي، إلا أنه مع ظهور تقنيات الذكاء الإصطناعي فقد تحول الاهتمام إلى استخدام هذه التقنيات الحديثة في التنبؤ بالتعثر المالي للشركات، وذلك نظراً لما توفره هذه التقنيات من دقة تنبؤية مرتفعة مقارنة بالأساليب التقليدية التي كانت مستخدمة من قبل في عملية التنبؤ.

٣- إطار مقترح لاستخدام المراجع الخارجي لأساليب التنقيب في البيانات للتنبؤ بالتعثر المالي:

في ضوء الحاجة إلى تطبيق أساليب التنقيب في البيانات في مجال المراجعة، وفي ظل التأكيد على أهميتها في العديد من الدراسات والأبحاث، تم وضع تصور مقترح لنموذج متكامل لتنفيذ عملية التنبؤ بالتعثر المالي باستخدام هذه الأساليب، حتى يتمكن المراجع الخارجي من التقرير عن استمرارية المنشأة. وذلك كما هو موضح في الشكل رقم (٢) التالي:

شكل رقم (٢) إطار مقترح لاستخدام المراجع الخارجي
لأساليب التنقيب في البيانات بغرض التنبؤ بالتعثر المالي

١/٣ – فهم طبيعة النشاط

من البديهي أنه قبل استخدام أدوات التنقيب في البيانات في التنبؤ بالتعثر المالي؛ فإنه يتعين على المراجع الخارجي فهم طبيعة عمل الشركة التي يقوم بمراجعة حساباتها، حتى يمكن تحقيق المنفعة الأعظم من تطبيق أساليب التنقيب في البيانات. بالإضافة إلى ذلك يتعين على المراجع الخارجي فهم طبيعة المنافسة داخل الصناعة التي تعمل فيها الشركة محل المراجعة، حتى يتمكن من تحديد البيانات والمعلومات اللازم توافرها للتنبؤ بالتعثر المالي للشركة محل المراجعة.

٢/٣- جمع البيانات اللازمة للتنبؤ بالتعثر المالي

وهنا يبدأ المراجع في جمع البيانات والمعلومات المالية وغير المالية التي تفيده في الحكم على مدى قدرة المنشأة على الاستمرارية في المستقبل، والتي تتسم كذلك بأهمية نسبية مرتفعة في التنبؤ بالتعثر المالي، ومن ثمّ التقرير عن الاستمرارية.

ومن أهم المؤشرات المالية التي يمكن للمراجع الاعتماد عليها عند التنبؤ بالتعثر المالي، وبالتالي التقرير عن استمرارية المنشأة (على سبيل المثال لا الحصر) ما يلي:

- مؤشرات عن مدى قدرة المنشأة على الوفاء بالتزاماتها قصيرة الأجل. مثال ذلك: (نسبة الأصول المتداولة إلى الالتزامات المتداولة، نسبة الأصول سريعة التحول إلى نقدية إلى الالتزامات المتداولة، معدل النقدية، نسبة الأصول المتداولة إلى إجمالي الأصول)
- مؤشرات عن القدرة على سداد الالتزامات طويلة الأجل. مثال ذلك: (نسبة إجمالي الالتزامات إلى إجمالي الأصول، نسبة إجمالي الالتزامات إلى إجمالي حقوق الملكية).
- مؤشرات عن نمط الأرباح أو الخسائر ومدى تكرارها. مثال ذلك: (نسبة الأرباح أو الخسائر المرحلة إلى إجمالي حقوق الملكية، نسبة الأرباح أو الخسائر المرحلة إلى إجمالي الأصول، نسبة صافي الربح قبل الفوائد والضرائب إلى صافي المبيعات، نسبة صافي الربح النهائي إلى صافي المبيعات، نسبة صافي الربح قبل الفوائد والضرائب إلى إجمالي الأصول، نسبة صافي الربح النهائي إلى إجمالي الأصول، نسبة صافي الربح النهائي إلى إجمالي حقوق الملكية).
- مؤشرات عن توزيعات الأرباح النقدية.
- مؤشرات عن القدرة على سداد استحقاقات الدائنين في مواعيدها.
- مؤشرات عن نمط تعامل الشركة مع الموردين والدائنين، ومدى قدرتها على الحصول على الائتمان التجاري.
- مؤشرات عن مدى قدرة المنشأة في الحصول على تمويل جديد.

ومن أهم المؤشرات غير المالية التي يمكن للمراجع الاعتماد عليها عند التنبؤ بالتعثر المالي، وبالتالي التقرير عن استمرارية المنشأة (على سبيل المثال لا الحصر) ما يلي:

- مؤشرات عن فقد الشركة لمديرين مهمين دون إيجاد من يحل محلهم.
- مؤشرات عن فقدان سوق رئيسية أو إمتياز أو مورد رئيسي أو عميل رئيسي.

- مؤشرات عن القوى العاملة ومدى استقرارها.
- مؤشرات عن التغير بالزيادة أو النقص في المستلزمات الهامة والرئيسية التي تعتمد عليها الشركة.

٣/٣- معالجة البيانات

لكي يتمكن المراجع الخارجي من التنبؤ بالتعثر المالي للشركة محل المراجعة فإنه لا بد من أن يقوم أولاً بمعرفة ماهية وطبيعة البيانات التي قام بجمعها، حيث أن معرفة طبيعة البيانات بصورة جيدة تعني مساعدة المراجع في استخدام الخوارزميات المناسبة لطبيعة البيانات بدقة عالية، مما يساهم في تعظيم فرص النجاح بالإضافة إلى رفع الفاعلية والكفاءة في التطبيق.

أيضاً يجب أن يقوم المراجع ببعض المهام الفرعية مثل إلغاء البيانات المتكررة، التصحيح الشكلي لبعض البيانات، معالجة القيم المفقودة، معالجة القيم الشاذة أو المتطرفة، إضافة بيانات جديدة، وذلك حتى لا يتعرض نموذج التنبؤ إلى الخلل وعدم الدقة بسبب وجود أخطاء في عملية معالجة البيانات، مما قد يؤدي في النهاية إلى الوصول إلى نتائج مضللة قد تؤدي إلى إصدار المراجع لرأي مخالف للحقيقة فيما يتعلق بتعثر أو عدم تعثر الشركة محل المراجعة.

٤/٣- اختيار أفضل المؤشرات للتنبؤ بالتعثر المالي

فبعد أن يقوم المراجع الخارجي بفهم النشاط، وجمع البيانات التاريخية للشركة محل المراجعة والشركات العاملة معها في نفس الصناعة، ومعالجة هذه البيانات، يصبح بإمكانه تصنيف الشركات إلى شركات معثرة وشركات أخرى غير متعثرة بناءاً على البيانات التاريخية التي تم جمعها.

وبذلك يمكن للمراجع استخدام إحدى التقنيات القوية والمستخدمة على نطاق واسع لاختيار أفضل المؤشرات ذات التأثير المعنوي على التنبؤ بالتعثر المالي للشركات، ومن هذه الخوارزميات التي يمكن للمراجع الاعتماد عليها خوارزمية (Recursive Feature Elimination) ويمكن التعبير عنها باختصار (RFE) وهي عبارة عن خوارزمية تكرارية تعمل بشكل منهجي على إزالة المؤشرات الأقل أهمية من مجموعة البيانات، مما يؤدي في النهاية إلى اختيار أفضل المؤشرات وبالتالي تحسين أداء النموذج المستخدم في التنبؤ.

٥/٣- اختيار أفضل خوارزميات التنقيب في البيانات للتنبؤ بالتعثر المالي

وهنا يقوم المراجع بتقسيم البيانات التي تم جمعها إلى نوعين هما: (بيانات تدريب، وبيانات اختبار)، حيث يمكن للمراجع أن يستخدم (٧٥ %) من البيانات التي تم جمعها لتدريب الخوارزميات على التنبؤ، ومن ثمّ اختيار الخوارزميات الأكثر دقة في التنبؤ بالتعثر المالي، ثم بعد ذلك يقوم باستخدام باقي البيانات التي تم جمعها وهي (٢٥ %) من البيانات لاختبار مدى دقة هذه الخوارزميات في التنبؤ بالتعثر المالي.

وهنا ينبغي الإشارة إلى أن المراجع يمكنه اختيار أكثر من خوارزمية بشكل مبدئي يرى أنها قد تكون مناسبة في عملية التنبؤ، ويمكن في هذه الحالة أن يقوم باختيار هذه الخوارزميات بناءاً على

مدى توافر مجموعة من المعايير وهي – كما سبق أن ذكرنا – تتمثل في (قدرة النموذج على التنبؤ بشكل صحيح، الوقت المستغرق لإعداد النموذج، قدرة النموذج على التنبؤ بشكل صحيح في حالة أن البيانات تحتوي على قيم شاذة و/ أو قيم مفقودة، قدرة النموذج على التعامل مع كمية كبيرة من البيانات، مستوى الفهم الذي يوفره النموذج، سهولة فهم الخوارزميات).

ومن خوارزميات التنقيب في البيانات الأكثر شيوعاً التي يمكن للمراجع الخارجي الاعتماد عليها عند التنبؤ بالتعثر المالي للشركة محل المراجعة (على سبيل المثال لا الحصر) ما يلي:

- الشبكات العصبية الإصطناعية Artificial Neural Networks.
- خوارزمية الجار الأقرب K–Nearest Neighbor.
- خوارزمية الغابة العشوائية Random Forest.
- خوارزمية آلة ناقل الدعم Support Vector Machine (SVM).

وبعد أن يقوم المراجع باختيار الخوارزميات المستخدمة في التنبؤ بشكل مبدئي، يقوم بالمفاضلة بين هذه الخوارزميات، لاختيار الخوارزمية الأكثر دقة في التنبؤ، ويمكنه الاعتماد على واحدة أو أكثر من معايير تقييم الأداء التالية:

- معامل الدقة Accuracy.
- معامل Kappa.
- معامل F1.Score.

ووفقاً لهذه المعاملات يتم المفاضلة بين الخوارزميات المستخدمة في التنبؤ بشكل مبدئي، ثم اختيار الخوارزمية الأكثر دقة في التنبؤ، والتي سوف يعتمد عليها المراجع في التنبؤ بالتعثر المالي للشركة محل المراجعة.

٦/٣- اختيار النموذج التنبؤي

وفي هذه المرحلة يكون المراجع قد توصل إلى اختيار أفضل المؤشرات التي يمكنه الاعتماد عليها في عملية التنبؤ بالتعثر المالي، وكذلك اختيار أفضل الخوارزميات التي يمكنه استخدامها في عملية التنبؤ (وهي الخوارزميات الأكثر دقة تنبؤية) والتي تم اختيارها بناء على معايير تقييم الأداء. وبالتالي يمكن للمراجع أن يقوم بإدخال المؤشرات عن الفترة الزمنية التي يرغب في التنبؤ بالتعثر المالي عنها، وبالتالي يصل إلى درجة تنبؤ بأعلى دقة ممكنة عن مدى وجود أو عدم وجود تعثر مالي للشركة محل المراجعة، ومن ثمّ يمكنه التقرير عن استمرارية المنشأة بأعلى دقة ممكنة.

تحليل نتائج الدراسة التطبيقية

تستهدف الدراسة التطبيقية بناءًا على ما تم عرضه في الدراسة النظرية إلى قياس أثر استخدام المراجع الخارجي لأساليب التنقيب في البيانات على دقة التنبؤ بالتعثر المالي للشركات المقيدة في البورصة المصرية. ولتحقيق هذه الهدف تم اتباع الخطوات التالية:

١- توصيف مجتمع وعينة البحث:

تكون مجتمع الدراسة من الشركات المقيدة في سوق الأوراق المالية المصرية في مؤشر EGX100، وذلك خلال الفترة من ٢٠١٨م حتى ٢٠٢٢م، وتم اختيار عينة الدراسة في ضوء الاعتبارات التالية:

- استبعاد المؤسسات المالية نظراً لخضوعها لقواعد تنظيمية ومتطلبات قياس وإفصاح خاصة بها نتيجة لاختلاف طبيعة نشاطها، قياساً على (علي، ٢٠٢٠؛ محمد، ٢٠٢١؛ أمين، ٢٠٢٣).
- استبعاد الشركات التي لم نتمكن من الحصول على بياناتها خلال أي سنة من سنوات الدراسة.
- استبعاد الشركات التي تقوم بنشر قوائمها المالية بعملة خلاف الجنيه المصري (علي، ٢٠٢٠؛ محمد، ٢٠٢١؛ أمين، ٢٠٢٣).
- التركيز على الشركات التي ينتهي العام المالي في ٣١ ديسمبر من كل عام، للتغلب على عدم اتساق الأعوام المالية (محمد، ٢٠٢١).

ويوضح الجدول رقم (٢) التالي إجراءات اختيار حجم العينة في ضوء الاعتبارات السابق ذكرها، وذلك على النحو التالي:

جدول رقم (٢) إجراءات اختيار العينة محل الدراسة

اجراءات اختيار العينة	عدد الشركات	عدد المشاهدات
الشركات المقيدة في البورصة المصرية في مؤشر EGX100	١٠٠	٥٠٠
يستبعد		
المؤسسات المالية	(٢٢)	(١١٠)
الشركات التي لا ينتهي عامها المالي في ٣١ ديسمبر	(١٧)	(٨٥)
الشركات التي تنشر قوائمها المالية بعملة خلاف الجنيه المصري	(٢)	(١٠)
الشركات التي لم نستطع الحصول على بياناتها خلال أي سنة من سنوات الدراسة	(٩)	(٤٥)
= الشركات الخاضعة للدراسة والتحليل	٥٠	٢٥٠
يستبعد: مشاهدات تحتوي على قيم شاذة (إشارات غير منطقية)	-	(١٥)
حجم العينة محل الدراسة		٢٣٥

وبذلك بلغ عدد شركات العينة محل الدراسة (٥٠) شركة، بإجمالي مشاهدات (٢٣٥) مشاهدة على مدار الفترة من ٢٠١٨م حتى ٢٠٢٢م.

٢- النموذج المبدئي لمتغيرات الدراسة ومؤشرات القياس

يمكن توضيح متغيرات الدراسة على النحو التالي:

١/٢- المتغيرات المستقلة:

اعتماداً على الدراسات السابقة تم التوصل إلى مجموعة من المؤشرات المالية وغير المالية، والتي أمكن الحصول عليها من واقع التقارير والقوائم والمالية المنشورة للشركات محل الدراسة، وذلك بغرض التنبؤ بالتعثر المالي للشركات المقيدة في البورصة المصرية. ويوضح الجدول رقم (٣) النموذج المبدئي لمتغيرات الدراسة، على النحو التالي:

جدول رقم (٣) النموذج المبدئي لمتغيرات الدراسة

رمز المتغير	المتغير والقياس
X1	الأصول المتداولة / الالتزامات المتداولة
X2	الأصول السريعة / الالتزامات المتداولة
X3	النقدية بالبنوك والصندوق / الالتزامات المتداولة
X4	الأصول المتداولة / إجمالي الأصول
X5	صافي المبيعات / إجمالي الأصول
X6	صافي المبيعات / رأس المال العامل
X7	صافي المبيعات / رصيد العملاء وأوراق القبض
X8	ت. البضاعة المباعة / قيمة المخزون
X9	مجمل الربح / صافي المبيعات
X10	صافي الربح قبل الفوائد والضرائب / صافي المبيعات
X11	صافي الربح النهائي / صافي المبيعات
X12	صافي الربح قبل الفوائد والضرائب / إجمالي الأصول
X13	صافي الربح النهائي / إجمالي الأصول
X14	صافي الربح النهائي / إجمالي حقوق الملكية
X15	الأرباح المحتجزة / إجمالي حقوق الملكية
X16	الأرباح المحتجزة / إجمالي الأصول
X17	إجمالي الالتزامات / إجمالي الأصول
X18	حقوق الملكية / إجمالي الأصول
X19	إجمالي الالتزامات / إجمالي حقوق الملكية
X20	عمر الشركة = لوغاريتم عدد سنوات عمر الشركة
X21	حجم الشركة = لوغاريتم إجمالي الأصول
X22	المراجعة بواسطة BIG4 = يأخذ القيمة (١) في حالة المراجعة من قبل BIG4 والقيمة (صفر) خلاف ذلك
X23	وجود خسائر السنة الحالية = يأخذ القيمة (١) في حالة وجود خسائر السنة الحالية والقيمة (صفر) خلاف ذلك
Y	تقدير مراقب الحسابات الفعلي بشأن التعثر المالي.. يأخذ القيمة (١) في حالة وجود تعثر مالي، والقيمة (صفر) خلاف ذلك.

وكما هو موضح بالجدول رقم (٣) السابق، تم الاعتماد في بناء النموذج المبدئي للدراسة على (٢٣) متغير مستقل، منهم (٢١) متغير كمي تشمل المتغيرات من (X1:X21)، (٢) متغير نوعي أو وصفي تشمل المتغيرات (X22, X23).

٢/٢- المتغير التابع: التنبؤ بالتعثر المالي للشركات

تم قياس التعثر المالي للشركات، من خلال تحليل محتوى تقرير مراقب الحسابات، سواء كان استقرار مالي أو تعثر مالي، حيث تم إعطاء القيمة (١) في حالة وجود تعثر مالي (عدم وجود استقرار مالي)، والقيمة (صفر) في حالة عدم وجود تعثر مالي (وجود استقرار مالي).

٣- أدوات التحليل الإحصائي:

في ضوء طبيعة وأنواع البيانات والمتغيرات وطرق القياس وأغراض التحليل، تم استخدام لغة البرمجة (R Language) كأداة للتحليل، كما تم الاعتماد على مجموعة من الأساليب الإحصائية لتحليل البيانات وإختبار الفروض، تتمثل فيما يلي:

١/٣- التحليل الوصفي Descriptive Analysis:

وذلك بغرض إعطاء صورة عامة متكاملة عن الخصائص الإحصائية لمتغيرات الدراسة، من خلال المعلمات الإحصائية الآتية: (المتوسط الحسابى Mean – الإنحراف المعيارى Standard Deviation – أقل قيمة Minimum – الربيع الأدنى (IQ1) – الربيع الأعلى (IQ3) – أعلى قيمة Maximum).

٢/٣- معامل تضخم التباين (VIF):

بغرض الكشف عما إذا كان هناك مشكلة تعدد أو إزدواج خطي بين البيانات، حيث أنه كلما كانت قيمته أقل من (١٠) فإن هذا يعني عدم وجود مشكلة إزدواج خطي بين المتغيرات المستقلة بعضها البعض.

٣/٣- خوارزمية Recursive Feature Elimination (RFE):

وذلك بغرض إزالة المتغيرات الأقل أهمية من مجموعة البيانات، مما يؤدي في النهاية إلى تحسين أداء النموذج، من خلال التوصل إلى أفضل المتغيرات ذات التأثير على المتغير التابع.

٤/٣- خوارزميات التنقيب في البيانات:

حيث سيتم الاعتماد على خوارزميات الغابات العشوائية (RF)، الجار الأقرب (KNN)، الشبكة العصبية الإصطناعية (ANN) كأساليب حديثة للتنبؤ بالتعثر المالي، بينما سيتم استخدام الإنحدار اللوجستي كأحد الأساليب التقليدية للتنبؤ بالتعثر المالي.

٣/٥- معايير تقييم أداء الخوارزميات:

في نماذج التعلم الآلي، يتم تقييم الخوارزميات بشكل أساسي من خلال مصفوفة التصنيف أو الارتباك. وبالنسبة للمتغير التابع؛ عندما يكون ثنائي الفئة (أي يأخذ القيمة صفر أو ١)، تكون المصفوفة عبارة عن مربع مكون من إثنين في إثنين، وكما يتضح من الجدول رقم (٤) نجد أنه توجد أربعة حالات للتصنيف، يمكن توضيحها على النحو التالي:

- True Positive (TP): وتشير إلى عدد الحالات الإيجابية الصحيحة، وتتحقق هذه الحالة عندما يوجد تعثر مالي، وتم التنبؤ بوجود تعثر مالي.
- True Negative (TN): وتشير إلى عدد الحالات السلبية الصحيحة، وتتحقق هذه الحالة عندما لا يوجد تعثر مالي، وتم التنبؤ بعدم وجود تعثر مالي.
- False Positive (FP): وتشير إلى عدد الحالات الإيجابية الخاطئة، وتتحقق هذه الحالة عندما يوجد تعثر مالي، وتم التنبؤ بعدم وجود تعثر مالي، (وهذا يمثل خطأ من النوع الأول).
- False Negative (FN): وتشير إلى عدد الحالات السلبية الخاطئة، وتتحقق هذه الحالة عندما لا يوجد تعثر مالي، وتم التنبؤ بوجود تعثر مالي، (وهذا يمثل خطأ من النوع الثاني).

جدول رقم (٤) مصفوفة التصنيف أو الارتباك (Confusion matrix)

		تنبؤي	
		إيجابي Positive	سلبي Negative
فعلي	إيجابي Positive	TP	FP
	سلبي Negative	FN	TN

وفيما يتعلق بمعايير تقييم أداء الخوارزميات؛ فإنه تم الاعتماد على العديد من المقاييس لقياس أداء كلا من الشبكات العصبية الاصطناعية ANN، الغابات العشوائية RF، الجار الأقرب KNN، الإنحدار اللوجستي LR. وتتضمن هذه المقاييس معيار الدقة (Accuracy)، ومعيار (Kappa)، ومعيار (F1.Score). وذلك كما هو موضح على النحو التالي:

١/٥/٣- معامل الدقة (Accuracy):

يعتبر معامل الدقة من أكثر الأساليب شيوعاً واستخداماً في مجال التصنيف، ويعتمد معامل الدقة في حسابه على العلاقة التالية (Mahmoud et al., 2021):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

٣/٥/٢- معامل (Kappa):

يقيس معامل Kappa مدى اتساق نموذج التنبؤ، حيث يقارن نتيجة النموذج المتوقع مع النتائج الفعلية. ويأخذ قيمة إحصائية تتراوح بين (صفر ، ١). وكلما اقتربت القيمة من (١) كلما كان النموذج ذات اتساق كبير، حيث تشير قيمة Kappa الأعلى إلى تصنيف أفضل. وتتمثل معادلة قيمة Kappa على النحو التالي (Seliem, 2022):

$$\boldsymbol{kappa} = \frac{\left[\frac{TP+TN}{N}\right] - \left[\frac{(TP+FN)\times(TP+FP)\times(TN+FN)}{N^2}\right]}{1 - \left[\frac{(TP+FN)\times(TP+FP)\times(TN+FN)}{N^2}\right]}$$

٣/٥/٣- معامل (F1.Score):

يُعرّف معيار F1.Score على أنه الوسط التوافقي لمعيار (Precision) ومعيار (Recall)، ويتم حسابه من خلال العلاقة التالية (Mahmoud et al., 2021):

$$F1.Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

حيث أن:

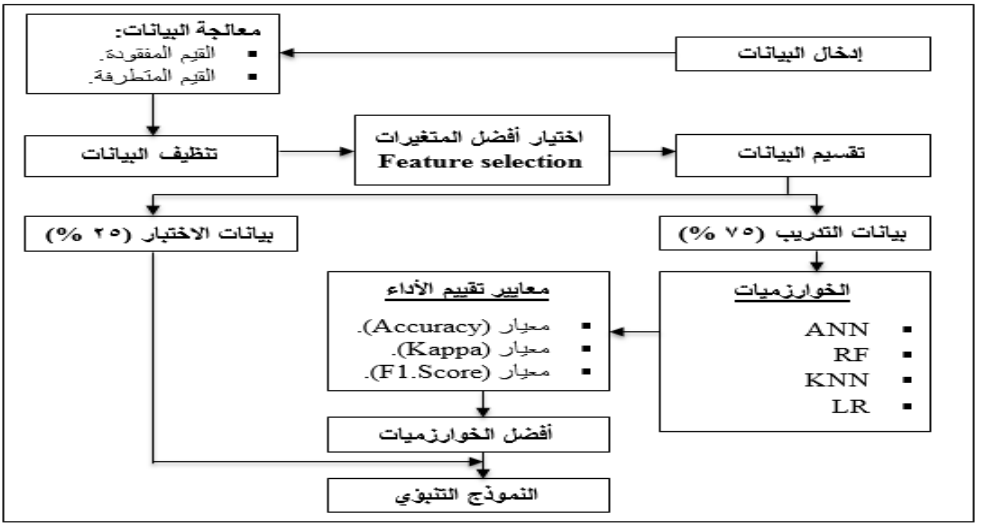
$$Precision = \frac{TP}{TP+FP} \quad , \quad Recall = \frac{TP}{TP+FN}$$

٤- منهجية إجراء الدراسة التطبيقية:

يوضح الشكل رقم (٣) رسم توضيحي لمنهجية إجراء الدراسة التطبيقية، على النحو التالي:

شكل رقم (٣)

رسم توضيحي لمنهجية إجراء الدراسة التطبيقية



يوضح الشكل رقم (٣) المنهجية المتبعة عند إجراء الدراسة التطبيقية، حيث تم أولاً إدخال البيانات الخاصة بعينة الدراسة، ثم تم إجراء معالجة للبيانات للتأكد من عدم وجود قيم مفقودة أو متطرفة، وكذلك للتأكد من عدم وجود مشكلة إزدواج خطي بين المتغيرات المستقلة، ثم بعد ذلك تم تنظيف البيانات ومعالجة مشكلات الإزدواج الخطي والقيم المتطرفة، وبعدها تم اختيار أفضل المتغيرات المستقلة التي تفسر بدرجة أكبر المتغير التابع، وذلك باستخدام خوارزمية (RFE)، ثم تم اختبار الخوارزميات (الشبكات العصبية الإصطناعية، الغابات العشوائية، الجار الأقرب، الإنحدار اللوجستي) باستخدام معايير للتقييم الأداء (Accuracy، Kappa، F1.Score)، وذلك للمفاضلة بين هذه الخوارزميات واستخلاص أيهم أكثر دقة، وترتيبهم من حيث درجة الدقة في التنبؤ، وأخيراً تحديد أفضل الخوارزميات وبناء النموذج التنبؤي.

٥- نتائج التحليل الإحصائي لمتغيرات الدراسة:

وفي هذا الجزء سوف نستعرض أهم نتائج التحليل الوصفي لمتغيرات الدراسة، وطرق معالجة البيانات تمهيداً لاختيار أفضل المتغيرات المستقلة التي يمكن الاعتماد عليها عند التنبؤ بالتعثر المالي باستخدام أساليب التنقيب في البيانات.

١/٥- توصيف متغيرات الدراسة (الإحصاء الوصفي):

يوضح الجدول رقم (٥) مقاييس الإحصاءات الوصفية لمتغيرات الدراسة، وذلك من خلال حساب بعض مقاييس النزعه المركزية ومقاييس التشت وهي: (المتوسط الحسابي، الانحراف المعياري، أقل قيمة، والربيع الأدني، والربيع الأعلى، وأكبر قيمة).

جدول رقم (٥) مقاييس الاحصاء الوصفي لمتغيرات الدراسة

المتغير	المتوسط	الانحراف المعياري	أقل قيمة	الربيع الأدني	الربيع الأعلى	أكبر قيمة
X1	1.6	1.3	0.36	1	1.9	14
X2	1.1	0.96	0.2	0.52	1.2	6.5
X3	0.29	0.44	0.0034	0.066	0.33	3.5
X4	0.57	0.24	0.065	0.41	0.77	0.98
X5	0.72	0.61	0.013	0.22	0.98	3.5
X6	1.5	1.3	0.019	0.47	2	7.6
X7	17	32	0.089	2.3	12	200
X8	9.8	26	0.058	1.4	7.2	287
X9	0.25	0.28	-2.3	0.14	0.37	0.77
X10	0.056	0.83	-8.4	0.06	0.23	1.1
X11	0.029	0.75	-8.2	0.031	0.16	1.3
X12	0.081	0.087	-0.25	0.027	0.13	0.35
X13	0.053	0.073	-0.26	0.018	0.094	0.3
X14	0.037	1	-14	0.051	0.2	3.3
X15	-0.043	1.9	-23	0.027	0.34	2.4
X16	0.078	0.16	-0.7	0.013	0.14	1.2
X17	0.54	0.2	0.093	0.39	0.7	1
X18	0.46	0.2	0.014	0.3	0.61	0.95
X19	2.4	5.8	0.1	0.65	2.3	72
X20	1.4	0.27	0.48	1.3	1.6	2.1
X21	9.5	0.8	6.6	9	9.9	11

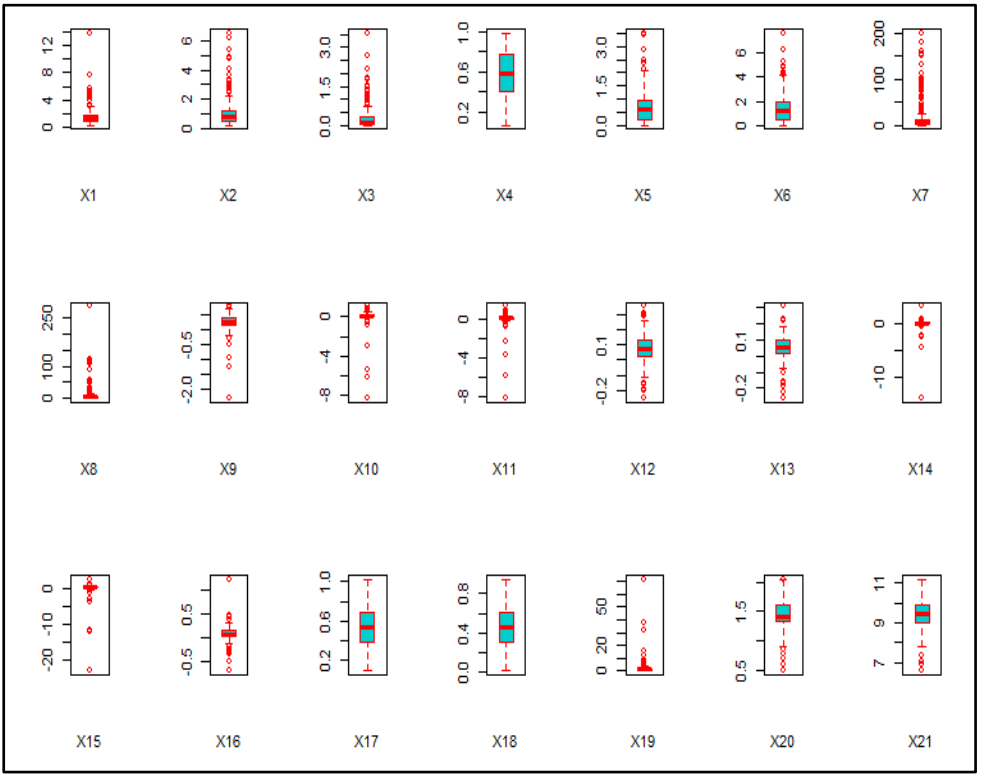
المصدر: نتائج التحليل الإحصائي.

يوضح الجدول رقم (٥) أن هناك متغيرات عند مقارنة متوسطها الحسابي بالحد الأعلى يتضح أن هناك فرقاً كبيراً بين المتوسط والحد الأعلى، مما يشير إلى وجود قيم شاذة أو متطرفة. فعلى سبيل نجد أن المتغير (X1) والذي يمثل (نسبة الأصول المتداولة إلى الالتزامات المتداولة) بلغ المتوسط الحسابي له (١.٦)، وانحرافه المعياري (١.٣)، وعند مقارنة المتوسط الحسابي له بالحد الأعلى (١٤) يتضح أن هناك فرقاً كبيراً مما يوحي إلى أن هناك قيم شاذة، أيضاً في المتغير (X2) والذي يمثل (نسبة الأصول السريعة إلى الالتزامات المتداولة) بلغ المتوسط الحسابي له (١.١)، وانحرافه المعياري (٠.٩٦). كما يتضح أن المتوسط الحسابي للمتغير (X7) والذي يمثل (نسبة

صافي المبيعات إلى رصيد العملاء وأوراق القبض) يساوي (١٧)، بينما الربيع الأعلى يساوي (١٢)، وأكبر قيمة له تساوي (٢٠٠)، كما نجد أن المتوسط الحسابي للمتغير (X8) والذي يمثل (نسبة ت. البضاعة المباعة إلى قيمة المخزون) يساوي (٩.٨)، بينما الربيع الأعلى يساوي (٧.٢)، وأكبر قيمة له تساوي (٢٨٧).

ومن خلال ما سبق نلاحظ أنه ربما يكون المتوسط الحسابي متأثراً بهذه المشاهدات الكبيرة، ويوضح الشكل رقم (٦) التالي رسم توضيحي للمتغيرات المستقلة (الكمية).

شكل رقم (٦) رسم توضيحي للمتغيرات المستقلة من خلال (Boxplot)



المصدر: نتائج التحليل الإحصائي.

يتضح من الشكل رقم (٦) وجود مشكلة القيم الشاذة أو القيم المتطرفة في البيانات، مما يؤدي إلى بناء خوارزميات ينتج عنها نماذج تنبؤيه غير دقيقة. ولعل ظهور هذه المشكلة يرجع بدرجة كبيرة إلى أن هذه البيانات تشمل قطاعات متنوعة تختلف فيما بينها من حيث طبيعة النشاط، ومن

ثمّ حجم رأس المال المستثمر بها، وكذلك تختلف من حيث مؤشرات الربحية والتشغيل والسيولة والمديونية وغيرها من الخصائص الأخرى.

ويوضح الجدول رقم (٦) الإحصاءات الوصفية للمتغيرات النوعية، حيث يتضح أن المتغير المستقل (X22) والذي يعبر عن المراجعة بواسطة BIG4 ويأخذ القيمة (١) إذا كانت الشركة تتم مراجعتها بواسطة أحد مكاتب المراجعة الكبرى، والقيمة (صفر) خلاف ذلك، نجد أن هناك (١٠٧) مشاهدة بنسبة (٤٥.٣٦ %) تمت مراجعتها بواسطة BIG4، بينما هناك (١٢٨) مشاهدة بنسبة (٥٤.٦٤ %) لم يتم مراجعتها بواسطة BIG4.

جدول رقم (٦) الإحصاءات الوصفية للمتغيرات النوعية

المتغير	الفئة	التكرار	النسبة المئوية
X22	0	128	54.64%
	1	107	45.36%
الإجمالي		**235**	**100%**
X23	0	206	87.66%
	1	29	12.34
الإجمالي		**235**	**100%**
Y	0	213	90.63%
	1	22	9.37%
الإجمالي		**235**	**100%**

المصدر: نتائج التحليل الإحصائي.

كما يتضح أن المتغير المستقل (X23) والذي يعبر عن وجود خسائر في السنة الحالية، ويأخذ القيمة (١) في حالة وجود خسائر في السنة الحالية، والقيمة (صفر) خلاف ذلك، نجد أن هناك (٢٩) مشاهدة بنسبة (١٢.٣٤ %) تحقق بها وجود خسائر في السنة الحالية، بينما هناك (٢٠٦) مشاهدة بنسبة (٨٧.٦٦ %) لا توجد بها خسائر في السنة الحالية.

ويتضح أيضاً أن المتغير التابع (Y) والذي يعبر عن تقدير مراقب الحسابات الفعلي بشأن التعثر المالي، ويأخذ القيمة (١) في حالة وجود تعثر مالي، والقيمة (صفر) خلاف ذلك، نجد أن هناك (٢٢) مشاهدة بنسبة (٩.٣٧ %) تحقق بها وجود تعثر مالي، بينما هناك (٢١٣) مشاهدة بنسبة (٩٠.٦٣ %) لا يوجد بها تعثر مالي.

٢/٥- معالجة مشكلة الإزدواج الخطي (Multicollinearity):

يقصد بالإزدواج الخطي وجود علاقة ارتباط قوية ومعنوية بين إثنين أو أكثر من المتغيرات التفسيرية، ويعتبر من أهم الآثار السلبية المترتبة على وجود إزدواج خطي بين المتغيرات التفسيرية هو عدم بناء نماذج تنبؤية دقيقة يمكن الاعتماد عليها في عملية التنبؤ.

وتم التأكد من تواجد / أو عدم تواجد مشكلة الإزدواج الخطي بين المتغيرات المستقلة بعضها البعض من خلال الاعتماد على معامل تضخم التباين (VIF) Varience Inflation Factore لكل متغير من المتغيرات المستقلة، حيث إذا كانت قيمة (VIF) أقل من (١٠) فإنه يمكن الحكم بعدم وجود إزدواج خطي. ويوضح الجدول رقم (٧) قيمة معامل تضخم التباين (VIF) للمتغيرات المستقلة، وذلك كما هو موضح على النحو التالي:

جدول رقم (٧) قيمة معامل تضخم التباين (VIF) للمتغيرات المستقلة

المتغير	X1	X2	X3	X4	X5	X6	X7	X8
VIF	**14.41**	**13.38**	2.44	**12**	**13.54**	**28.54**	8.57	8.23
المتغير	X9	X10	X11	X12	X13	X14	X15	X16
VIF	**10.59**	**43.97**	**43.45**	**25.12**	**22.61**	**12.18**	**57.85**	7.37
المتغير	X17	X18	X19	X20	X21	X22	X23	--
VIF	270.95	245.87	64.47	2.8	2.65	1.77	3.33	--

المصدر: نتائج التحليل الإحصائي.

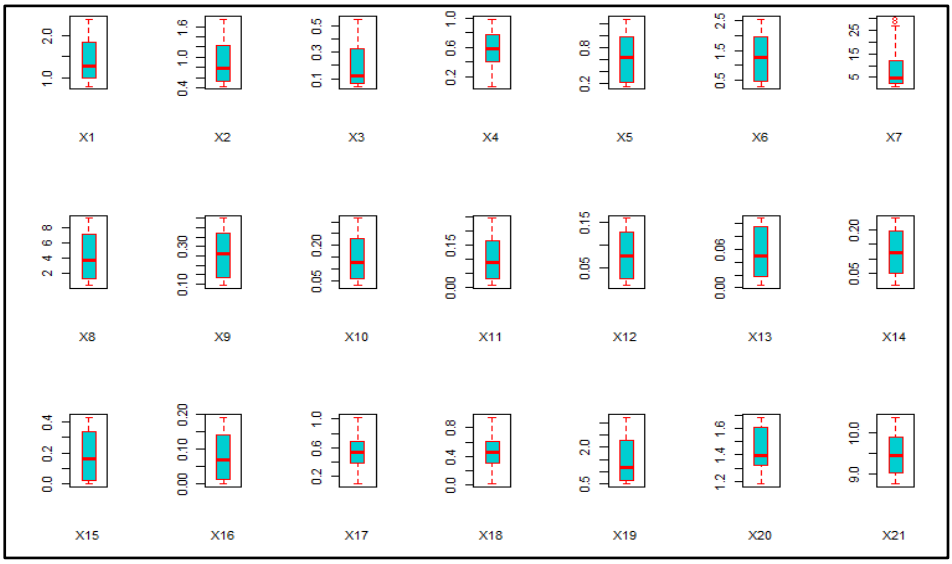
يوضح الجدول رقم (٧) السابق أن هناك متغيرات كثيرة قيمة (VIF) الخاص بها أكبر من (١٠)، مما يؤكد على أن هناك مشكلة (Multicollinearity)، مما يؤكد على أن نموذج الانحدار اللوجستي لن يفي بالغرض، كما أن نتائجه تكون متدنية مقارنة بالخوارزميات الأخرى. وتم معالجة هذه المشكلة باستخدام انحدار الحرف اللوجستي (Lasso logistic regression) بدلاً من الإنحدار اللوجستي العادي.

٣/٥- معالجة مشكلة البيانات المتطرفة (Outlier Data):

لكي يتم بناء نماذج تنبؤيه دقيقة، لا بد وأن يتم بناؤها على بيانات سليمة خالية من أي مشاكل مثل مشكلة القيم المتطرفة. ولذا سوف يتم معالجة القيم الشاذه باستخدام الربعيات الأدني (IQ1) والربعيات الأعلى (IQ3) للتخلص من القيم الشاذه. ويوضح الشكل رقم (٧) رسم توضيحي للمتغيرات المستقلة من خلال (Boxplot) بعد أن تمت معالجة القيم الشاذة بها، مما يجعل بناء النماذج أكثر دقة، وبالتالي يمكن الاعتماد عليها فى عملية التنبؤ.

شكل رقم (٧)

رسم توضيحي للمتغيرات المستقلة من خلال (Boxplot) بعد معالجة القيم المتطرفة



المصدر: نتائج التحليل الإحصائي.

٤/٥ – اختيار أفضل المتغيرات المستقلة (Feature Selection):

لا شك أن إحدى الخطوات الحاسمة في بناء نماذج تنبؤية دقيقة وفعالة هي اختيار المتغيرات ذات التأثير المعنوي على المتغير التابع. وفي سبيل تحقيق ذلك تم استخدام خوارزمية Recursive Feature Elimination (RFE)، حيث أنها تعتبر إحدى التقنيات القوية والمستخدمة على نطاق واسع لاختيار المتغيرات المعنوية، وهي عبارة عن خوارزمية تكرارية تعمل بشكل منهجي على إزالة المتغيرات الأقل أهمية من مجموعة البيانات، مما يؤدي في النهاية إلى تحسين أداء النموذج.

ويوضح الجدول رقم (٨) نتائج خوارزمية (RFE) للمتغيرات المستقلة الأكثر أهمية، حيث يتضح أن خوارزمية (RFE) صنفت تسعة متغيرات على أنها مهمة ولها تأثير جوهري على المتغير التابع، وهي كالتالي: (X1, X2, X3, X4, X8 ,X15, X16, X19, X21). كما يتضح أن هذه المتغيرات التسعة بالفعل كانت متنوعة واشتملت جميع مؤشرات الأداء، حيث أنها تضم مؤشرات تقيس درجة السيولة وهي (X1, X2, X3, X4)، ومؤشراً لقياس مستوى النشاط أو الكفاءة أو التشغيل (X8)، ومؤشرات لقياس مستوى الربحية (X15, X16)، ومؤشراً لقياس نسبة المديونية (X19)، ومؤشراً لقياس حجم الشركة من حيث إجمالي أصولها (X21).

جدول رقم (٨) نتائج خوارزمية (RFE) للمتغيرات المستقلة الأكثر أهمية

رمز المتغير	المتغير والقياس	درجة الأهمية	ترتيب الأهمية	القرار
X1	الأصول المتداولة / الالتزامات المتداولة	8.57	4	أفضل المتغيرات المستقلة
X2	الأصول السريعة / الالتزامات المتداولة	9.81	1	
X3	النقدية بالبنوك والصندوق / الالتزامات المتداولة	7.22	7	
X4	الأصول المتداولة / إجمالي الأصول	7.39	6	
X8	ت. البضاعة المباعة / قيمة المخزون	6.87	8	
X15	الأرباح المحتجزة / إجمالي حقوق الملكية	7.59	5	
X16	الأرباح المحتجزة / إجمالي الأصول	8.61	2	
X19	إجمالي الالتزامات / إجمالي حقوق الملكية	6.57	9	
X21	حجم الشركة = لوغاريتم إجمالي الأصول	8.59	3	

المصدر: نتائج التحليل الإحصائي.

ويوضح الجدول رقم (٩) مقاييس الإحصاء الوصفي للمتغيرات المستقلة النهائية، والذي يتضح من خلاله عدم وجود قيم متطرفة، مما يعني أنه يمكننا بناء خوارزميات ينتج عنها نماذج تنبؤية دقيقة. وذلك على النحو التالي:

جدول رقم (٩) مقاييس الإحصاء الوصفي لمتغيرات الدراسة

المتغير	المتوسط	الانحراف المعياري	أقل قيمة	الربيع الأدني	الربيع الأعلى	أكبر قيمة
X1	1.5	0.55	0.8	1	1.9	2.4
X2	0.93	0.46	0.43	0.52	1.2	1.7
X3	0.21	0.18	0.042	0.066	0.33	0.55
X4	0.57	0.24	0.065	0.41	0.77	0.98
X8	4.3	3.2	0.49	1.4	7.2	9.2
X15	0.19	0.16	0	0.027	0.34	0.43
X16	0.082	0.071	0	0.013	0.14	0.19
X19	1.5	1	0.49	0.65	2.3	3.2
X21	9.5	0.56	8.7	9	9.9	10

المصدر: نتائج التحليل الإحصائي.

بالإضافة إلى ذلك نجد أن الجدول رقم (١٠) يوضح قيمة معامل تضخم التباين (VIF) للمتغيرات المستقلة النهائية، حيث يتضح من الجدول أن هذه البيانات لا تعانى من مشكلة الإزدواج الخطي (Multicollinearity)، حيث جاءت قيمة (VIF) لجميع المتغيرات أقل من (١٠).

جدول رقم (١٠) قيمة معامل تضخم التباين (VIF) للمتغيرات المستقلة النهائية

المتغير	X1	X2	X3	X4	X8	X15	X16	X19	X21
VIF	4.57	4.58	2.06	2.58	2.09	9.94	9.24	2.49	1.81

المصدر: نتائج التحليل الإحصائي.

ومن خلال ما سبق يمكننا الآن البدء في التحليل باستخدام خوارزميات التنقيب في البيانات للتبؤ بالتعثر المالي الشركات، حيث أصبحت البيانات المستخدمة في التحليل تمكننا من بناء نماذج تنبؤية أكثر دقة. وهو ما سيتم توضيحه فيما يأتي.

٦- نتائج التحليل باستخدام أساليب التنقيب في البيانات:

١/٦- عينة التدريب وعينة الاختبار:

تم بناء النموذج النهائي باستخدام المتغيرات المستقلة الناتجة من خوارزمية (RFE) للتنبؤ بالمتغير التابع، وبعد أن تم اختيار المتغيرات المستقلة لبناء نموذج تصنيف، تم تقسيم مجموعة البيانات المجمعة التي تحتوى على (٩) متغيرات مستقلة إلى مجموعتين فرعيتين هما: بيانات تدريب (Training Dataset) بنسبة (٧٥ %)، وبيانات اختبار (Test Dataset) بنسبة (٢٥ %). ويوضح الجدول رقم (١١) تقسيم البيانات لتدريب واختبار الخوارزميات، وذلك على النحو التالي:

جدول رقم (١١) تقسيم البيانات لتدريب واختبار الخوارزميات

القيمة	حجم البيانات	بيانات التدريب (٧٥ %) (Training Dataset)	بيانات الاختبار (٢٥ %) (Test Dataset)
	٢٣٥	١٧٧	٥٨
٠	٢١٣	١٦٠	٥٣
١	٢٢	١٧	٥

المصدر: نتائج التحليل الإحصائي.

يوضح الجدول رقم (١١) السابق أن البيانات الكلية (٢٣٥) مشاهدة تم تقسيمها إلى (٢١٣) مشاهدة ليس لديها تعثر مالي، (٢٢) مشاهدة بها تعثر مالي. كما يوضح أن بيانات التدريب والتي تم استخدامها لتدريب واختيار أفضل الخوارزميات تم تقسيمها إلى (١٦٠) مشاهدة ليس لديها تعثر مالي، (١٧) مشاهدة بها تعثر مالي، بينما بيانات الاختبار (والتي يُختبر بها مدى كفاءة هذه الخوارزميات) تم تقسيمها إلى (٥٣) مشاهدة ليس لديها تعثر مالي، (٥) مشاهدات بها تعثر مالي.

٢/٦- نتائج خوارزميات التنقيب في البيانات:

يوضح الجدول رقم (١٢) التالي مقارنة بين أداء الخوارزميات المستخدمة في التنبؤ (ANN، RF، KNN، LR) بناءاً على معايير تقييم الأداء لهذه الخوارزميات (معامل الدقة Accuracy، معامل Kappa، معامل F1.Score).

١/٢/٦- نتائج التحليل باستخدام خوارزمية الشبكة العصبية الإصطناعية (ANN):

يوضح الجدول رقم (١٢) التالي نتائج التصنيف الناتجة عن خوارزمية ANN وكذلك نتائج جودة أداء الخوارزمية، كما يأتي:

يتضح من الجدول أن خوارزمية "الشبكة العصبية الإصطناعية" نتج عنها أنها صنفت (٣) مشاهدات على أنها يوجد بها تعثر مالي وهي يوجد بها تعثر مالي، كما أنها صنفت (٥٣) مشاهدة على أنه لا يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي. ولكنها صنفت أيضاً أن هناك (٢) مشاهدة يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي، وبالتالي فإن معدل خطأ التصنيف للخوارزمية ككل يساوي (٣.٤ %).

جدول رقم (١٢)نتائج التحليل باستخدام خوارزمية الشبكة العصبية الإصطناعية (ANN)

مصفوفة التصنيف			معايير تقييم الأداء		
خوارزمية ANN		تنبؤي		معامل الدقة Accuracy	معامل Kappa	معامل F1.Score
		يوجد تعثر مالي	لا يوجد تعثر مالي			
فعلي	يوجد تعثر مالي	٣	٠	%٩٦.٦	%٧٣.٣	%٩٨.١
	لا يوجد تعثر مالي	٢	٥٣			

المصدر: نتائج التحليل الإحصائي.

كما يتضح من الجدول أن هذه الخوارزمية حققت معدل دقة بنسبة (٩٦.٦ %) من عمليات التصنيف الصحيح، بينما حقق معامل Kappa (٧٣.٣ %) من عمليات التصنيف الصحيح، أما معامل (F.1Score) فقد حقق (٩٨.١ %) من عمليات التصنيف الصحيح، وكلما اقتربت قيم هذه المعاملات من الواحد الصحيح كلما دلت على جودة بناء الخوارزمية.

ويوضح الجدول رقم (١٣) التالي الأهمية النسبية لأوزان المتغيرات المستقلة الناتجة من خوارزمية الشبكة العصبية الإصطناعية، وذلك كما يأتي:

جدول رقم (١٣)الأهمية النسبية لأوزان المتغيرات المستقلة (%) الناتجة من خوارزمية (ANN)

المتغير	X1	X2	X3	X4	X8	X15	X16	X19	X21
الأهمية النسبية	٢٨.٣٨	٥٨.٥٩	٩٩.٠٩	٥٩.٥٩	٠.٠٠	١٠٠	٢٧.٢٨	١٠.٥٩	٣٩.٦
ترتيب الأهمية	٦	٤	٢	٣	٩	١	٧	٨	٥

المصدر: نتائج التحليل الإحصائي.

وكما هو موضح بالجدول رقم (١٣) السابق نجد أن أقوي متغير يفسر المتغير التابع (التعثر المالي للشركة) هو المتغير (X15)، ويليه المتغير (X3)، كما نلاحظ أن أقل متغير يفسر المتغير التابع هو المتغير (X19)، ولا يوجد أي تأثير للمتغير (X8).

٢/٢/٦- نتائج التحليل باستخدام خوارزمية الغابة العشوائية (RF):

يوضح الجدول رقم (١٤) التالي نتائج التصنيف الناتجة عن خوارزمية RF وكذلك نتائج جودة أداء الخوارزمية، كما يأتي:

جدول رقم (١٤)نتائج التحليل باستخدام خوارزمية الغابة العشوائية (RF)

مصفوفة التصنيف				معايير تقييم الأداء		
خوارزمية RF		تنبؤي		معامل الدقة Accuracy	معامل Kappa	معامل F1.Score
		يوجد تعثر مالي	لا يوجد تعثر مالي			
فعلي	يوجد تعثر مالي	٢	٠	%٩٤.٨	%٥٤.٩	%٩٧.٢
	لا يوجد تعثر مالي	٣	٥٣			

المصدر: نتائج التحليل الإحصائي.

يتضح من الجدول أن خوارزمية "الغابات العشوائية" نتج عنها أنها صنفت (٢) مشاهدات على أنها يوجد بها تعثر مالي وهي يوجد بها تعثر مالي، كما أنها صنفت (٥٣) مشاهدة على أنه لا يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي. ولكنها صنفت أيضاً أن هناك (٣) مشاهدة يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي، وبالتالي فإن معدل خطأ التصنيف للخوارزمية ككل يساوي (٥.٢ %).

كما يتضح من الجدول أن هذه الخوارزمية حققت معدل دقة بنسبة (٩٤.٨ %) من عمليات التصنيف الصحيح، بينما حقق معامل Kappa (٥٤.٩ %) من عمليات التصنيف الصحيح، أما معامل (F.1Score) فقد حقق (٩٧.٢ %) من عمليات التصنيف الصحيح، وكلما اقتربت قيم هذه المعاملات من الواحد الصحيح كلما دلت على جودة بناء الخوارزمية.

ويوضح الجدول رقم (١٥) التالي الأهمية النسبية لأوزان المتغيرات المستقلة الناتجة من خوارزمية الغابة العشوائية، وذلك كما يأتي:

جدول رقم (١٥)الأهمية النسبية لأوزان المتغيرات المستقلة (%) الناتجة من خوارزمية (RF)

المتغير	X1	X2	X3	X4	X8	X15	X16	X19	X21
الأهمية النسبية	٣٤.٠٩	٥٥.٧	٣١.٤٩	٧٤.٣٣	٠.٠٠	٦٤.٢٧	١٠٠	٥.٥٦	٦٨.٣٨
ترتيب الأهمية	٦	٥	٧	٢	٩	٤	١	٨	٣

المصدر: نتائج التحليل الإحصائي.

وكما هو موضح بالجدول رقم (١٥) السابق نجد أن أقوي متغير يفسر المتغير التابع (التعثر المالي للشركة) هو المتغير (X16)، ويليه المتغير (X4)، كما نلاحظ أن أقل متغير يفسر المتغير التابع هو المتغير (X19)، ولا يوجد أي تأثير للمتغير (X8).

٣/٢/٦- نتائج التحليل باستخدام خوارزمية الجار الأقرب (KNN):

يوضح الجدول رقم (١٦) التالي نتائج التصنيف الناتجة عن خوارزمية KNN وكذلك نتائج جودة أداء الخوارزمية، كما يأتي:

جدول رقم (١٦)نتائج التحليل باستخدام خوارزمية الجار الأقرب (KNN)

مصفوفة التصنيف				معايير تقييم الأداء		
خوارزمية KNN		تنبؤي		معامل الدقة Accuracy	معامل Kappa	معامل F1.Score
		يوجد تعثر مالي	لا يوجد تعثر مالي			
الفعلي	يوجد تعثر مالي	١	٠	%٩٣.١	%٣١.٤	%٩٦.٤
	لا يوجد تعثر مالي	٤	٥٣			

المصدر: نتائج التحليل الإحصائي.

يتضح من الجدول أن خوارزمية "الجار الأقرب" نتج عنها أنها صنفت (١) مشاهدة على أنها يوجد بها تعثر مالي وهي يوجد بها تعثر مالي، كما أنها صنفت (٥٣) مشاهدة على أنه لا يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي. ولكنها صنفت أيضاً أن هناك (٤) مشاهدات يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي، وبالتالي فإن معدل خطأ التصنيف للخوارزمية ككل يساوي (٦.٩ %).

كما يتضح من الجدول أن هذه الخوارزمية حققت معدل دقة بنسبة (٩٣.١ %) من عمليات التصنيف الصحيح، بينما حقق معامل Kappa (٣١.٤ %) من عمليات التصنيف الصحيح، أما

معامل (F.1Score) فقد حقق (٩٦.٤ %) من عمليات التصنيف الصحيح، وكلما اقتربت قيم هذه المعاملات من الواحد الصحيح كلما دلت على جودة بناء الخوارزمية.

ويوضح الجدول رقم (١٧) التالي الأهمية النسبية لأوزان المتغيرات المستقلة الناتجة من خوارزمية الجار الأقرب، وذلك كما يأتي:

يتضح من الجدول أن أقوي متغير يفسر المتغير التابع (التعثر المالي للشركة) هو المتغير (X2)، ويليه المتغير (X1)، كما نلاحظ أن أقل متغير يفسر المتغير التابع هو المتغير (X21)، ولا يوجد أي تأثير للمتغير (X8).

جدول رقم (١٧)الأهمية النسبية لأوزان المتغيرات المستقلة (%) الناتجة من خوارزمية (KNN)

المتغير	X1	X2	X3	X4	X8	X15	X16	X19	X21
الأهمية النسبية	٩٥.٤٨	١٠٠	٨٦.٦٤	٦٠.٢٢	٠.٠٠	٤١.١٦	٧٣.٣٨	٦٢.٣٨	٣٣.٢٨
ترتيب الأهمية	٢	١	٣	٦	٩	٧	٤	٥	٨

المصدر: نتائج التحليل الإحصائي.

٤/٢/٦- نتائج التحليل باستخدام خوارزمية الإنحدار اللوجستي (LR):

يوضح الجدول رقم (١٨) التالي نتائج التصنيف الناتجة عن خوارزمية LR وكذلك نتائج جودة أداء الخوارزمية، كما يأتي:

جدول رقم (١٨)نتائج التحليل باستخدام خوارزمية الجار الأقرب (LR)

مصفوفة التصنيف			معايير تقييم الأداء		
خوارزمية LR		تنبؤي		معامل الدقة Accuracy	معامل Kappa	معامل F1.Score
		يوجد تعثر مالي	لا يوجد تعثر مالي			
الواقع	يوجد تعثر مالي	١	١	%٩١.٤	%٢٤.٩	%٩٥.٤
	لا يوجد تعثر مالي	٤	٥٢			

المصدر: نتائج التحليل الإحصائي.

يتضح من الجدول أن خوارزمية "الإنحدار اللوجستي" نتج عنها أنها صنفت (١) مشاهدة على أنها يوجد بها تعثر مالي وهي يوجد بها تعثر مالي، كما أنها صنفت (٥٢) مشاهدة على أنه لا يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي. ولكنها صنفت أيضاً أن هناك (٤) مشاهدات يوجد بها تعثر مالي وهي لا يوجد بها تعثر مالي، كما أنها صنفت أن هناك (١) مشاهدة لا يوجد

بها تعثر مالي وهي يوجد بها تعثر مالي، وبالتالي فإن معدل خطأ التصنيف للخوارزمية ككل يساوي (٨.٦ %).

كما يتضح من الجدول أن هذه الخوارزمية حققت معدل دقة بنسبة (٩١.٤ %) من عمليات التصنيف الصحيح، بينما حقق معامل Kappa (٢٤.٩ %) من عمليات التصنيف الصحيح، أما معامل (F.1Score) فقد حقق (٩٥.٤ %) من عمليات التصنيف الصحيح، وكلما اقتربت قيم هذه المعاملات من الواحد الصحيح كلما دلت على جودة بناء الخوارزمية.

ويوضح الجدول رقم (١٩) التالي الأهمية النسبية لأوزان المتغيرات المستقلة الناتجة من خوارزمية الإنحدار اللوجستي، وذلك كما يأتي:

جدول رقم (١٩)الأهمية النسبية لأوزان المتغيرات المستقلة (%) الناتجة من خوارزمية (LR)

المتغير	X1	X2	X3	X4	X8	X15	X16	X19	X21
الأهمية النسبية	١٠.٥٩	١٢.٣٧	٠.٠٠	١٠٠	٥٣.١٩	٥٣.٨٤	٥٣.٣٦	٧٨.٠٣	٩٥.٤٤
ترتيب الأهمية	٨	٧	٩	١	٦	٤	٥	٣	٢

المصدر: نتائج التحليل الإحصائي.

وكما هو موضح بالجدول رقم (١٩) السابق نجد أن أقوي متغير يفسر المتغير التابع (التعثر المالي للشركة) هو المتغير (X4)، ويليه المتغير (X21)، كما نلاحظ أن أقل متغير يفسر المتغير التابع هو المتغير (X1)، ولا يوجد أي تأثير للمتغير (X3).

٦- نتائج الدراسة:

توصلت الدراسة إلى مجموعة من النتائج تمثلت فيما يأتي:

١) أن أساليب التنقيب في البيانات (الشبكة العصبية الإصطناعية، والغابات العشوائية، والجار الأقرب) كانت جميعها أكثر دقة في عملية التنبؤ بالتعثر المالي للشركات مقارنة بالأساليب التقليدية (الإنحدار اللوجستي).

٢) أن أكثر الأساليب دقة في عملية التنبؤ كانت خوارزمية الشبكة العصبية الإصطناعية، يليها في المرتبة الثانية خوارزمية الغابات العشوائية، والمرتبة الثالثة خوارزمية الجار الأقرب، وأخيراً خوارزمية الإنحدار اللوجستي.

٣) أن أفضل المؤشرات التي تساعد المراجع الخارجي في التنبؤ بدقة بالتعثر المالي للشركات كانت على التوالي كما يلي: (نسبة السيولة السريعة، نسبة الأرباح المحتجزة إلى إجمالي

الأصول، حجم الشركة، نسبة التداول، نسبة الأرباح المحتجزة إلى إجمالي حقوق الملكية، نسبة الأصول المتداولة إلى إجمالي الأصول، معدل النقدية، معدل دوران المخزون، إجمالي الالتزامات إلى إجمالي حقوق الملكية).

٤) أن الأهمية النسبية للمؤشرات التسعة التي تم ذكرها في النقطة رقم (٣) السابقة، اختلفت أهميتها النسبية وفقاً لكل خوارزمية من الخوارزميات التي استخدمت في عملية التنبؤ بالتعثر المالي.

٧- توصيات الدراسة:

في ضوء النتائج التي تم استخلاصها من الدراسة التطبيقية، وسعياً نحو تطبيق الهدف الأساسي للدراسة، فقد جاءت توصيات الدراسة على النحو التالي:

١) ضرورة عقد دورات تدريبية من قبل الهيئات المهنية لتدريب وزيادة خبرة مراجعي الحسابات في استخدام أساليب التنقيب في البيانات في عملية المراجعة، وذلك لتحسين أحكامهم المهنية، بما يساهم في تحسين كفاءة عملية المراجعة الخارجية.

٢) ضرورة استعانة مراجعي الحسابات في مصر بذوي الخبرة التكنولوجية في تطوير نماذج لدعم القرار، مما يزيد من جودة عملية المراجعة بصفة عامة، وجودة حكم المراجع الخارجي بشأن الاستمرارية بصفة خاصة.

٣) ضرورة تشجيع مراجعي الحسابات على استخدام أساليب التنقيب في البيانات في عملية المراجعة، بصورة تجعل هذه الأساليب أداة مساعدة للمراجع الخارجي، بما تمكّن المراجع من الاستفادة منها في تحديد مخاطر المراجعة، أو في تقييم استمرارية الشركة، أو في الحصول على مؤشرات أولية لوجود أو عدم وجود غش في القوائم المالية.

٤) ضرورة التفكير جدياً في صياغة نماذج تنبؤية حديثة تستند على تقنيات التنقيب في البيانات، بحيث تكشف عن التعثر المالي حسب طبيعة كل قطاع من القطاعات الاقتصادية داخل الدولة.

٥) ضرورة تدريس مادة عن المراجعة الالكترونية في أقسام المحاسبة داخل الجامعات المصرية، حتى يتسنى لخريجي هذه الجامعات من اكتساب أساس علمي ونظري مصحوباً بالتطبيق العملي على استخدام هذه التقنيات، بما يساهم من تعظيم الاستفادة من أدوات التنقيب في البيانات في مجال المراجعة.

٦) ضرورة قيام الجهات المعنية بإصدار معايير المراجعة بإلزام المراجع الخارجي بتطوير الأساليب التي يتبعها عند قيامه بعملية المراجعة، وذلك من خلال التركيز على الأساليب الحديثة لمواكبة التطورات التكنولوجية.

قائمة المراجع

أولاً: المراجع باللغة العربية

١. إبراهيم، محمد زيدان؛ أبو السعود، محمد عبدالمنعم. (٢٠١٨). "أثر إدارة التدفقات النقدية على احتمال تعرض الشركات لمخاطر التعثر المالي: دراسة تطبيقية على الشركات المدرجة بالبورصة المصرية"، المؤتمر العلمي الثاني، كلية التجارة – جامعة الأسكندرية.

٢. أبو الخير، أسامة أحمد محمد. (٢٠١٩). "دور إستخدام أساليب التنقيب فى البيانات لتحسين تقديرات مراقب الحسابات فى مدى وجود أخطاء جوهرية بالقوائم المالية: دراسة ميدانية فى بيئة الأعمال المصرية"، مجلة الدراسات التجارية المعاصرة، جامعة كفر الشيخ – كلية التجارة، العدد(٧).

٣. أمين، عصام حمدي مصطفى. (٢٠٢٣). "أثر اعتماد مراقب الحسابات على نموذج هجين من أدوات الذكاء الاصطناعي على تحسين جودة حكمه بشأن الاستمرارية: دراسة تطبيقية على الشركات المقيدة بالبورصة المصرية". مجلة الأسكندرية للبحوث المحاسبية، المجلد (٧)، العدد (١).

٤. بسيوني، عبدالرحيم عوض عبدالخالق. (٢٠٢٢). "تحليل زمن البقاء باستخدام أسلوب الشبكات العصبية الإصطناعية ونموذجي انحدار كوكس والانحدار اللوجستي (دراسة تطبيقية)". المجلة العلمية للتجارة والتمويل، كلية التجارة – جامعة طنطا، المجلد (٤٢)، العدد (٤).

٥. الجبالي، عصام الدين محمد علي. (٢٠٢٠). "أثر حساسية عوائد أسهم المنشأة للتغيرات في بيئتها الخارجية على تعثرها المالي: دراسة تطبيقية على الشركات المقيدة بالبورصة المصرية". مجلة البحوث المالية والتجارية، جامعة بور سعيد – كلية التجارة، العدد (٢).

٦. الجبلي، وليد سمير عبدالعظيم. (٢٠٢٠). "أثر استخدام الإنحدار اللوجستي كأحد أساليب التنقيب في البيانات (Data Mining) في دعم الرأي المهني لمراجعي الحسابات: دراسة تطبيقية"، مجلة الإسكندرية للبحوث المحاسبية، جامعة الإسكندرية – كلية التجارة، المجلد (٤)، العدد (٢).

٧. الجبيلي، راميا. (٢٠٢٣). "التنبؤ بتعثر القروض المصرفية باستخدام التحليل التمييزي وخوارزميات التنقيب في البيانات: دراسة مقارنة بين الأساليب الإحصائية وأساليب الذكاء الإصطناعي في تقييم القروض المصرفية". مجلة جامعة البعث، المجلد (٤٥)، العدد (٤).

٨. خطاب، نهى محمد شحات سليمان؛ علي، سمية أمين. (٢٠١٩). "استخدام الشبكات العصبية الضبابية Fuzzy Neural Networks في التنبؤ بالتعثر المالي للمنشآت"، رسالة ماجستير غير منشورة – كلية التجارة جامعة القاهرة.

٩. الشورى، عمرو إبراهيم عوض؛ الغندور، محمد مصطفى عطية. (٢٠٢٠). "إطار مقترح لتحقيق التكامل بين أسلوب تحليل النظم وتقنيات التنقيب عن البيانات بهدف دعم المراجع الخارجي لاكتشاف الأخطاء الجوهرية بالقوائم المالية والتقرير عن استمرارية المنشأة (دراسة ميدانية)"، المجلة العلمية للدراسات والبحوث المالية والتجارية، جامعة دمياط – كلية التجارة، المجلد (١)، العدد (١).

١٠. الشيخ، عمر الشيخ الطاهر. (٢٠٢٠). "الاتجاهات الحديثة للمراجعة الداخلية ودورها في الحد من التعثر المالي: دراسة ميدانية على عينة من المصاف السودانية"، مجلة كلية التنمية البشرية، العدد (٩) الجزء الثاني.

١١. الصفراني، محمد فرج؛ زايد، عبدالفتاح المختار؛ كشيّم، سعاد الهادي. (٢٠٢٠). "إمكانية استخدام نموذج Kida في التنبؤ بالفشل المالي لشركة الإنماء للاستثمارات المالية القابضة ٢٠١٤-٢٠١٧"، مجلة البحوث الأكاديمية (العلوم التطبيقية)، العدد (١٦).

١٢. عثمان، حسام محمد محمد؛ اسماعيل، طارق محمد حسانين. (٢٠٢٢). "استخدام الشبكات العصبية متعددة الطبقات في التنبؤ بمخاطر الائتمان لمنشآت الأعمال: دراسة تطبيقية"، مجلة الدراسات المالية والتجارية، العدد (١).

١٣. العطار، حسن عبدالحميد؛ إبراهيم، صفاء محمد عبدالسلام؛ فارس، دعاء أحمد سعيد. (٢٠٢٠). "أثر كثافة رأس المال على العلاقة بين الرافعة المالية والتعثر المالي بالتطبيق على الشركات المدرجة بالبورصة المصرية: دراسة أمبريقية". مجلة البحوث التجارية، جامعة الزقازيق – كلية التجارة، المجلد (٤٢)، العدد (١).

١٤. علي، هبه جمال هاشم. (٢٠٢٣). "انعكاسات استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع الخارجي وأثرها على عدالة القوائم المالية: دراسة تطبيقية". المجلة العلمية للدراسات والبحوث المالية والتجارية، كلية التجارة – جامعة دمياط، المجلد (٤)، العدد (٢).

١٥. فرج، أشرف رمضان عبدالجواد؛ الصباغ، أحمد عبدالمولى؛ عوض، محمد حمدي. (٢٠٢٢). "دراسة أثر التقارب مع المعايير الدولية إعداد التقارير المالية على تكلفة رأس المال وانعكاسه على التعثر المالي". المجلة الأكاديمية للبحوث التجارية المعاصرة، المجلد (٢)، العدد (٣).

١٦. محمد، دلال محمد إبراهيم. (٢٠٢١). "التأثير المشترك للتعثر المالي والتجنب الضريبي على كل من القيمة السوقية وعوائد الأسهم: دراسة اختبارية على الشركات دولية النشاط المسجلة في البورصة المصرية"، المجلة العلمية للدراسات والبحوث المالية والتجارية، كلية التجارة، جامعة دمياط، المجلد (٢)، العدد (٢) الجزء الثاني.

١٧. مشابط، نعمة حرب. (٢٠٢٠). "أثر جودة المراجعة الخارجية على سلامة رأي مراقب الحسابات بشأن الاستمرارية – دراسة تطبيقية على الشركات الصناعية المسجلة بالبورصة المصرية"، مجلة الأسكندرية للبحوث المحاسبية، كلية التجارة – جامعة الاسكندرية، المجلد (٤)، العدد (٢).

١٨. معلم، رقية؛ طيار، أحسن. (٢٠١٩). "استخدام نموذج ألتمان للتنبؤ بالتعثر المالي لشركات قطاع التأمين في الجزائر"، مجلة الباحث الاقتصادي، المجلد (٧)، العدد (١١).

١٩. النقودي، سوزي فاروق. (٢٠٢٢). "تقييم دقة أساليب التنقيب في البيانات Data Mining (DM) في الحد من المخاطر الائتمانية وانعكاساتها على جودة القوائم المالية بالقطاع المصرفي – دراسة حالة". مجلة البحوث المالية والتجارية، كلية التجارة – جامعة بور سعيد، المجلد (٢٣)، العدد(٤).

٢٠. هاشم، محمد محمود. (٢٠٢٠). "استخدام الشبكات العصبية الإصطناعية والخوارزميات الجينية الوراثية في تقدير هامش ربح الاكتتاب في شركات تأمينات الممتلكات والمسئولية في السوق المصري (بالتطبيق على تأمين أجسام السفن)"، المجلة العلمية للدراسات التجارية والبيئية، المجلد (١١)، العدد (٣) الجزء الأول.

ثانياً: المراجع باللغة الأجنبية

1. Alareeni, B., & Hamdan, A. (2022). The Impact of Artificial Intelligence on Accounting and Auditing in Light of the COVID-19 Pandemic. In *Artificial Intelligence and COVID Effect on Accounting* (pp. 3-7). Singapore: Springer Nature Singapore.
2. Ashraf, S., GS Félix, E., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress?. *Journal of Risk and Financial Management*, *12*(2), 55.

3. Awad, S. S., & Wathik, I. M. (2022). Using data mining tools to the prediction of going concerns on auditor opinion-empirical study in Iraqi commercial. *Academy of Accounting and Financial Studies Journal*, *26*(3), 1-13.
4. Chen, S. (2019). An effective going concern prediction model for the sustainability of enterprises and capital market development. *Applied Economics*, *51*(31), 3376-3388.
5. Chu, M. K., & Yong, K. O. (2021). Big data analytics for business intelligence in accounting and audit. *Open Journal of Social Sciences*, *9*(9), 42-52..
6. Çığşar, B., & Ünal, D. (2019). Comparison of data mining classification algorithms determining the default risk. *Scientific Programming*, *2019*.
7. Dawalekar, P., Gharat, N., Gaikwad, S., & Kundale, J. (2019). Data Mining Techniques for Smart Fitness. In 2nd International Conference on Advances in Science & Technology (ICAST).
8. Elewa, M. M. (2022). Using Altman Z-Score Models for Predicting Financial Distress for Companies–The Case of Egypt panel data analysis. *Alexandria Journal of Accounting Research*, 6(1).
9. Fredrick. I. & Osazemen. E. (2018).Capital structure and corporate financial distress of manufacturing firms in Nigeria. Journal of Accounting and Taxation, 10(7), 78-84.
10. Hertina, D., & Dari, F. W. (2023). Comparative Analysis of Financial Distress Models in Predicting Bankruptcy during Covid-19 Pandemic. Budapest International Research and Critics Institute-Journal (BIRCI-Journal), 6(1), 180-191.
11. John, A.T., & Ogechukwu, L. (2018)."Corporate Governance and Financial Distress in the Banking Industry: Nigerian Experience". Corporate Governance, 10(1).
12. Kok, D., Caron, S., & Acun, A. (2021). Building a quantum kNN classifier with Qiskit: theoretical gains put to practice.
13. Kuloba, L., & Ombaba, K. B. M. (2019).Determining the effect of information and communication on financial distress of unlisted banks in Kenya. *International Journal of Management IT and Engineering*, *9*(10).
14. Kušter, D. (2023). Construction of bankruptcy prediction model using discriminant analysis and financial ratios. *Ekonomija: teorija i praksa*, *16*(1), 1-22.
15. Ledhem, M. A. (2022). Data mining techniques for predicting the financial performance of Islamic banking in Indonesia. *Journal of Modelling in Management*, *17*(3), 896-915.
16. Mahmoud, W. A., Aborizka, M., & Amer, F. A. E. (2021). Heart Disease Prediction Using Machine Learning and Data Mining Techniques: Application of Framingham Dataset. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(14), 4864-4870.
17. Máté, D., Raza, H., & Ahmad, I. (2023). Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies. *Risks*, *11*(10), 176.
18. Michalkova, L., Adamko, P., & Kovacova, M. (2018). The analysis of causes of business financial distress. In Third International Conference on Economic and Business Management (FEBM 2018) (pp. 49 Atlantis Press.
19. Mohammadi, M., Yazdani, S., Khanmohammadi, M. H., & Maham, K. (2020). Financial reporting fraud detection: An analysis of data mining algorithms. *International Journal of Finance & Managerial Accounting*, *4*(16), 1-12.

20. Pandimurugan, V., Usha, D., Guptha, M. N., & Hema, M. S. (2022). Random forest tree classification algorithm for predicating loan. *Materials Today: Proceedings*, *57*, 2216-2222.
21. Papik, M., & Papikova, L. (2022). Detecting accounting fraud in companies reporting under US GAAP through data mining. *International Journal of Accounting Information Systems*, 45, 100559.
22. Paramartha, P. A., & Wiagustini, N. L. P. (2021). Determination of financial distress in manufacturing companies on the Indonesia Stock Exchange. *International Journal of Management Studies and Social Science Research*, *3*(3).
23. Seliem, M. M. (2022). HandlingOutlier data as missing values by imputation methods: application of machine learning algorithms. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *13*(1), 273-286.
24. Shahwan, T., Fadel, M. (2020). Machine Learning Models and Financial Distress Prediction of Small and Medium-Sized Firms: Evidence from Egypt. Journal of Alexandria University for Administrative Sciences, 57(1).
25. Sharma, A., Patel, M., & Tiwari, M. (2019). A comparative study to detect fraud financial statement using data mining and machine learning algorithms. *International Research Journal of Engineering and Technology*, *6*(8), 1492-1495.
26. Singh, J., & Banerjee, R. (2019). A study on single and multi-layer perceptron neural network. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE.
27. Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, *134*, 93-101.
28. Stankevičienė, J., & Prazdeckaitė, G. (2021). Analysis of the accuracy of bankruptcy prediction models: the case of Lithuanian companies. Apskaitos ir finansų mokslas ir studijos: problemos ir perspektyvos, (1), 44-53.
29. Widhiadnyana, I. K., & Ratnadi, N. M. D. (2019). The impact of managerial ownership, institutional ownership, proportion of independent commissioner, and intellectual capital on financial distress. Journal of Economics, Business & Accountancy Ventura, 21(3).
30. Yang, T., & Nazareth, D. (2022). Genetic Algorithm-based Feature Selection for Auditing Decisions. *Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL)*.