انعكاسات استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع الخارجي وأثرها على عدالة القوائم المالية: دراسة تطبيقية

إعداد

د. هبه جمال هاشم علي

أستاذ مساعد المحاسبة والمراجعة

كلية التجارة بالإسماعيلية، جامعة قناة السويس

drhebagamala@gmail.com

المجلة العلمية للدراسات والبحوث المالية والتجارية

كلية التجارة – جامعة دمياط

المجلد الرابع - العدد الثاني – الجزء الثاني - يوليو ٢٠٢٣

التوثيق المقترح وفقاً لنظام APA:

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، ٤(٢)٢، ١٥٥-٢٠٤.

رابط المجلة: https://cfdj.journals.ekb.eg/

انعكاسات استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع الخارجي وأثرها على عدالة القوائم المالية: دراسة تطبيقية

د. هبه جمال هاشم علي

الملخص:

استهدف البحث دراسة انعكاسات استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع وأثرها علي عدالة القوائم المالية حيث تناول استراتيجيات وأساليب التنقيب عن البيانات من خلال عرض مفهوم التنقيب عن البيانات واستراتيجياتها، وأنواعها، وأساليبها، ثم انعكاس استخدام أساليب التنقيب عن البيانات علي التنبؤ برأي المراجع الخارجي في عدالة القوائم المالية، ويقتصر مجتمع الدراسة في كافة الشركات المتحولة إلى النظم الرقمية في الآونة الأخيرة وبخاصة تلك المسجلة في سوق الأوراق المالية المصري، ومن خلال تحليل المعلومات المتعلقة بالعديد من الشركات الصناعية المسجلة في سوق الأوراق المالية المصري تبين للباحثة تحول العديد من الشركات إلى تطبيق النظم الرقمية حيث بلغ عددهم ٢٥ شركة بإجمالي ١١٢مشاهدة، وتوصل البحث إلي وجود فروق جوهرية بين التنبؤ برأي المراجع باستخدام كلاً من الشبكات العصبية، وشجرة القرارات، ونموذج تعلم الآلة كتقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، بينما أتضح عدم وجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام تقنية الأقرب جار كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية" وذلك نتيجة لتطابق النتائج للتنبؤ مع العينة الفعلية للدراسة حيث أن مستويات الدقة كانت ١٠٠ % ومن ثم فهي لم تقم بالتنبؤ.

الكلمات الدالة: التنقيب عن البيانات، شجرة القرارات، الشبكات العصبية، الأقرب جار، تعلم الآلة، التنبؤ برأي المراجع.

أولاً: المقدمة وطبيعة المشكلة

نتيجة للنمو النوعي والتعقيدات المتزايدة للأنشطة الاقتصادية التي تسببت في أن تلعب القوائم المالية دوراً مهماً في تقييم الشركات، حيث تعد عدالة هذه القوائم المالية ضرورية للمستثمرين والدائنين الذين يعتمدون علي القوائم المالية التي يقدمها مديري الشركات لاتخاذ القرارات الاستثمارية، فأحياناً ينشأ تعارض بين وجهة نظر مقدمي هذه القوائم ومستخدميها الأمر الذى يعتبر الدافع لتنفيذ عملية المراجعة لتعزيز القدرة علي اعتماد القوائم المالية المقدمة من الشركات، ولذلك ينبغي علي المراجع الخارجي القيام بعمله وفقاً لمعايير المراجعة المصرية والدولية عند إبداء رأيه المهني بشكل واضح حول صدق وعدالة القوائم المالية ومدي خلوها من التحريفات الجوهرية (Wu, H. P., & Li, L., 2021).

فقد ازداد الطلب علي التنبؤ برأي المراجع حول موثوقية وعدالة القوائم المالية للتخصيص الأمثل للموارد الاقتصادية ومع الاستخدام الواسع لتكنولوجيا المعلومات والتقنيات الحديثة في المعاملات الاقتصادية ظهرت تحديات في طرق المراجعة، ولكي تواكب عملية المراجعة هذه التحديات ينبغي عليها التعامل مع التقنيات الحديثة فمن الممكن أن تمنع هذه التقنيات التلاعب من قبل شركات مختلفة ويمكن أن تساعد المراجع في الاستجابة المستمرة لمطالب بيئة الأعمال ، كما يمكن استخدام المراجع لمخرجات هذه النماذج في تخطيط عملية المراجعة والاستفادة منها في الإجراءات التي يمكن تطبيقها لتحقيق مستوي مقبول من مخاطر المراجعة، وتعمل نماذج اتخاذ القرار أيضاً علي مساعدة المراجع في فحص العديد من الشركات مما يحقق وفورات له في الوقت والجهد والأتعاب (Wu, H. P., & Li, L., 2021).

ولذلك ظهرت أساليب حديثة تساعد المراجع علي كشف الغش والاحتيال في القوائم المالية من خلال مساهمة هذه الأساليب في تقليل إجمالي عدد ساعات عملية المراجعة، بالرغم من أنها تمكنه من اختبار كامل لجميع البيانات التي تم اختيارها بدلاً من اختبار عينة في ظل الأساليب التقليدية لعملية المراجعة، وبالتالي يزداد اعتماد المراجع علي النتائج التي تم التوصل إليها ومن ثم تساهم في تحقيق جودة المراجعة (Yaşar, A., and et al., 2015)، ومن الأساليب الحديثة أسلوب التنقيب عن البيانات الذي يعد أحد الأدوات الهامة والناجحة في تحليل محتوي البيانات في الماضي والمستقبل حيث يعتبر امتداد للوسائل التقليدية ولكن يتميز عنها بقدرته علي تحليل كمية بيانات كبيرة بالإضافة إلي قدرته علي التنبؤ بالمستقبل.

حيث أشارت دراسة (Amani, F. A., & Fadlalla, A. M., 2017) إلى أن التنقيب عن البيانات أحد أهم النماذج الحالية لتحليلات الأعمال الذكية المتقدمة وأدوات دعم القرار، حيث تم الاعتراف بها من قبل الهيئات المهنية المحاسبية الكبري فالمعهد الأمريكي (AICPA) اعتبرها واحد من أفضل عشر تقنيات للمستقبل، وقد أدرج معهد المراجعين الداخليين (IIA) أن استخراج البيانات يعتبر أحد أولويات البحث الأربعة، بالإضافة إلي (CGMA) أن ٥٠ ٪ من إدارة الشركات يصنفون البيانات الضخمة واستخراج البيانات من بين الأعلى عشر أولويات مؤسسية أساسية لعصر الأعمال المبني على البيانات.

ويعرف التنقيب عن البيانات بأنه عبارة عن عملية تحليل مجموعة كبيرة من البيانات بهدف إيجاد قواعد ونماذج يمكن أن تستخدم في مساعدة متخذي القرارات بالإضافة إلي قدرته علي التنبؤ (Papík, M., & Papíková, L., 2022)، وتتلخص خطوات عملية التنقيب عن البيانات في فهم طبيعة الأعمال من خلال فهم المشاكل التي تواجه الأعمال ثم العمل علي وجود صيغة محددة لأهداف الأعمال، ثم تأتي خطوة فهم البيانات عن طريق القيام بجمع البيانات وتوصيف البيانات والتأكد من جودتها وعمل التحليل الاسترشادي لذلك، وبعد ذلك خطوة صياغة نماذج الحل وثبوتها ثم خطوة التقييم وتعديل نتائج النموذج ثم خطوة نشر وتوزيع النموذج.

ويوجد عدة أساليب لتقنية التنقيب عن البيانات **ومن أهمها أسلوب الشبكات العصبية الصناعية** وهو عبارة عن تقنيات حسابية مصممة تشبه طريقة عمل العقل البشري لأداء عملية معينة من خلال عمليات معالجة ضخمة موزعة بالتوازي وهذه العمليات مكونة من وحدات معالجة بسيطة تسمي العقد وتعبر هذه الوحدات عن عناصر حسابية لها خاصية عصبية، ويقوم هذا الأسلوب بتخزين المعرفة عند اكتسابها وتخزن المعلومات الحالية أيضاً ولذلك هو يشبه العقل البشري لأنها تكتسب المعرفة بالتدريب (Salehi, M., and et al., 2016)، بينما يقوم **أسلوب شجرة القرارات** بعرض جميع الكائنات بشكل شجرة ثم يتم تصنيفها ثم تقوم بالبحث عن المعلومات المفيدة علي التوالي، وفي هذا الأسلوب يتم استخدام الخرائط بشكل كبير، ويتسم هذا الأسلوب بأنه طريقة سريعة وتتصف بالتصنيف البسيط لذلك تكون قابلة للتطبيق علي معالجة البيانات واسعة النطاق ولكن ينبغي عند بناء شجرة القرارات أن يؤخذ بعين الاعتبار أن تكون قابلة للتطبيق بقدر الإمكان علي كل البيانات المتوفرة (Ortmeier, C., and et al., 2021).

ونظراً لقيام تقنية التنقيب عن البيانات باختبار شامل للنظم والرقابة للتأكد من أن الشركة يمكنها إنتاج قوائم مالية عادلة وصادقة، ونتيجة لزيادة المعلومات المحاسبية من خلال الإنترنت زادت أهمية هذه التقنية في تأكيد قدرة المستخدمين لاسترجاع المعلومات المحاسبية بمستوي عالي من الدقة، حيث تعتمد تقنية التنقيب عن البيانات علي نظم المعلومات المحاسبية للاستفادة من تجميع بيانات المعاملات المالية بكفاءة ومن ثم توفر معلومات لمتخذي القرار، كما توفر تقدير للرقابة الداخلية وبالتالي يؤدي التنقيب عن البيانات إلي زيادة القدرة علي تقدير الأداء المالي وهذا يسهل مهمة التنبؤ والتوقع بالمتغيرات المالية (Gao, B., 2022).

كما يمكن استخدام تقنيات التنقيب عن البيانات في اكتشاف التلاعب بالممارسات والتحريفات والغش المحاسبي، عن طريق أدوات معالجة البيانات والتحقق من جودة وسلامة البيانات وهذا بدوره يؤدي إلي تصحيح التحريفات، **وتعتقد الباحثة أنه إذا تم استخدام تقنيات التنقيب عن البيانات سوف يؤثر بشكل إيجابي علي التنبؤ برأي المراجع في عدالة القوائم المالية، ومن ثم تتمثل مشكلة البحث في الإجابة على التساؤل التالي:**

- هل توجد فروق بين أراء مراقبي الحسابات في عدالة القوائم المالية قبل استخدام التقنيات المختلفة للتنقيب عن البيانات وبعد استخدامها؟

ثانياً: أهمية البحث

في ضوء مشكلة البحث تتمثل أهميته العلمية في تقديم منهجية جديدة في البحث المحاسبي تساعد على التنبؤ برأي المراجع الخارجي بمدي عدالة وصدق القوائم المالية من خلال استخدام تقنيات التنقيب عن البيانات كما يوفر دليل من البيئة المصرية على هذه العلاقة، كما تتمثل أهميته من أنه يتناول موضوع هام في مجال التنبؤ بعدالة القوائم المالية وأنها خالية من أي تحريفات باستخدام تقنيات التنقيب عن البيانات ومدي رأي المراجع في ذلك. **بينما يستمد البحث أهميته العملية** من استخدام التقنيات الحديثة كالتنقيب عن البيانات في التنبؤ برأي المراجع الخارجي في عدالة القوائم المالية وهذا الأمر سيكون له تداعياته التي تزيد من ثقة المستثمرين ومن ثم مساعدتهم في اتخاذ قرارات استثمارية وزيادة كفاءة سوق المال أيضاً مما يعني زيادة الفرص الاستثمارية الجديدة.

ثالثاً: أهداف البحث

يتمثل الهدف الرئيسي لهذا البحث في دراسة دور تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع الخارجي في عدالة القوائم المالية، ويمكن تحقيق هذا الهدف من خلال تحقيق الأهداف التالية:

١- دراسة أهمية استراتيجيات التنقيب عن البيانات في التأكد من عدالة القوائم مالية.
٢- توضيح أثر التنقيب عن البيانات على التنبؤ برأي المراجع في عدالة القوائم المالية.
٣- إجراء دراسة تطبيقية على الشركات المقيدة بالبورصة المصرية لمعرفة مدي تأثير التنقيب عن البيانات على التنبؤ برأي المراجع الخارجي.

رابعاً: فروض البحث

في ضوء مشكلة البحث وأهدافه وأهميته، تتمثل فروضه على النحو التالي:

الفرض الأول: توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام شجرة القرارات كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية.

الفرض الثاني: توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام الشبكات العصبية كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية.

الفرض الثالث: توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام نموذج تعلم الآلة كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية.

الفرض الرابع: توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام تقنية الأقرب جار كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية.

خامساً: حدود البحث

يقتصر هذا البحث على انعكاسات استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع وأثرها على عدالة القوائم المالية، ويقتصر مجتمع الدراسة في كافة الشركات المتحولة إلى النظم الرقمية في الآونة الأخيرة ولا سيما تلك المسجلة في سوق الأوراق المالية المصري، ومن خلال تحليل المعلومات المتعلقة بالعديد من الشركات الصناعية المسجلة في سوق الأوراق المالية المصري تبين للباحثة تحول العديد من الشركات إلى تطبيق النظم الرقمية حيث بلغ عددهم ٢٥ شركة بإجمالي ١١٢مشاهدة.

سادساً: خطة البحث

يتم تناول هذا البحث من خلال ثلاثة أقسام، حيث يستعرض **القسم الأول** الدراسات السابقة التي تناولت المتغيرات البحثية وهي تقنيات التنقيب عن البيانات والتنبؤ برأي المراجع الخارجي في عدالة القوائم المالية لاستنتاج الفجوة البحثية، بينما يتناول القسم الثاني **الدراسة النظرية** والتي تشتمل علي استراتيجيات التنقيب عن البيانات، وعرض مفهوم التنقيب عن البيانات واستراتيجياتها، وأنواعها، وأساليبها، ثم إبراز أثر دور أساليب التنقيب عن البيانات علي التنبؤ برأي المراجع الخارجي في عدالة القوائم المالية، ويتناول القسم الثالث تصميم الدراسة التطبيقية واختبار الفروض البحثية.

القسم الأول: الدراسات السابقة

يتمثل الهدف من عرض وتحليل الدراسات السابقة للمتغيرات البحثية محل الدراسة وهما تقنيات التنقيب عن البيانات والتنبؤ برأي المراجع الخارجي في عدالة القوائم المالية في معرفة ما تم التوصل إليه من نتائج وتوصيات للاستفادة منها في استنتاج الفجوة البحثية العلمية في مجالي التنقيب عن البيانات والتنبؤ برأي المراجع الخارجي في عدالة وصدق القوائم المالية، حيث أجريت **دراسة (Amani, F. A., & Fadlalla, A. M., 2017)** فحص منهجي للبحوث المنشورة حول تطبيقات التنقيب عن البيانات في المحاسبة لفهم الوضع الحالي، كما قدمت إطار تنظيمي يعمل علي توفير رؤية شاملة لما تم إنجازه باستخدام تطبيقات التنقيب عن البيانات وما هي المجالات المحاسبية الأكثر استخدام لهذه التطبيقات وما هي الأقل في الاستخدام، حيث اعتمدت الدراسة في التفسير علي تحليل المحتوي لفهم الأدبيات ذات الصلة، وتوصلت الدراسة إلي وجود فوائد لتطبيق التنقيب عن البيانات في الموضوعات الرئيسية في المحاسبة مثل حوكمة الشركات والتقارير المتكاملة والتعليم المحاسبي فقد يفيد نظام المعلومات المحاسبي بشكل كبير من دمج قدرات استخراج البيانات وتحسين دقة التسجيل وبالتالي دقة التقارير ومن ثم تساعد تقنيات التنقيب عن البيانات بفوائد كبيرة علي اتخاذ القرار في المحاسبة.

واستعرضت **دراسة (Kopun, D., 2018)** الأدبيات السابقة في مجال الكشف عن الغش في البيانات المالية باستخدام طرق التنقيب عن البيانات التي تنفيذها في الخمسة عشر عاماً الماضية حيث أكدت الدراسة علي أن نسب التحليل المالي هي مؤشرات قوية في كشف الغش في البيانات المالية، وقد أظهرت نتائج الدراسة أن الغش في البيانات المالية هي نوع من الغش الذي يتسبب في خسائر كبيرة للشركات ونتيجة لحالات الغش في الخمسة عشر عام الماضية كان هناك اهتمام متزايد من قبل المجتمع والأكاديميون، كما أوضحت الدراسة أن لطرق التنقيب عن البيانات فاعلية في التنبؤ وكشف الغش في البيانات المالية.

واستهدفت **دراسة (Stanisic, N., Radojevic, T., & Stanic, N., 2019)** التغلب علي القيود المنهجية المحددة للدراسات السابقة التي تهدف إلي التنبؤ بنوع رأي المراجع واستخلاص استنتاجات محددة بشأن الأداء التنبئي النسبي والطرق التنبؤية المختلفة لهذا الغرض، واستخدمت الدراسة الأداء التنبئي لاثني عشر نموذجاً مرشحاً من مجالات الإحصاء كما تم تقييم نماذج التعلم الآلي بشكل منفصل وذلك بالتطبيق علي السيناريوهات الآتية أولهما: عند وجود معلومات مسبقة عن العميل مثل أنواع رأي المراجع الذي تم تلقيه في الماضي حيث يمكن استخدامه في التنبؤ، وثانيهما: عندما لا تكون هذه المعلومات متاحة أي في حالات الشركات الجديدة، وتضم الدراسة ١٣٤٥٦زوجاً من البيانات المالية السنوية وتقارير المراجعة المقابلة، وتوصلت الدراسة في السيناريو الأول عدة طرق من كليهما تحقق العوالم أداء تنبؤياً مشابهاً يبلغ حوالي ٨٩٪ مثل المنطقة الواقعة تحت المنحني AUC ، أما في السيناريو الثاني حققت خوارزميات التعلم الآلي وشجرة القرارات أداء أفضل بشكل ملحوظ، وأوصت الدراسة بالسماح للمحاسبين والمراجعين في جميع أنحاء العالم بالتطوير واختبار النماذج التنبؤية التي تساعدهم في إجراءات تخطيط المراجعة وتقييم المخاطر.

كما استهدفت **دراسة (Sánchez-Serrano, J. R., 2020)** توفير نموذج جديد للتنبؤ برأي المراجع حول التقارير المالية، وتمثلت عينة الدراسة في ٢٩٨ شركة من الشركات العاملة في السوق الإسبانية وتم تقسيم العينة إلى مجموعات الرأي غير المؤهلة ومجموعات الرأي المؤهلة كما تم تطبيق تقنية الشبكات العصبية الاصطناعية علي المعلومات التي تم الحصول عليها من العينة، وأظهرت الدراسة أن تقنية الشبكات العصبية تنبأت برأي المراجعة بدقة ٨٦٪ وهذه نسبة كبيرة لتوفير صورة عن مدي عدالة القوائم المالية.

وناقشت **دراسة (Werner, M., Wiese, M., & Maas, A. 2021)** تقنيات التنقيب عن البيانات كأساليب تحليل بيانات جديدة وتم التركيز علي ممارسات المراجعة حيث يتيح التنقيب عن البيانات إمكانية تحليل العمليات التجارية بطريقة آلية، كما بحثت الدراسة في كيفية دمج تقنيات التنقيب عن البيانات في عمليات المراجعة المعاصرة من خلال معايير المراجعة ذات الصلة وذلك من خلال دراسة ميدانية، وتوصلت الدراسة إلي أن تقنيات التنقيب عن البيانات وفقاً لمعايير المراجعة المعاصرة وممارسات المراجعة تزيد من موثوقية استنتاجات المراجعة ويحسن من متانة أدلة المراجعة عن طريق استبدال إجراءات المراجعة التقليدية كما أنها توفر الوسائل اللازمة لمواكبة تطورات التكنولوجيا والتحديات الحالية.

كما تناولت **دراسة (Saglar, J., & Ilker, K. E. F. E., 2021)** طرق التنقيب عن البيانات المستخدمة في أنشطة المراجعة كما ناقشت طرق التنقيب عن البيانات في الكشف عن الاحتيال وركزت الدراسة على الشبكات العصبية الاصطناعية وشجرة القرارات والانحدار اللوجيستي وآلية ناقلات الدعم والخوارزميات الجينية، وتوصلت الدراسة إلي أن تقنيات التنقيب عن البيانات تساعد المراجع الخارجي في أداء مهام المراجعة بكفاءة عالية وتكلفة وجهد أقل، حيث أنها يمكن استخدامها في إجراءات تخطيط المراجعة وتقييم المخاطر.

فيما قامت **دراسة (Zeng, J., & Yang, Y., 2021)** بمعرفة إمكانية التنبؤ برأي المراجع في تحسين تخصيص الموارد في سوق الأوراق المالية والحفاظ علي ترتيب اقتصاد السوق كما حللت هذه الدراسة نتائج المؤشرات المالية المؤثرة في نتائج المراجعة من عدة اتجاهات وهي المؤشرات المالية وبيئة السوق والقيمة النسبية للسوق، حيث تم اختيار القطاع المالي للفترة من ٢٠١٧ إلي ٢٠١٩ كعينة للدراسة، كما استخدمت الدراسة الشبكات العصبية لتحديد المخاطر المالية والتنبؤ برأي المراجع، وتوصلت الدراسة إلي أنه يمكن للمديرين استخدام نموذج التنبؤ برأي المراجع لتقييم حالة التشغيل المرحلية للشركة مما يساعد في تخصيص الموارد في سوق الأوراق المالية.

وقدمت **دراسة (خميس،٢٠٢٢)** إطار مقترح لدور تقنية التنقيب عن العمليات في دعم منهجية ستة سيجما وكيفية دمجها في المراحل الخمس لنموذج DMAIC من أجل تحقيق استراتيجية التحسين المستمر لكايزن وتم ذلك من خلال دراسة حالة علي إحدى الشركات المساهمة المصرية بهدف الإجابة علي تساؤل الدراسة وتقديم نموذج عملي يوضح كيفية دعم تقنية التنقيب عن العمليات لمنهجية ستة سيجما ودمجها في المراحل الخمس لنموذج DMAIC لتحقيق استراتيجية كايزن للتحسين المستمر للعمليات، وتوصلت الدراسة إلي أن تطبيق تقنية التنقيب عن العمليات لها قدرة فائقة علي استخراج المعلومات الخاصة بالعمليات لتوفير رؤية تفصيلية لسوك العمليات التي تم تنفيذها، مما يجعلها أداة قوية داعمة للمراحل المختلفة لنموذج

DMAIC لتحليل التعقد الحقيقي لخرائط تدفق العمليات، كما توصلت إلي أن الإطار المقترح يقوم علي أن تقنية التنقيب عن العمليات لا تحل محل منهجية ستة سيجما وإنما هي تقنية حديثة وأداة فعالة قادرة علي دعم أساليب وأدوات هذه المنهجية من خلال قدرتها علي التمكن من فهم العملية بشكل سريع ونمذجة العمليات بما يساهم في زيادة كفاءة وفعالية استراتيجية كايزن في تحقيق التحسين المستمر للعمليات.

وساهمت **دراسة (Papík, M., & Papíková, L., 2022)** في أدبيات التنقيب عن البيانات الخاصة بكشف الغش في تقارير بموجب مبادئ المحاسبة المقبولة عموماً GAAP كما هدفت إلي تطوير التنقيب عن البيانات كمنهج يحقق أداء نموذج أعلي من الموجود في الأدبيات السابقة، وتم استخدام البيانات المالية من نماذج K- 10 لعدد٣٢٠ بياناً مالياً احتيالياً من ٨٠ شركة احتيالية و١٢٠٠ بياناً مالياً احتيالياً من ٢٤٠ شركة غير احتيالية من الولايات المتحدة، وقد أتضح أن تقنيات التنقيب عن البيانات هي الأكثر شيوعاً تم تطبيقها في تطوير نماذج التنبؤ، وأظهرت نتائج الدراسة أن تقنية التنقيب عن البيانات تساهم في تحسين أداء النموذج للشركات التي تقدم تقارير بموجب مبادئ المحاسبة المقبولة عموماً في الولايات المتحدة.

وركزت **دراسة (Zeng, S., Li, Y., & Li, Y., 2022)** الضوء علي عملية التنبؤ برأي المراجع ودورها في الوقاية من مخاطر سوق الأمن عن طريق أساليب التعلم الآلي، وذلك لتحسين جودة المراجعة وتقليل مخاطر المراجعة وبالتالي رفع كفاءة المراجعة، حيث اشتملت عينة دراسة علي جميع شركات التصنيع المدرجة في بورصة الصين من عام ٢٠١٢ الي ٢٠١٩ التنبؤ، وأشارت نتائج الدراسة إلي أن التنبؤ برأي المراجع يحل بشكل فعال التناقض بين فعالية المراجعة ومخاطر المراجعة مما يجعل شركات المراجعة المدرجة أكثر قيمة، كما أوضحت أن تطبيق آليات التعلم الآلي تعطي إنذار مبكر لمخاطر المراجعة وبالتالي يساعد ذلك أصحاب المصلحة في التنبؤ برأي المراجع حول التقارير المالية.

أما **دراسة (Zdolsek, D., Jagric, T., & Kolar, I. 2022)** قدمت نموذج للتنبؤ برأي المراجع في استمرارية نشاط الشركة من خلال استخدام مجموعات مختلفة من المتنبئين الاقتصاديين للشركة، حيث يقوم المراجع بإصدار رأيه بشأن استمرارية الشركة في تقريره وفقاً لمدي كفاية إفصاحات الشركة بخصوص استمرار نشاطها في التقرير السنوي للشركة المفصح عنه، وذلك بالتطبيق علي الشركات بسلوفينيا خلال الفترة ٢٠٠٥- ٢٠١٣وبلغت عينة الدراسة ١٤٧٦١ملاحظة، وأوضحت نتائج الدراسة أن الشركات ذات المؤهلات المستمرة لديها هيكل مالي سئ أي تمويل أقل لمعدلات الأسهم مقارنةً بالشركات غير المؤهلة، كما أشارت إلي أن استخدام نموذج الانحدار اللوجيستي للتنبؤ برأي المراجع بشأن استمرارية الشركة في تقريره يُمكن من التنبؤ بمؤهلات الشركة بدقة كافية.

بينما هدفت **دراسة (Awad, S. S., 2022)** إلي اختبار مجموعة من المتغيرات المالية علي عدد من المتغيرات الخاصة للبنوك التجارية في العراق التي أثبتت حقيقة الوضع المالي لديهم وتعثر بعضهم بغرض تأكيد المراجع في تقريره علي قدرة الشركة علي الاستمرار، كما استخدمت الباحثين شجرة القرار كأحد تقنيات التنقيب عن البيانات وأساليب أخري للبحث في بيانات ١٣بنك منهم ٧ متعثر مالياً حيث تعتبر هذه الدراسة بمثابة المحاولة الأولي لاختبار هذه التقنية في بيئة البنوك العراقية، وأظهرت نتائج الدراسة أن تقنية شجرة القرارات حققت تطابق في تنبؤها بنسبة ١٠٠٪ مع الواقع الفعلي.

التعليق على الدراسات السابقة

في ضوء الدراسات السابقة التي تناول معظمها العلاقة بين تقنيات التنقيب عن البيانات واكتشاف الغش والاحتيال المالي كدراسة دراسة (Kopun, D., 2018) التي استعرضت الأدبيات السابقة في مجال الكشف عن الغش في البيانات المالية باستخدام طرق التنقيب عن البيانات التي تنفيذها في الخمسة عشر عاماً وأوضحت الدراسة أن لطرق التنقيب عن البيانات فاعلية في التنبؤ وكشف الغش في البيانات المالية، كما بحثت دراسة (Papík, M., & Papíková, L., 2022) في أدبيات التنقيب عن البيانات الخاصة بكشف الغش في تقارير بموجب مبادئ المحاسبة المقبولة عموماً GAAP، وأظهرت نتائج الدراسة أن تقنية التنقيب عن البيانات تساهم في تحسين أداء النموذج للشركات التي تقدم تقارير بموجب مبادئ المحاسبة المقبولة عموماً في الولايات المتحدة.

فيما تناولت بعض الدراسات التنبؤ برأي المراجع ومنها دراسة (Zeng, J., & Yang, Y., 2021) التي تناولت إمكانية التنبؤ برأي المراجع في تحسين تخصيص الموارد في سوق الأوراق المالية والحفاظ علي ترتيب اقتصاد السوق، وتوصلت إلي أنه يمكن للمديرين استخدام نموذج التنبؤ برأي المراجع لتقييم حالة التشغيل المرحلية للشركة مما يساعد في تخصيص الموارد في سوق الأوراق المالية، كما أوضحت دراسة (Zeng, S., Li, Y., & Li, Y., 2022) دور عملية التنبؤ برأي المراجع في الوقاية من مخاطر سوق الأمن عن طريق أساليب التعلم الآلي، وذلك لتحسين جودة المراجعة وتقليل مخاطر المراجعة وبالتالي رفع كفاءة المراجعة، كما أوضحت أن تطبيق آليات التعلم الآلي تعطي إنذار مبكر لمخاطر المراجعة وبالتالي يساعد ذلك أصحاب المصلحة في التنبؤ برأي المراجع حول التقارير المالية.

في حين تناولت دراسات أخري علاقة تقنيات التنقيب عن البيانات بمهام المراجعة كدراسة **(Saglar, J., & Ilker, K. E. F. E., 2021)** التي توصلت إلي أن تقنيات التنقيب عن البيانات تساعد المراجع الخارجي في أداء مهام المراجعة بكفاءة عالية وتكلفة وجهد أقل، حيث أنها يمكن استخدامها في إجراءات تخطيط المراجعة وتقييم المخاطر، كما أظهرت دراسة (Werner, M., Wiese, M., & Maas, A. 2021) أن تقنيات التنقيب عن البيانات وفقاً لمعايير المراجعة المعاصرة وممارسات المراجعة تزيد من موثوقية استنتاجات المراجعة ويحسن من متانة أدلة المراجعة عن طريق استبدال إجراءات المراجعة التقليدية كما أنها توفر الوسائل اللازمة لمواكبة تطورات التكنولوجيا والتحديات الحالية.

وأخيراً أظهرت دراسة (Sánchez-Serrano, J. R., 2020) أن تقنية الشبكات العصبية تنبأت برأي المراجعة بدقة ٨٦٪ وهذه نسبة كبيرة لتوفير صورة عن مدي عدالة القوائم المالية، وقامت دراسة (Sánchez-Serrano, J. R., and et al., 2020) بتوفير نموذج جديد للتنبؤ برأي المراجع حول التقارير المالية وتوصلت إلى أن تقنية الشبكات العصبية تنبأت برأي المراجعة بدقة ٨٦٪ وهذه نسبة كبيرة لتوفير صورة عن مدي عدالة القوائم المالية.

ويتضح مما سبق أن مراجعة البيانات المالية عملية معقدة ومتخصصة للغاية ولذلك تخلق الرقمنة المتزايدة لمعالجة المعاملات تحديات جديدة للمراجعين في أداء مهامهم ولذلك تتمثل الفجوة البحثية في إمكانية استخدام تقنيات التنقيب عن البيانات لتحليل البيانات الجديدة لتحسين عملية مراجعة البيانات المالية والتغلب على قيود المراجعة التقليدية ومن ثم تساعد تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع الخارجي عن مدي عدالة القوائم المالية.

القسم الثاني: الدراسة النظرية

أولاً: استراتيجيات وأساليب التنقيب عن البيانات في ضوء الأدبيات المحاسبية

يعتبر التنقيب عن البيانات أداة متطورة لتحليل البيانات كما تسمح بتطبيق فكر جديد لتنظيم البيانات التي لها استخدامات عملية، حيث يعرف التنقيب عن البيانات بأسلوب اكتشاف المعرفة أو استخراج المعلومات فهي كتقنية تسعي لاستنتاج المعرفة من كميات هائلة من البيانات كما أنها تعتمد علي الخوارزميات الرياضية والتي تعتبر أساس التنقيب عن البيانات فهي مستمدة من العديد من العلوم كعلم الرياضيات، والإحصاء، وعلم المنطق، والذكاء الاصطناعي، وعلم الآلة، فهذه التقنية تستطيع أن تجيب علي أسئلة في الماضي وأسئلة في الحاضر وفي المستقبل (عبيد الله، ٢٠١٦).

فعادةً ما يقصد بالتنقيب عن البيانات الأساليب اللازمة لجمع المعرفة الضمنية وغير المعروفة حيث يمكن اعتباره شكلاً من أشكال اكتشاف المعرفة اللازمة لحل العديد من المشاكل في نطاق معين، كما يعرف بعملية تحليل كامل لكم هائل من البيانات حيث يتم فرزها لإيجاد معلومات مفيدة لتحسين عملية صنع القرار، ولذلك يعد التنقيب عن البيانات نظام معلومات يعتمد على الحاسب الآلي يقوم بفحص كميات هائلة من البيانات لتوليد المعلومات واكتشاف معرفة أعمق لاكتشاف روابط جديدة بين البيانات الضخمة (Dahiya, N., and et al., 2021).

وتقوم هذه التقنية باكتشاف نماذج للبيانات ويعمل علي هيكلة قواعد الارتباط وينتج عن ذلك مخرجات لتقنية التنقيب عن البيانات تكون داعماً قوياً لمتخذي القرار، كما تعرف بأنها عملية يتم استخدامها للتحليل الكامل للبيانات وخاصةً البيانات المالية المتاحة للشركة لاختيار معلومات تفيد المستخدم في تحديد العلاقات بين البيانات مما يحسن المعلومات المتاحة لمتخذي القرار (Albashrawi, M., 2016) ، ومن الممكن تعريفها علي 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(Papík, M., & Papíková, L., 2022).

ويتضح مما سبق أن تقنية التنقيب عن البيانات هي العملية التي يتم فيها استخدام الذكاء الاصطناعي والرياضي والإحصائي بالإضافة إلى تقنيات التعلم الآلي لاستخراج المعلومات المفيدة واكتساب المعرفة من قواعد البيانات الكبيرة، فهي تعتبر عملية تحليلية للبحث في بيانات ضخمة وهائلة لاستخراج أنماط مفيدة وإيجاد علاقات بين هذه الأنماط ومدي الارتباط بينهما. كما تتلخص خطوات عملية التنقيب عن البيانات في **فهم طبيعة الأعمال** من خلال فهم المشاكل التي تواجه الأعمال ثم العمل على وجود صيغة محددة لأهداف الأعمال، ثم تأتي **خطوة فهم البيانات** عن طريق القيام بجمع البيانات وتوصيف البيانات والتأكد من جودتها وعمل التحليل الاسترشادي لذلك، وبعد ذلك **خطوة صياغة نماذج الحل وثبوتها ثم خطوة التقييم وتعديل نتائج النموذج ثم خطوة نشر وتوزيع النموذج**.

حيث يتمثل الهدف من التنقيب عن البيانات في التنبؤ والتصنيف أو المزيج بينهما حيث يهدف التنبؤ إلي التركيز علي قوة التوقع، كما يمكن **استخدام التنقيب عن البيانات في حل أربع مشكلات ، تتمثل المشكلة الأولى في اكتشاف العلاقة** حيث يقوم التنقيب عن البيانات بتحليل الارتباط وهو من أشهر التطبيقات ويعد هذا التحليل واضح في توجيه المبيعات، ففي حالة كانت منتجات المبيعات هي نفس

المنتجات التي يرغب العملاء في شرائها فمن المتوقع أن ترتفع المبيعات (Gaurab, T., 2015)، والمشكلة الثانية هي **اتخاذ القرارات** حيث يمثل اتخاذ القرار اختيار طريقة واحدة من بين عدة طرق وهذا ما يقوم به صانعي القرار، ومن خلال استخدام استراتيجيات التنقيب عن البيانات مثل التصنيف يمكن عن طريقها تقييم البيانات من أجل اتخاذ القرار وطرق التفضيل واقتراح أفضل الطرق، وتتمثل المشكلة الثالثة في **التنبؤ** من خلال الاختيار من بين عدة طرق حول المستقبل وهذا ما يستخدم في التنبؤ بتوقعات العميل في المستقبل واحتياجات السوق المستقبلية، وأخيراً **مشكلة تحسين إجراءات العمل** حيث يخضع العمل في نظام معقد من مقدمي العمل وضغط السوق ومن خلال استخدام التنقيب عن البيانات يمكن تسليط الضوء علي الأبعاد المتعددة للإجراءات التي تتسم بالكفاءة (Kopun, D., 2018) .

فمن المزايا التي يوفرها التنقيب عن البيانات المتمثلة في مساعدة اتخاذ القرار حيث يساعد متخذ القرار في تفعيل الترابط بين الأقسام المختلفة كما يساعد في الاستخدام الأمثل لمصادر البيانات المتاحة وفي التخطيط من خلال تحسين وتطوير نظم المعلومات المحاسبية المستخدمة، ويساعد أيضاً في معرفة قدرة الشركة على النمو ومواكبة التطور فبالتالي يوفر التنقيب عن البيانات معلومات تساعد علي زيادة المعرفة والحد من البدائل ومن ثم تخفيض حالة عدم التأكد التي تتمثل في هذه البدائل (Changpetch, P., & Reid, M., 2021).

وعلى ذلك تشتمل منهجية التنقيب عن البيانات علي خمس مراحل وهما (أخذ العينة، والاستكشاف، والتعديل، والنمذجة، والتقييم)، كما يوجد أربع استراتيجيات للتنقيب عن البيانات وهما (التنبؤ، والتصنيف، وتحليل التجميع، واكتشاف العلاقات) ، وتعبر **استراتيجية التنبؤ** عن طرق منبثقة من التنبؤات الإحصائية التقليدية للتنبؤ بقيمة متغير معين، حيث يتضمن التنبؤ استخدام بعض المتغيرات في قاعدة البيانات للتنبؤ ولذلك يركز التنبؤ علي إيجاد قيمة رقمية مستقبلية أو قيمة غير رقمية، بينما تعد **استراتيجية التصنيف** هي الأكثر استخدام في التنقيب عن البيانات فهذه الاستراتيجية تعمل علي تمييز فئات مختلفة من الإجراءات أو الأشياء، حيث يمكن من خلال هذه الاستراتيجية تصنيف المعاملات المالية إلي عادية أو بها احتيال وينتج عن ذلك حماية مبالغ كبيرة من الأموال، كما يمكن استخدام هذه الاستراتيجية في اكتشاف وجود ميزة خاصة غير موجودة في منتجات المنافسين، ويتضح من ذلك أن التصنيف يركز علي تعيين البيانات لمجموعة صفات نوعية منفصلة محددة مسبقاً من الفئات والتي يمكن أن تكون ثنائية أو متعددة الطبقات (Gepp, A., and et al., 2018).

وتشبه **استراتيجية تحليل التجميع** استراتيجية التصنيف حيث يستخدم التجميع لتوليد عناوين للتصنيفات وذلك بهدف تكوين مجموعات متشابهة في شكل نماذج ويتم تحديد هذه المجموعات عن طريق برنامج وهذا ما يحدث في المواقع الإلكترونية التي تتيح عمليات البحث، كما تستخدم استراتيجيات تحليل التجميع لتقسيم الأشياء في مجموعات لها معني، وتتسم هذه الاستراتيجية بالقوة حيث يمكن استخدامها في تجميع المنتجات إلي مجموعات لها نفس المواصفات، ويتضح من ذلك أن هذه الاستراتيجية تستخدم للتعرف علي المعلومات القابلة للمقارنة أي التي تشبه بعضها البعض، أما **استراتيجية اكتشاف العلاقات** للتنقيب عن البيانات تعمل علي اكتشاف البيانات التي ترتبط مع بعضها البعض والتي لا يمكن رؤيتها بوضوح إلا من خلال هذه الاستراتيجية، كما تركز هذه الاستراتيجية علي إيجاد البيانات التي تنحرف بشكل كبير عن المعتاد (Dos Santos Garcia, C., and et al., 2019).

وتقوم تقنية التنقيب عن البيانات علي عدة خطوات أساسية ينبغي تطبيقها بعناية من أجل تحقيق نتائج فعالة من التحليلات التي تمت حيث تتمثل الخطوة الأولي من التنقيب عن البيانات في تحديد مشكلة العمل بوضوح وبعد ذلك يتم تجميع البيانات وإعدادها للنمذجة وتأخذ هذه الخطوة وقتاً طويلاً وتعتبر خطوة أساسية من معالجة البيانات(Amani, F. A., & Fadlalla, A. M., 2017)، ثم يتم اختيار أسلوب التنقيب عن البيانات الأنسب للمشكلة المطروحة وتطبيقه وبعد ذلك يتم تقسيم البيانات بشكل عام إلي مجموعات تدريب وتوثيق ويتم استخدام تدريب البيانات لاشتقاق القواعد والصيغ كما يتم استخدام تحقق البيانات للتأكد من مدي تعامل القواعد التي تم إعدادها علي مجموعة مختلفة من البيانات اعتماداً علي نتائج تطبيق النموذج، ثم يتم تقييم أداء النموذج أو تعديله أو نشره لحل مشكلة العمل الأمر الذى يفيد في التوصل لحالات الغش في القوائم المالية من خلال تحليل العمليات التجارية التي تتم وما حدث عليها من تغييرات لاحقة (Chen, S., 2016).

فيما **تنقسم تقنية التنقيب عن البيانات** إلى نوعين **أولهما: التنقيب الوصفي** وهو عبارة عن وصف وتصنيف للبيانات المتاحة حسب تواجدها والعلاقة بينهما عن طريق التفاعل الطبيعي لكي يتم توضيح هذه البيانات، كما يعتمد على إعادة تنظيم البيانات والبحث في داخلها لاستخراج النماذج الموجودة بها ومن أهم أدوات التنقيب الوصفي التصنيف، قاعدة الربط، التحليل التسلسلي، العنقدة. **وثانيهما: التنقيب التنبؤي** ويقصد به إيجاد أفضل التنبؤات من خلال المعطيات كمعرفة المنتج الأفضل لعميل معين فهو يقوم باستخدام المعلومات القديمة للتنبؤ بما يحدث في المستقبل ويتضمن الأساليب التالية كالتصنيف والانحدار وتحليل السلاسل الزمنية (Popov, A. A., and et al., 2020).

فبالنسبة لأدوات التنقيب الوصفي يقوم **التصنيف** علي حل المشاكل عن طريق تحليل مجموعة من البيانات ووضعها علي شكل أقسام يمكن استخدامها بعد عملية التصنيف ومن الأساليب التي يمكن استخدامها في تصنيف البيانات كالخوارزميات الإحصائية والشبكات العصبية، بينما تعد **قاعدة الربط** أحدي أدوات التنقيب عن البيانات فهي تستخدم كأداة من أدوات اكتشاف المعرفة من قواعد البيانات فلديها القدرة علي تصفح كميات هائلة من البيانات كما يمكن اعتبارها قواعد ارتباطية معينة بين مجموعة من البيانات في قاعدة البيانات، أما **التحليل التسلسلي** الذي يتم من خلاله البحث لاكتشاف نماذج تحدث بالتسلسل فالمدخلات تتمثل في بيانات تشكل مجموعة متسلسلة وتعد كل سلسلة من البيانات قائمة منظمة من العمليات وعندما تمثل العملية مجموعة من المصطلحات يلزم أن يحسب معها الوقت المصاحب لكل عملية ولكن تعتبر مشكلة هذا النموذج هي إيجاد كل النماذج المتسلسلة مع أقل دعم يخصصه المستفيد عندما يكون الدعم لهذا النموذج يعادل نسبة تسلسل البيانات التي يحتويها النموذج، وأخيراً **عملية العنقدة** التي تقوم بتقسيم البيانات إلي مجموعة من الأصناف وفقاً لخصائصها المتشابهة فهي تقسيم غير موجه للبيانات يختلف عن التصنيف فهي تساعد علي فهم التركيب الطبيعي للمجموعات من البيانات (Dutta, I., and et al., 2017).

بينما تحتوي أدوات التنقيب التنبؤي على **التصنيف** الذي يقوم بتحديد الأقسام قبل فحص البيانات ويتم تصنيف البيانات وفقاً للخصائص المشتركة بينهما، و**الانحدار** الذي يسمح بتحليل لوصف العلاقة بين متغيرين أو أكثر حيث يقوم بوضع البيانات بنوع معروف من الدوال وبعد ذلك يفترض أفضل دالة للبيانات المعطاة، **وتحليل السلاسل الزمنية** الذي يتم من خلاله مشاهدة البيانات عبر الزمن ويترتب على ذلك تحليل مفيد وذلك لأنه يتم من خلال مشاهدة سلوك البيانات عبر الزمن تكون أساسية (Popov, A. A., and et al., 2020).

التقنيات المستخدمة في التنقيب عن البيانات

تحتوي تقنية التنقيب عن البيانات على عدة أساليب ومنها الشبكات العصبية، وشجرة القرارات، شبكة دعم التمييز، أقرب جار، الخوارزمية الجينية، والتي يمكن عرضها على النحو التالي:

١- **الشبكات العصبية** (ANN)

تعتبر نوع من الذكاء الاصطناعي فهي عبارة عن نموذج رياضي قائم علي أساس عدد كبير من عناصر المعالجة شديدة الترابط تسمي الخلايا العصبية حيث تستخدم هذه التقنية التفكير المنطقي، فهي مفيدة للتعرف علي الأنماط في البيانات خاصةً عندما يكون شكل العلاقات بين المتغيرات التابعة والمستقلة معقد أو غير معروف تم تصميمها علي نفس سياق الدماغ البشري، كما تستخدم الشبكات العصبية في نمذجة الوظائف التي لها تعبير رياضية غير معروفة فهي تتكون من عدد من المعالجات المستقلة والبسيطة للخلايا العصبية، كما تقوم بالتعرف علي الأنماط لمحاكاة الطريقة التي يعالج بها العقل البشري المعلومات وتخزينها، حيث يمكن استخدامها كتقنيات استخراج للبيانات من أجل التصنيف وتم تطويرها بطريقة منظمة تُمكن من التعلم والتعميم من خبرة البيانات من خلال توزيع هذه الخلايا العصبية في طبقات هرمية قليلة وعادة ما تشتمل علي ثلاثة أنواع من الطبقات تتمثل في طبقة المدخلات وطبقة خفية وطبقة المخرجات(Wu, H. P., & Li, L., 2021)، فبعد تلقي المدخلات من جميع الخلايا العصبية من طبقة الإدخال يتم إضافة القيم عن طريق الأوزان التطبيقية وتحويلها إلي قيمة الإخراج من خلال تطبيق وظيفة التنشيط، ومن ثم يتم تمرير النتيجة إلي جميع الخلايا العصبية في الطبقة التالية وبذلك توفر مسار للتغذية من الأمام لطبقة الإخراج (Salehi, M., and et al., 2016).

وتري الباحثة أن الشبكات العصبية تمتلك تنظيماً ذاتياً جيداً وقابل للتكييف حيث تستخدم بشكل أساسي في معالجة البيانات غير الخطية بدرجة عالية من خلال متغيرات التأثير التفاعلي وتنقسم إلي نموذج تغذية أمامية ونموذج التغذية الراجعة للشبكة العصبية وطريقة رسم الخرائط ذاتية التنظيم، كما أنها تكون الشبكات العصبية قادرة علي تعلم خصائص القوائم المالية ومن ثم تستطيع معرفة إذا كانت احتيالية عن طريق مقارنة المعلومات الجديدة بالبيانات المخزنة واكتشاف الأنماط المخفية داخل مجموعة البيانات الكبيرة، وبعد ذلك تقوم الشبكات العصبية بتقييم إشارات البيانات الفردية لإنشاء نمط سلوك متميز يصنف بيانات الإدخال على أنها احتيالية أو غير احتيالية ثم تقوم بتطبيق النمط الناتج للكشف عن وجود غش في القوائم المالية.

٢- **شجرة القرارات** (DT)

تعد تقنية شجرة القرارات كأحد أساليب التنقيب عن البيانات بل وتمثل التقنية الشائعة للتصنيف والتنبؤ فهي مناسبة لاكتشاف المعرفة حيث تتكون من هيكل شجري يشبه المخطط الانسيابي ولا يكون طرفياً وتمثل العقد اختبارات علي صفة واحدة أو أكثر وتعكس العقد الطرفية نتائج القرار من السهل فهمها، حيث تقابل كل عقدة في شجرة القرارات مجموعة من السجلات من مجموعة البيانات الأصلية وتسمي العقدة العليا بالجذر وتقوم هذه العقدة بوصف جميع الصفوف في مجموعة البيانات المحددة فهي العقدة التي يحدث فيها الانقسام الرئيسي، وتعرف العقد باسم العقد الداخلية وتسمي العقد المتبقية بالعقد الطرفية وتدل علي فئة القرار، وبتقسيم الصفوف في العقدة إلى عقدتين فرعيتين، ويعتبر اختيار السمة التي تفصل العينة بشكل

أفضل مصدر قلق كبير في بناء شجرة القرار حيث يتم تقسيم كل عقدة إلى عقد فرعية إلى أن تكون المجموعات الفرعية صغيرة جدًا ويترتب علي هذا التقسيم المتتالي للعينة شجرة كبيرة وقد تعكس بعض الفروع حالات شاذة في شكل القيم المتطرفة أو القيم الخاطئة، ويتم إزالة هذه الفروع بطريقة لا تؤثر على معدل دقة النموذج (Ortmeier, C., and et al., 2021).

وتتمثل المزايا الرئيسية لشجرة القرارات في أنها توفر طريقة مفيدة لتمثيل المعرفة المكتسبة ويكون من السهل تكوينها من خلال قواعد التصنيف، ويعد الغرض من أسلوب شجرة القرارات هو تصنيف البيانات وفقاً للقيم المنفصلة باستخدام عدة متغيرات تنبؤية، وتعتبر عقد شجرة القرارات نقاط اختبار متغيرات التنبؤ وبناءً على نتيجة الاختبار في العقدة يتم تقسيم عقد شجرة القرارات إلى مزيد من نقاط القرار بمستويات أقل(Dahiya, N., and et al., 2021) .

وتري الباحثة أن شجرة القرارات تقوم بعرض جميع الكائنات بشكل شجرة ثم يتم تصنيفها ثم تقوم بالبحث عن المعلومات المفيدة على التوالي، وفي هذا الأسلوب يتم استخدام الخرائط بشكل كبير، ويتسم هذا الأسلوب بأنه طريقة سريعة وتتصف بالتصنيف البسيط لذلك تكون قابلة للتطبيق على معالجة البيانات واسعة النطاق، ولكن ينبغي عند بناء شجرة القرارات أن يؤخذ بعين الاعتبار أن تكون قابلة للتطبيق بقدر الإمكان على كل البيانات المتوفرة.

٣- **آلية نقل الدعم** (SVM)

ومن ناحية أخري يوجد **أسلوب آلية نقل الدعم** (SVM) فهي طرق تعلم خاضعة للإشراف تستخدم للتصنيف حيث يتم استخدام هذا الأسلوب في تصنيف مجموعات العينات إلي فئتين منفصلتين أو أكثر، وهذا الأسلوب تم تقديمه من خلال نظرية التعلم الإحصائي، ويعمل هذا الأسلوب علي زيادة الأداء في التعرف علي الأنماط ثم التنبؤ بالسلاسل الزمنية المالية والتسويق وتقدير إنتاجية التصنيع، كما ينتج عنه مصنفاً ثنائياً ويسمي بالطبقات الفائقة الفاصلة المثلي عن طريق رسم خرائط غير خطية للغاية لموجهات الإدخال، ثم يبني نموذج خطي لتقدير وظيفة القرار باستخدام حدود الفئة غير الخطية علي أساس دعم التمييز، وأخيراً يقوم بتدريب الشبكات الخطية للحصول علي المستوي الأمثل الذي يفصل البيانات بدون خطأ (Salehi, M., and et al., 2016).

٤- **خوارزمية أقرب جار**(KNN)

أما أسلوب **خوارزمية أقرب جار**(KNN) يحاول القيام بتصنيف عينة غير معروفة اعتماداً علي التصنيف المعروف لجيرانها بافتراض أن مجموعة من العينات ذات التصنيف المعروف تشبه بشكل بديهي تصنيف كل عينة تكون متشابهة للعينات المحيطة بها إذا كان تصنيف العينة غير معروف، وبالتالي يمكن التنبؤ به من خلال مقارنة السجلات الشبيهة بالسجل المراد التنبؤ له وتقدير القيمة المجهولة لهذا السجل وفقاً لمعلومات تلك السجلات، فبشكل عام يمكن أن تكون قاعدة التصنيف هذه ضعيفة لأنها تستند إلي قاعدة واحدة في عينة معروفة، ويمكن أن يكون دقيقاً إذا كانت العينة غير المعروفة محاطة بعدة عينات معروفة لها نفس التصنيف (Salehi, M., and at al., 2016).

٥- **أسلوب تعلم الآلة**

بينما يعد **أسلوب تعلم الآلة** أحد فروع الذكاء الاصطناعي الذي يُمكن الحاسب الآلي من أداء مهام تتصف بالذكاء حيث يقوم هذا الأسلوب بعمليات معقدة من خلال التعلم من البيانات بدلاً من اتباع القواعد المبرمجة مسبقاً وذلك عن طريق استخدام البيانات التاريخية كمدخلات والتنبؤ بالنتائج غير المبرمجة بشكل صحيح، كما يستخدم هذا الأسلوب **خوارزميات خاضعة للإشراف (التعلم الآلي خاضع للإشراف)** في حالة كانت تصنيفات البيانات محددة يتعلم النظام التصنيف حسب النموذج الذي يقدمه المستخدم، أو **يستخدم خوارزميات غير خاضعة للإشراف (التعلم الآلي غير الخاضع للإشراف)** في حالة أن يكون لدي المستخدم أمثلة فقط ولكن ليس لديه تصنيفات للبيانات سابقاً وبذلك ينبغي علي الخوارزميات الاكتشاف بنفسها بنية البيانات المخفية بشكل أو بأخر لتستخرج منها التصنيفات، ويستخدم أيضاً **خوارزميات شبه خاضعة للإشراف (التعلم الآلي شبه خاضع للإشراف)** وذلك عبارة عن خوارزميات تتعلم السلوك من خلال الملاحظة ثم التكييف وتعلم النتيجة من بيئتها وتحسين خطواتها المستقبلية بشكل مستمر (Ding, K., and et al., 2020).

ويتضح مما سبق أن التعلم الآلي يعمل علي تحسين عملية تعلم الحاسب الآلي اعتماداً على تجاربه بدون أن تتم برمجته فعلياً أو بدون مساعدة بشرية عن طريق بناء نماذج تعلم الآلة باستخدام البيانات والخوارزميات المختلفة التي تعتمد على نوع البيانات ونوع المهمة التي يمكن تنفيذها بشكل تلقائي. كما يعمل أسلوب التعلم الآلي على تغيير الصناعة المالية حيث يتم استبدال ممارسات إدخال البيانات المعتادة وإدارة الفواتير ومهام مسك الدفاتر منخفضة المستوي بتعلم الآلة وهذا بدوره يؤدي إلى تقليل فرصة الإدخال غير الصحيح للمعلومات المحاسبية وتقليل عبء العمل على المحاسبين (Hunt, J. O., and et al., 2021).

٦- **الخوارزمية الجينية**

يعتمد **أسلوب الخوارزمية الجينية** بشكل أساسي علي نظرية الانتقاء الطبيعي البيولوجي، حيث يتبع هذا الأسلوب الأسس التطورية للعلوم الطبيعية والوراثية فهي تنص علي أن المجموعة الوراثية لمجتمع معين ربما تتضمن حل لمشكلة معينة حيث يتم تمثيل لحل كل مشكلة معينة من خلال عدة خطوات ففي الخطوة الأولي يتم في التوليد العشوائي لمجموعة من الحلول التي يطلق عليها عدد العناصر وتعمل هذه الخوارزمية علي افتراض أن العناصر الجديدة ستكون أفضل من العناصر القديمة بمعني أنها تقوم بإنشاء مجموعة جديدة من المجموعة الحالية، ويتم تطبيق العوامل الوراثية كالاختيار والتهيئة والطفرة والتقاطع علي مجموعة الحلول للحصول علي حل متطور من أجل إيجاد أفضل حل ممكن ومن ثم يتم تطوير حلول جديدة وتستمر عملية التطوير هذه حتي يتم استيفاء شروط معينة (Dutta, I., and et al., 2017) .

٧- **الشبكة البايزية**

يعرف **أسلوب الشبكة البايزية** باسم الشبكات السببية أو شبكات الاحتمالات البيانية فهي تقوم بتمثيل مجموعات فرعية من الصفات، حيث تمثل برسم بياني موجه وتمثل كل عقدة في الرسم خاصية ويمثل كل سهم احتمالاً مستقلاً.

يتضح من خلال عرض أساليب التنقيب عن البيانات أن الخوارزمية الجينية واحدة من أفضل التقنيات للتعامل مع القيم المفقودة في البيانات ولكنها ذات تكلفة مرتفعة، بينما يعتبر أسلوب شجرة القرارات معتدل التكلفة كما توفر شجرة القرارات وسيلة مهمة لتمثيل المعرفة المكتسبة، أو يمكن استخدام نموذج تعلم الآلة الذي يعمل علي استبدال ممارسات إدخال البيانات المعتادة وإدارة الفواتير ومهام مسك الدفاتر منخفضة المستوي بتعلم الآلة وهذا بدوره يؤدي إلي تقليل فرصة الإدخال غير الصحيح للمعلومات المحاسبية، كما تعد الشبكات العصبية تمتلك تنظيماً ذاتياً جيداً وقابل للتكييف، وبناءً علي ذلك يمكن استخدام تقنيات التنقيب عن البيانات في المساعدة بالتنبؤ برأي المراجع والتحقق من عدالة القوائم المالية.

ثانياً: تحليل وتوضيح العلاقة بين تقنيات التنقيب عن البيانات وتنبؤات رأى مراقب الحسابات في عدالة القوائم المالية

يتضمن تقرير المراجعة رأي المراجع الخارجي الفني المحايد والمستقل حول مدي مصداقية وعدالة القوائم المالية، ومن الممكن أن يمتنع المراجع الخارجي عن إبداء الرأي في الحالات التي يتعذر فيها الإفصاح عن هذا الرأي مع ذكر أسباب ذلك، فبشكل عام يتوقف تحديد نوع رأي المراجع الخارجي حول نوع الأخطاء التي يواجهها خلال قيامه بعمله حيث يوجد أربعة أنواع لرأي المراجع عند إعداد تقريره وهما: **رأي نظيف** ويقصد به الرأي الذي لا يتضمن تحفظات، **ورأي متحفظ** وهو الرأي الذي يحتوي علي بعض التحفظات، **والامتناع عن الرأي** ويقصد به عدم إبداء رأي في القوائم المالية، **والرأي السلبي** وهو الرأي الذي يفصح علي أن القوائم المالية لا تعبر بصورة صحيحة وصادقة عن المركز المالي ونتائج أعمال الشركة (Abad, D., and et al., 2017).

ونظراً للدور الهام الذي يلعبه المراجع الخارجي في دعم الشفافية والموثوقية بالقوائم المالية وذلك لأنه يوفر التحقق المستقل من البيانات المالية مما يُمكنه من اكتشاف أوجه القصور في نظام المعلومات المحاسبية ومن ثم كتابة ذلك في تقريره (Zeng, S., and et al., 2022)، كما يقوم بالتحقق من أن القوائم المالية تتوافق مع المعايير المحاسبية، وبعد ذلك يتم الحصول علي أحد نوعي لرأي المراجع سواء رأي نظيف أو مقيد، ففي حالة الحصول علي رأي نظيف يؤكد علي عدم وجود اختلافات بين الأرقام المحاسبية التي تقدمها إدارة الشركة وتلك التي قام المراجع بالتحقق منها، أما في حالة الحصول علي رأي متحفظ يدل علي وجود فروق بين أرقام إدارة الشركة والأرقام التي قام المراجع بالوصول إليها (Yaşar, A., and et al., 2015).

ونتيجة لكثرة التحديات الإقليمية والعالمية التي تواجهها الشركات وما ترتب علي ذلك من ضرورة التركيز علي أهمية الإدراك الصحيح لهذه التحديات وكيفية التعامل معها ومدي الاستجابة لمعطيات هذه التحديات عن طريق تبني التقنيات الحديثة لمواجهة هذه التحديات ومن ثم تستطيع الشركات الحفاظ علي استمرارها وتطورها وهذا بدوره يزيد من حجم الضغوط علي عاتق المراجع الخارجي وبالتالي يؤثر علي رأيه في عدالة القوائم المالية، ويعتبر تقرير المراجع الخارجي هو الناتج لعملية المراجعة ويقوم المراجع بكتابته بعد استيفاءه لكافة إجراءات المراجعة والتحقق من أنه استوفي أهداف المراجعة وأن لديه تأكيد بمدي عدالة القوائم المالية، وبعد ذلك يقوم بكتابة التقرير لكي يوصل المعلومات في نطاق مهام المراجعة التي قام بها ونتائجها كما ينبغي علي المراجعة عند كتابة التقرير أن يشير إلي درجة التأكد في عدالة القوائم

المالية (Salehi, M., & Dehnavi, N., 2018)، وذلك وفقاً لما ورد في معيار المراجعة المصري رقم ٢٠٠ في الفقرة ٢١ بأن يقوم المراجع الخارجي بالحصول علي أدلة المراجعة وتقييمها للحصول علي درجة التأكد المناسبة بخصوص ما إذا كانت القوائم المالية التي تمت مراجعتها تتسم بالعدالة والتعبير بصدق عن حقيقة الوضع المالي للشركة وما إذا كانت خالية من الأخطاء الجوهرية، حيث تعرف درجة التأكد المناسبة باحتمالية وجود خطر يتمثل في إبداي المراجع الخارجي لرأي غير ملائم عند وجود تحريفات في القوائم المالية وهذا يسمي بخطر المراجعة.

فلقد أصبح اكتشاف الغش والاحتيال عنصراً مهماً في عمليات المراجعة للبيانات المالية ولذلك يمثل الكشف عن الغش مصدر قلق للمستثمرين وأصحاب المصالح، فبنك أوف أمريكا علي سبيل المثال يوافق علي دفع ١٦,٥ مليار دولار مقابل حل قضية الغش المالي، ويصنف الغش الداخلي إلي فئتين وهما الغش في البيانات المالية والغش في المعاملات ويعرف الغش في البيانات المالية بأنه تحريف متعمد لقيم مالية معينة لتعزيز الربحية ومن ثم خداع المساهمين بينما يقصد بالغش في المعاملات بعملية انتزاع الأصول التنظيمية ، ففي عام ٢٠٠٠ أشارت لجنة المراجعة التابعة لمجلس الرقابة العامة AICPA إلي فاعلية مجموعة متنوعة من الإصلاحات الضرورية لضمان استمرارية المراجعة علي المدي الطويل، حيث يتطلب القسم ٢٠٠ من بيان معايير المراجعة رقم ١٢٢/١٢٣ أن يحصل المراجعين الخارجيين علي تأكيد معقول حول ما إذا كانت القوائم المالية خالية من الأخطاء الجوهرية سواء كانت ناتجة عن الغش والاحتيال أو الخطأ ومن خلال اعتماد أساليب التنقيب عن البيانات يمكن للمراجعين توفير ذلك التأكيد (Gepp, A., and et al., 2018).

كما أصدر مجلس معايير المراجعة ASB لمعيار SAS 99 بعنوان النظر في الاحتيال والغش في مراجعة البيانات المالية، حيث يشير هذا المعيار إلي أن التحريف الجوهري في البيانات المالية يمكن أن ينتج من تقارير مالية احتيالية والتي يتم تعريفه بتحريفات متعمدة أو حذف للمبالغ أو الإفصاحات في البيانات المالية المصممة لخداع مستخدمي البيانات المالية، ويتطلب هذا المعيار من المراجع أن يقوم بمجموعة متنوعة من المهام التحليلية والتخطيطية وإجراءات المراجعة الموضوعية لدعم اكتشاف الأخطاء الناشئة من التقارير المالية الاحتيالية، وبالتالي نتج عن هذا المعيار زيادة مسئولية المراجعين عن معالجة الغش والاحتيال كجزء لا يتجزأ من عمليات المراجعة المالية (Papík, M., & Papíková, L., 2022).

كما تم تطوير مثلث الغش الذي ورد بهذا المعيار بواسطة (Cressey, D.R., 1973) مما وسع استخدامه من قبل المتخصصين كنموذج نظري لشرح سبب حدوث عمليات الغش، حيث يفترض ذلك المثلث احتمالية أن يحدث الغش بسبب توافر عنصر واحد أو أكثر من العناصر الثلاثة وهما الدافع والفرص والضغوط ولذلك يعد فهم مثلث الغش أمراً ضرورياً لتقييم الغش المالي نظراً لاستناد مخاطر الغش التي وردت في المعايير علي سبيل المثال SAS 99 ، ISA 240 و ASA 240 علي عناصر هذا المثلث، وتعد الشبكات العصبية وشجرة القرارات وسيلة حاسمة لجمع المعرفة الضمنية وغير المعروفة والمحتملة مما يساعد علي كشف الغش بالقوائم المالية (Lin, C. C., and et al., 2015).

ونظراً لاستخدام عملاء المراجعة معلومات الأنظمة المعقدة التي تولد كميات هائلة من الأدلة الالكترونية وبالتالي يمثل العثور علي دليل يساعد في كشف الغش في داخل بيئة تلك المعلومات تحدياً كبيراً، فعند الاعتماد علي أساليب التنقيب عن البيانات لكشف الغش ينبغي التعرف علي نماذج الغش وتعريف مؤشراتها التي تشير إلي وجودها وبعد تحديد تلك المؤشرات التي تدل علي وجود غش في البيانات المالي يمكن ربط أساليب التنقيب عن البيانات التي تؤدي في النهاية إلي إنشاء نموذج يكتشف الغش المحتمل في البيانات المالية، وبالتالي يترتب علي استخدام تقنيات التنقيب عن البيانات القدرة علي تحسين كفاءة وفاعلية فرق المراجعة في إجراء مهام المراجعة المتعلقة بالغش(Kopun, D., 2018) .

ومن خلال استخدام التنقيب عن البيانات في الكشف عن الغش والاحتيال بالقوائم المالية يمكن الوصول إلي الاتجاهات والعلاقات بين البيانات المحاسبية التي تكشف عن الغش، ولذلك تساعد أساليب التنقيب عن البيانات كأداة تحليلية متطورة المراجع الخارجي في التأكد من مدي سلامة البيانات (Albashrawi, M., 2016)، وكشف الغش والاحتيال المتواجد بالقوائم المالية، والتحقق من نظام الرقابة الداخلية، ومن ثم تمكنه من الحصول علي المعلومات بكفاءة وانتاجية عالية ويترتب علي ذلك إنجاز عملية المراجعة وتحليل البيانات بكفاءة وفاعلية مما يوفر الوقت بالإضافة إلي المحافظة علي أعلي درجات الجودة والدقة في البيانات، كما تساعده في المساهمة في كفاءة المراجعة الشاملة وذلك لقدرة التنقيب عن البيانات علي تغير دور أنظمة التحكم الإدارية وتنفيذها بشكل جذري في الشركات، وأخيراً تساعد المراجع الخارجي في معرفة العلاقات بين البيانات التي قد تستغرق وقت كبير لاكتشافها بدلاً عن استخدام التقنيات القديمة.

كما تساعد أساليب التنقيب عن البيانات المراجع الخارجي في استرجاع البيانات وتحليل البيانات وإعداد التقارير بواسطة برامج قياسية التي توفر خاصية التعديل على البيانات عند الحاجة مما يوفر من جهد المراجع، وتساعد أيضاً في وضع خطة المراجعة بشكل مبكر مما يساهم في تحليل البيانات الأولية في وقت مبكر، كما تتيح العمل علي إعادة إجراء العمليات الحسابية للتأكد من صحة المعاملات ودقة النظم المحاسبية بالإضافة إلي تحديد العمليات المتكررة والمفقودة مما يكشف عمليات التلاعب والاحتيال إذا وجدت (ICAI, 2016).

وأشارت أحدي الدراسات إلى أنه نظراً لقدرة تقنية التنقيب عن البيانات علي إجراء اختبار شامل للنظم والرقابة للتأكد من أن الشركة يمكنها إنتاج قوائم مالية عادلة وصادقة، ونتيجة لزيادة المعلومات المحاسبية من خلال الإنترنت زادت أهمية هذه التقنية في تأكيد قدرة المستخدمين لاسترجاع المعلومات المحاسبية بمستوي عالي من الدقة، حيث تعمل تقنية التنقيب عن البيانات علي نظم المعلومات المحاسبية للاستفادة من تجميع بيانات المعاملات المالية بكفاءة وبالتالي توفر معلومات لمتخذي القرار، كما توفر تقدير للرقابة الداخلية ومن ثم يؤدي التنقيب عن البيانات إلي زيادة القدرة علي تقدير الأداء المالي وهذا يسهل مهمة التنبؤ بالمتغيرات المالية (Lin, C. C., and et al., 2015) ، كما أوضحت دراسة (Amani, F. A., & Fadlalla, A. M., 2017) أنه يمكن استخدام تقنيات التنقيب عن البيانات في مجال المحاسبة من خلال استخدام نماذج للكشف عن أقسام معينة في مجالات المحاسبة وذلك باكتشاف التلاعب بالممارسات والتحريفات والغش المحاسبي، عن طريق أدوات معالجة البيانات والتحقق من جودة وسلامة البيانات وهذا بدوره يؤدي إلي تصحيح التحريفات .

ولذلك ازداد الطلب علي المراجعة للتأكد من موثوقية التقارير المالية ومن ثم التخصيص الأمثل للموارد الاقتصادية وباستخدام واسع النطاق لتكنولوجيا المعلومات مثل التنقيب عن البيانات في المراجعة يساعد ذلك في منع ممارسات الغش من قبل الشركات، حيث يستخدم المراجعين هذه التقنية لمراقبة الجودة في المرحلة النهائية أو مرحلة المراجعة للمهمة ولإجراء تحليلات للطوارئ حول كيفية القيام بتغييرات يمكن أن تزيد أو تنقص من احتمال الحصول علي رأي متحفظ، ولذلك تعتبر نماذج اتخاذ القرار هذه تساعد المراجعين في تقديم رأي خاص بمدي عدالة القوائم المالية.

حيث تتمثل الخطوات التي يعتمد عليها مراقب الحسابات عند استخدام تقنيات التنقيب عن البيانات في عملية المراجعة علي **اختبار البيانات** من خلال قيام المراجع الخارجي بإنشاء المخرجات التي تعالج بطلب من العميل، وبعد ذلك تتأتي خطوة **اختبار المنشأة المتكامل** حيث تقوم تقنية التنقيب عن البيانات في هذه الخطوة بمقارنة البيانات المتوقعة مع البيانات الناتجة من المحاكاة، وتعتمد **المحاكاة المتوازية** علي قيام التنقيب عن البيانات التي قام العميل بمعالجتها ثم تُقارن بالبيانات الفعلية التي قام المراجع باستخدامها، ثم تأتي **خطوة مراقبة عملية المراجعة** وفيها يقوم المراجع برصد المعاملات والتأكد من استيفائها لمعايير معينة، ومن ثم يتم أخذ العينات التي تساعد بصورة إحصائية عشوائية وبالتالي تكون العينات أكثر دقة مما يوفر وقت وجهد المراجع الخارجي، ويأتي بعد ذلك **إدارة الملفات** التي يتم فيها تجميع الملفات ومقارنتها وإدارتها بواسطة برامج إدارة الملفات المحوسبة والمقبولة وإنجاز التعديلات علي التقارير بشكل أسهل، وأخيراً تأتي **خطوة إعداد التقارير** بعد التأكد من سلامة كل الملفات لكي يتمكن المراجع من إصدار تقارير موثوقة وعلي درجة عالية من العدالة والنزاهة (Oyedokun, P., & Emmanuel, G. 2015).

ويقوم التنقيب عن البيانات في عملية المراجعة بتجميع بيانات المراجعة الأصلية المطلوبة للاستخدام في قاعدة بيانات نظام المراجعة وبعد ذلك يتم تنظيف البيانات وتصنيفها قبل عملية التنقيب عن البيانات وذلك للتخلص من المعلومات غير المفيدة وإزالة البيانات غير المرتبطة أيضاً، وبعد ذلك يقوم التنقيب عن البيانات بتحليل لقاعدة البيانات من خلال اختيار أنسب أسلوب للتنقيب عن البيانات ثم يقوم بمعالجتها بواسطة المراجعين والخبراء ثم إضافتها إلي قاعدة المعرفة، وعند معالجة بيانات المراجعة المضافة حديثًا ينبغي معالجة البيانات السابقة باستخدام طريقة المعالجة السابقة للبيانات ثم يتم إجراء تحليل مراجعة البيانات باستخدام قاعدة المعرفة المحدثة كمعايير للحكم فإذا تم العثور على بيانات مشكوك فيها يتم تحليلها ومعالجتها من قبل المراجعين والخبراء، ثم يتم إضافتها إلى قاعدة المعرفة.

ونظراً لقدرة التنقيب عن البيانات علي التصنيف والتنبؤ واكتشاف العلاقات بين البيانات يمكن استخدامها في الإجراءات التحليلية والتي يقصد بها وفقا بمعيار المراجعة SAS 56 الصادر عن المعهد الأمريكي للمحاسبين القانونين بأنها النسب والمؤشرات المهمة كنتائج البحث للعلاقات والتقلبات التي تكون متعارضة مع المعلومات الأخرى ذات العلاقة، كما تمثل الإجراءات التحليلية جزء هام من عملية المراجعة فهي تقوم بمقارنات بسيطة واستخدام النماذج المعقدة التي تحتوي علي العديد من العلاقات وعناصر البيانات (Lin, C. C., and et al., 2015) ، ويتضح من ذلك أنها متقاربة مع التنقيب عن البيانات ويترتب علي ذلك سهولة التنبؤ برأي المراجع الخارجي.

كما يمكن استخدام تقنية التنقيب عن البيانات لتفادي مخاطر المراجعة التي تعد من الاتجاهات الحديثة في المراجعة فهي تتطلب فهم المراجع الخارجي للأهداف الاستراتيجية للشركة ونظراً لدور التنقيب عن البيانات في فهم مخاطر المراجعة، وذلك من خلال تمكين المراجعين من فحص العينات بسرعة وذلك لا يحسن من كفاءة أعمال المراجعة فقط بل تقلل ظهور مخاطر المراجعة أيضاً (Zeng, J., & Yang, Y., 2021)، فعند مراجعة البيانات الاقتصادية للشركات يمكن الحصول بسرعة علي العلاقة بين أنواع مختلفة بين البيانات باستخدام طريقة استقراء القاعدة، وفي نفس الوقت يمكن تحليل العلاقة بين قائمة التدفق النقدي وقائمة الدخل وتحليل الميزانية العمومية أيضاً وذلك لمساعدة المراجعين في معرفة بعض الأنشطة الاقتصادية غير المعروفة (Kopun, D., 2018) ، كما تساعد تقنية التنقيب عن البيانات تقييم مخاطر الشركات من خلال استخراج القوائم المالية السابقة من البيانات الجماعية في قاعدة بيانات الشركات لتحليلها باستخدام طريقة تعميم البيانات لتقنية التنقيب عن البيانات حتي يستطيع المراجع تحديد طبيعة المؤسسات المالية وبالتالي تُمكن المراجع الخارجي من إجراء التحليل للقوائم المالية والتأكد من مدي عدالتها (Papík, M., & Papíková, L., 2022).

وتري الباحثة أنه عند استخدام تقنيات التنقيب عن البيانات في فحص العوامل التي تؤثر علي تقارير المراجعة سوف تقوم بعمل نموذج لشرح رأي المراجع في تقارير المراجعة وبالتالي تتمكن بعض أساليب التنقيب عن البيانات مثل الشبكات العصبية في تطوير دور المراجع الخارجي في الكشف عن الغش في القوائم المالية وذلك من خلال تحسين دقة أحكام المراجع في تقييم مخاطر الغش والاحتيال ومن ثم التنبؤ بوجودهم في القوائم المالية أم لا ويترتب علي ذلك سد فجوة توقعات المراجعة حيث أنها تعمل باستخدام أسلوب المحاكاة للوصول إلى أفضل السيناريوهات الممكنة للتنبؤ بشكل رأي المراجع الخارجي في التقارير.

كما تساعد تقنية التنقيب عن البيانات ونموذج تعلم الآلة المراجعون علي فحص العديد من الشركات عن طريق الاستفادة من نماذج وأساليب التنقيب عن البيانات وبالتالي يمكنهم من جذب المزيد من الانتباه إلى الشركات التي لديها احتمالية أعلى لتلقي رأي مراجع نظيف يؤدي إلى تحقيق وفورات في وقت وأتعاب عملية المراجعة، حيث أوضحت دراسة (Zeng, J., & Yang, Y., 2021) نموذج التنبؤ برأي المراجع علي أساس الشبكة العصبية، كما استخدمت دراسة (Sánchez-Serrano, and et al., 2020) الشبكة العصبية لإنشاء نموذج التنبؤ، في حين بحثت دراسة (Jan C. L.., 2018) في استخدام أسلوب ناقلة الدعم في نموذج التنبؤ لعملية المراجعة.

وتري الباحثة أن الاتجاه البحثي لاستخدام تكنولوجيا المعلومات لحل مخاطر المراجعة حيث أن الشبكات العصبية تعمل علي تحديد المخاطر المالية والتنبؤ برأي المراجع من خلال البيانات المالية للشركات، كما يتضح أن المراجعين يستخدمون أساليب التنقيب عن البيانات في تخطيط إجراءات المراجعة، وكأداة لمراقبة الجودة في مرحلة مراجعة التزام الشركات، وباستخدام أسلوب شجرة القرارات يمكن اختبار الرقابة الداخلية حيث يتم تصنيف البيانات اللازمة للمراجعة في الشركات ثم يتم اختيار جزء من البيانات من كل نوع من البيانات للاختبار وعند اختبار الرقابة الداخلية من الممكن اكتشاف مشكلة بنظام الرقابة الداخلية إذا وجدت بيانات منفصلة .

كما استخدمت بعض الدراسات (Zeng, J., & Yang, Y. 2021 ; Dahiya, N., and et al., 2021) شجرة القرارات ونموذج تعلم الآلة كأحد تقنيات التنقيب عن البيانات من أجل التنبؤ بالأزمة المالية، في حين تناولت أحدي الدراسات التنقيب عن البيانات كطرق لتصميم نظام إنذار مبكر للأزمة المالية للمؤسسات الصغيرة والمتوسطة الحجم حيث قامت باختبار النموذج على أكثر من ٧٠٠٠ الشركات الصغيرة والمتوسطة الحجم وتطوير عدد من ملفات تعريف المخاطر، ومؤشرات المخاطر، وأنظمة الإنذار المبكر، وخرائط للطرق المالية التي يمكن استخدامها للتخفيف من المخاطر حيث تعد الشبكات العصبية وشجرة القرارات أدوات قوية لتحليل التفاعل غير الخطي للعلاقات المعقدة ولذلك تكمل الإحصاءات التقليدية (Gepp, A., and et al., 2018 ; Zeng, S., Li, Y., & Li, Y., 2022).

بينما اقترحت إحدى الدراسات منهجاً مختلفاً لتطبيق تقنيات التنقيب عن البيانات في شكل شبكات عصبية لبناء واختبار نماذج للتنبؤ بالأزمة المالية، وتدعم الأبحاث الإضافية منهجهم وتقترح أن أداء الشبكات العصبية أفضل في حالة نمذجة الأزمة المالية من أداء أشجار القرار والمناهج البديلة مثل آلات نقل الدعم (Geng, R. B., and et al., 2015)، ومع ذلك فإن الجمع بين كلا الأسلوبين يؤدي بشكل أفضل من كل منها بمفردها (Zhou, L. G., and et al., 2015).

ويحقق استخدام الشبكات العصبية قدرة عالية في مجال الكشف عن الممارسات الاحتيالية في التقارير المالية وهذا بدوره يزيد من فاعلية المراجعين في كشف هذه الممارسات، كما تتميز بالتنبؤ بما إذا كان عميل المراجعة ضحية للغش المالي أم لا وذلك قبل تخطيط عملية المراجعة مما يمكن المراجعين من إدخال معلومات للعميل في نموذج الشبكات العصبية ومن خلال نتيجة هذا النموذج يمكن التنبؤ بما إذا كان العميل ضحية للغش أم لا، وهذا الأمر يساعد المراجعين في مرحلة تخطيط المراجعة في معرفة حجم العمل المفترض القيام به ومجالات الفحص الإضافية، وذلك لقدرة هذه الشبكات علي استيعاب العديد من المتغيرات والكمية الهائلة من البيانات بالإضافة إلي مقارنتها والتوصل منها لقرار مناسب وبالتالي تساعد علي التنبؤ برأي المراجع في عدالة القوائم المالية كما توجهه نحو المزيد من المراجعات الإضافية لتقييم المخاطر (Jan C. L., 2018).

وقدمت أحدي الدراسات نموذج لـ 144 شركة يعمل علي كشف الغش والاحتيال المالي من خلال استخدام الشبكات العصبية، آلات نقل الدعم، والبرمجة الجينية لتحديد الشركات المتورطة في الاحتيال المالي، وأتضح تفوق الشبكات العصبية الاحتمالية والبرمجة الجينية على الأساليب الأخرى وهي دقيقة بالمثل، بينما قارنت أحدي الدراسات في الاختلافات بين أساليب التنقيب في البيانات وأوضح أن الشبكات العصبية وشجرة القرارات تحقق معدل التصنيف الصحيح لأكثر من90 ٪ على عينة محتجزة (Lin, C. C., and et al., 2015).

واستخدم (Lin, C. C., and et al., 2015) تقنيات التنقيب عن البيانات كالانحدار اللوجيستي وشجرة القرارات والشبكات العصبية الاصطناعية في كشف الغش وقامت بعمل مقارنات بين شجرة القرارات والشبكات العصبية الاصطناعية مع عينات التدريب والاختبار بشكل صحيح، وكان معدل التصنيف للشبكات العصبية ٩١,٢٪، ٩٠,٤٪ لشجرة القرارات، والذي هو أكثر دقة من النموذج اللوجيستي الذي يصل إلى ٨٣,٧٪ من التصنيف الصحيح في تقييم وجود الغش. في حين قدم (Sánchez-Serrano,

(J. R., 2020 نموذج جديد للتنبؤ برأي المراجع حول التقارير المالية، وتمثلت عينة الدراسة في ٢٩٨ شركة من الشركات العاملة في السوق الإسبانية وتم تطبيق تقنية الشبكات العصبية الاصطناعية علي المعلومات التي تم الحصول عليها من العينة، وأظهرت الدراسة أن تقنية الشبكات العصبية تنبأت برأي المراجعة بدقة ٨٦٪ وهذه نسبة كبيرة لتوفير صورة عن مدي عدالة القوائم المالية، بينما طور (Kim, Y. J., and et al., 2016)أحدي النماذج القادر علي التمييز بين البيانات المالية وممارسات الاحتيال والأخطاء غير المقصودة التي تؤدي إلي إعادة البيانات المالية بدون أخطاء وذلك باستخدام آلية نقل الدعم وحقق معدل دقة من ٨٢ إلى ٨٨٪ في كشف التحريف المالي متعدد المستويات. أما (Dutta, L., and et al., 2017) قدم نهجاً مشابهاً في نموذج من فئتين يجمع بين استخدام العديد من تقنيات التنقيب عن البيانات مثل (الأقرب جار، والشبكات العصبية، وشجرة القرارات، الخوارزمية الجينية) في الكشف عن عمليات إعادة الصياغة للقوائم المالية المقصودة وغير المقصودة، وتحليل المتغيرات المالية وغير المالية.

كما قام (Li, D., 2017) أيضاً ببناء نموذج التنبؤ برأي المراجع على أساس الشبكة العصبية، واستخدم (Zhiheng Z., and et al., 2017) الشبكة العصبية لإنشاء نموذج التنبؤ، ودرس (Jie, S., and et al., 2018) نموذج التنبؤ لمراجعة الحسابات متعددة الفئات على أساس على آلية نقل الدعم، ويتضح مما سبق توجه الاتجاه البحثي لاستخدام تكنولوجيا التعلم العميق لحل مخاطر المراجعة (Zeng, J., & Yang, Y., 2021). واستخدم أيضاً شجرة القرارات، والشبكات العصبية، وشبكات المعتقدات بايزي على الشركات البريطانية والأيرلندية، وأظهرت النتائج أن متغيرات الإعسار والربحية عالية تتعلق بآراء المراجع النظيفة، ومن ناحية أخرى اختار(Sánchez-Serrano, J. R., and et al., 2020) عينة من قائمة الشركات في بورصة لندن والتي تظهر قوة شرح أكبر لأساليب أقرب جار حيث كانت الإضافة والربحية وتصنيف الائتمان هي الأكثر تنبؤاً.

وتري الباحثة أن الحالات التي يتم خداع المراجعين من قبل المديرين ترجع إلى احتمالية احتواء التقارير المالية علي عوامل مخاطر الاحتيال أم لا حيث يمكن توضيح العوامل التي تساعد على اكتشاف غش القوائم المالية في العوامل التالية مثل (النمو السريع، وضعف بيئة الرقابة الداخلية، وانشغال الإدارة بتلبية توقعات الأرباح، وتباطئي الإدارة في تقديم الأدلة الخاصة بعملية المراجعة للمراجع الخارجي، ونوع الملكية، وموقف الإدارة العدواني تجاه التقارير المالية) وبعد تحديد هذه العوامل تستطيع تقنيات التنقيب عن البيانات تطبيقها على القوائم المالية ومن ثم تتمكن من تصنيف الشركات على أنها شركات احتيالية وغير احتيالية وبعد ذلك تتنبأ برأي المراجع في القوائم المالية. ويمكن القول أن لكل تقنية للتنقيب عن البيانات نقاط قوه وضعف ولا توجد تقنية فردية تمثل حلاً مثاليًا لمشكلة اكتشاف غش القوائم المالية ، وبالتالي قد يتم دمج تقنيات التنقيب عن البيانات المتعددة لتشكيل نظام هجين للكشف الناجح في الوقت المناسب عن الاحتيال، واستناداً لما سبق يتضح أن الاستخدام المتزايد لتكنولوجيا لمعلومات في عملية المراجعة وخاصةً في مجال تحليل البيانات والتنقيب عن البيانات يكون له أثر كبير علي كفاءة وفاعلية عملية المراجعة وذلك لمنع أو تقليل التلاعب بالبيانات المالية، فمن خلال استخدام التنقيب عن البيانات لأسلوب العينات والإجراءات التحليلية يمكن التنبؤ برأي المراجع الخارجي في عدالة القوائم المالية كما يوفر التنقيب عن البيانات العديد من المزايا كتدعيم أنظمة الرقابة.

القسم الثالث: الدراسة التطبيقية

تهدف الدراسة الحالية إلى دراسة أثر استخدام تقنيات التنقيب عن البيانات في التنبؤ برأي المراجع عن مدى عدالة القوائم المالية، ولكي تتمكن الباحثة من تحقيق هذا الهدف فإنها ستقوم بإجراء التحليل الإحصائي أولاً باستخدام تقنيات التنقيب عن البيانات المتمثلة في: شجرة القرارات، والشبكات العصبية، ونموذج تعلم الآلة وتقنية الأقرب جار. ثانياً، ستقوم الباحثة بإجراء تحليل الانحدار اللوجيستي للتعرف على مدى تفوق كل تقنية من التقنيات الأربعة في تحسين دقة التنبؤ برأي المراجع عن عدالة القوائم المالية. ويمكن للباحثة تحقيق ذلك من خلال: عرض مجتمع وعينة الدراسة، ثم تعريف المتغيرات ومعلمات القرار بالنسبة للمراجع الخارجي، ثم تحليل البيانات باستخدام تقنيات التنقيب عن البيانات، ثم تحليل البيانات باستخدام الانحدار اللوجيستي (الأساليب التقليدية)، وأخيراً عرض نتائج اختبارات فروض الدراسة من خلال مقارنة النتائج وذلك على خلاف ما استقرت عليه الأبحاث من الاستناد إلى المؤشرات الإحصائية.

أولاً: مجتمع وعينة الدراسة

يتمثل مجتمع الدراسة في كافة الشركات المتحولة إلى النظم الرقمية في الآونة الأخيرة ولا سيما تلك المسجلة في سوق الأوراق المالية المصري، ومن خلال تحليل المعلومات المتعلقة بالعديد من الشركات الصناعية المسجلة في سوق الأوراق المالية المصري تبين للباحثة تحول العديد من الشركات إلى تطبيق النظم الرقمية وهي مبينة بالجدول التالي:

م	اسم الشركة	تاريخ التحول إلى النظم الرقمية	عدد المشاهدات
1	مصر لإنتاج الأسمدة – موبكو	2018	4
2	شركة سيد كرير للبتروكيماويات (سيد بك)	2019	3
3	مصر للأسمنت – قنا	2017	5
4	جنوب الوادي للأسمنت	2017	5
5	شركة السويدي إلكتريك	2020	1
6	شركة جى . بى أوتو غبور	2017	5
7	شركة مجموعة عامر القابضة (عامر جروب)	2020	1
8	الشركة المصرية لخدمات النقل والتجارة (ايجيبترانس)	2011	11
9	شركة القناة للتوكيلات الملاحية	2020	1
10	شركة النصر للملابس والمنسوجات (كابو)	2019	2
11	شركة النساجون الشرقيون	2018	3
12	ايديتا للصناعات الغذائية	2018	3
13	جهينة للصناعات الغذائية	2019	2
14	اسيك للتعدين – اسكوم	2015	7
15	العز الدخيلة للصلب – الاسكندرية	2014	8
16	مستشفى النزهة الدولى	2016	6
17	مينا فارم للأدوية والصناعات الكيماوية	2016	6

م	اسم الشركة	تاريخ التحول إلى النظم الرقمية	عدد المشاهدات
18	شركة الصناعات الهندسية المعمارية للإنشاء والتعمير(إيكون)	2016	6
19	شركة أكرو مصر للشدات والسقالات المعنية	2017	5
20	شركة السادس من أكتوبر للتنمية والاستثمار(سوديك)	2016	6
21	الشركة المصرية لمدينة الإنتاج الإعلامي	2017	5
22	الشركة المصرية للاتصالات	2014	8
23	أوراسكوم للاستثمار القابضة	2017	5
24	شركة راية للاستثمارات المالية	2019	3
25	الشركة الشرقية للدخان	2020	1
إجمالي عدد المشاهدات			**١١٢ مشاهدة**

ثانياً: تعريف المتغيرات ومعلمات القرار

في إطار تحقيق هدف الدراسة الحالي ستصبح متغيرات ومعلمات الدراسة هي كافة المتغيرات التي يمكن للمراجع استخدامها لبناء الرأي على أساسها بشأن عدالة القوائم المالية على أن يتم تحليل أثرها على التنبؤ برأي المراجع واستخدامها أيضاً في التقنيات الرقمية للتنقيب عن البيانات للتنبؤ برأي المراجع. وبناء على ذلك تصبح (X1, X2, X..... Xn) هي المعبرة عن متغيرات القرار أما Y فهي المعبرة عن معلمات قرار المراجع بشأن عدالة القوائم المالية ويمكن تعريف كل منها على النحو التالي:

المتغير	التفسير
Y	رأي المراجع الخارجي ويأخذ القيمة ١ في حالة كونه نظيف والقيمة ٤ إذا كان نظيف مع فقرة تفسيرية والقيمة ٦ إذا كان متحفظاً
X1	معدل العائد على الأصول، ويتم حسابه بصافي ربح الشركة مقسوماً على إجمالي الأصول
X2	العائد على حقوق الملكية، ويتم حسابه بقسة صافي الربح على حقوق الملكية
X3	صافي الربح قبل الفوائد والضرائب مقسوماً على المبيعات
X4	صافي الدخل بعد الفوائد والضرائب
X5	صافي الدخل مقسوماً على صافي المبيعات
X6	الأرباح المحتجزة مقسومة على إجمالي الأصول
X7	نسبة النقدية ويتم حسابها بإجمالي النقدية وما يكافئها مقسومة على الاتزامات المتداولة
X8	نسب التداول السريع ويمكن حسابه بإجمالي الأصول سريعة السيولة مثل النقدية وما يكافئها وحسابات العملاء وأوراق القبض
X9	نسبة التداول ويمكن حسابها بإجمالي الأصول المتداولة على الالتزامات المتداولة
X10	رأس المال العامل / إجمالي الأصول
X11	التدفقات النقدية التقليدية وتساوي مجموع صافي الدخل والفوائد والضرائب والإهلاك

المتغير	التفسير
X12	مجموع صافي الدخل والفوائد والضرائب والإهلاك مقسوماً على صافي المبيعات
X13	معدل دوران النقدية ويتم قياسه بالمبيعات على متوسط النقدية
X14	نسبة المديونية ويمكن قياسها بقسمة إجمالي الديون على إجمالي الاصول
X15	دوران المخزون ويمكن قياسه بقسمة المبيعات على متوسط المخزون
X16	حسابات العملاء / إجمالي الأصول
X17	حسابات العملاء / المبيعات
X18	رأس المال السوقي ويمكن حسابه بقسمة عدد الأسهم المتداولة إجمالي أسعار الأسهم في نهاية العام
X19	إجمالي أسعار الأسهم على الأرباح في نهاية السنة
X20	متغير وهمي لاختلاف الصناعة
X21	متغير وهمي يأخذ القيمة ١ في حالة وجود مراجع خارجي عالمي والقيمة صفر فيما عدا ذلك
X22	متغير وهمي يأخذ القيمة ١ في حالة تغيير المراجع الخارجي والقيمة صفر فيما عدا ذلك
X23	لوغاريتم إجمالي الأصول
X24	لوغاريتم صافي المبيعات

ثالثاً: تحليل البيانات باستخدام تقنيات التنقيب عن البيانات

من خلال الدراسة النظرية يمكن للباحثة أن تقسم تقنيات التنقيب عن البيانات إلى أربعة مجموعات رئيسي تتمثل في: شجرة القرارات، والشبكات العصبية، ونموذج تعلم الآلة، وتقنية أقرب جار، والجدير بالذكر أن استخدام التقنيات الرقمية يسفر عن أشكال التنبؤ والأخطاء الموجودة به في إطار ثلاثة مستويات للملائمة للنموذج تتمثل في: Training set وهي المجموعة الإحصائية التي يتم الاستناد اليها في ملاءمة معلمات كل نموذج عند القيام بعملية التنبؤ وهي تمثل نسبة ٩٠٪ من عملية التنبؤ، ثم يتم الاستناد إلى Ten-fold cross-validation لأداء إجراءات تحديد ملاءمة النموذج وذلك بعد اكتمال نسبة ٩٠٪ من الملاءمة في الخطوة السابقة Training set وأخيراً تقسيم درجة الملاءمة بنسب ٨٠٪ للمؤشر Train و٢٠٪ للمؤشر Test وذلك لعدم المبالغة في درجات الملاءمة والسيطرة على عدم استقرار البيانات. ويمكن استخدام كافة الطرق السابقة في التحليل للوصول إلى المستويات الثلاث سالفة الذكر على النحو التالي:

- **<u>شجرة القرارات للتنبؤ برأي المراجع:</u>**

تعد أحد الأساليب التي تقوم بتحليل البيانات من خلال تقسيمها إلى مجموعات كل مجموعة منها تختص بخاصية معينة، وتستند جودة تقسيم البيانات إلى مجموعات فرعية على الخصائص التي تم تحديدها في مجموعة البيانات، وتعتمد شجرة القرارات استخدام مؤشر الكثافة للبيانات في اتجاه معين لتقسيم البيانات إلى مجموعات فرعية، وتعتمد تقنية شجرة القرارات في تحليل البيانات وعمليات التنبؤ على استخدام طريقة تحليل المكونات الرئيسية (Principle Component analysis) وهي طريقة إحصائية متعارف عليها تهدف إلى اختزال العوامل بين المجموعات المختلفة على أن تكون كل مجموعة مستقلة في حد ذاتها تشكل نموذج يمكن من خلاله التنبؤ بالنماذج سالفة الذكر. وبتشغيل طريقة المكونات الأساسية للوصول إلى المجموعات الفرعية لنماذج التنبؤ تم التوصل إلى مصفوفة العوامل المدورة التالية في الجدول التالي رقم (١-٣).

جدول رقم (٣-١): مصفوفة العوامل المدورة للمكونات الرئيسية للنموذج

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
X1	0.159	0.106	0.228	0.103	0.091	0.2	0.083	0.843	0.141	0.009	-0.067
X2	0.109	0.224	0.194	0.125	0.172	0.111	-0.15	0.588	0.022	0.397	0.194
X3	0.157	0.09	0.157	0.096	0.124	0.175	0.138	0.079	0.103	0.884	0.007
X4	0.225	0.088	0.07	0.186	0.159	0.783	0.107	0.165	0.03	0.208	0.101
X5	0.164	0.175	0.239	0.113	0.371	0.673	0.058	0.159	0.128	0.055	0.098
X6	0.161	0.117	0.137	0.155	0.778	0.177	0.027	0.102	0.189	0.05	0.126
X7	0.2	0.196	0.259	0.155	0.781	0.157	0.024	0.054	0	0.098	0.14
X8	0.132	0.21	0.41	0.084	0.594	0.118	0.181	0.046	0.118	0.1	-0.017
X9	0.112	0.277	0.738	0.019	0.271	0.121	0.02	0.154	0.092	0.149	0.031
X10	0.169	0.179	0.815	0.139	0.151	0.101	0.022	0.094	0.185	0.078	0.016
X11	0.268	0.156	0.17	0.156	0.217	0.162	0.162	-0.025	0.098	0.024	0.793
X12	0.241	0.193	0.655	-0.04	0.205	0.026	0.207	0.174	0.033	0.024	0.364
X13	0.304	0.147	0.235	0.071	0.226	0.084	0.179	0.149	0.748	0.113	0.065
X14	0.148	0.424	0.446	0.102	0.131	0.096	0.118	0.123	0.508	0.078	0.142
X15	0.342	0.271	0.063	0.22	0.099	0.156	0.689	0.023	0.136	0.165	0.071
X16	0.347	0.356	0.188	0.054	0.061	0.039	0.589	0.018	0.16	0.049	0.236

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
X17	0.28	0.629	0.216	0.122	0.11	0.149	0.223	0.099	0.29	0.198	0.085
X18	0.271	0.718	0.326	0.085	0.111	0.114	0.149	0.135	0.024	0.06	0.079
X19	0.383	0.73	0.114	0.121	0.25	0.008	0.119	0.096	0.103	0.038	0.138
X20	0.47	0.608	0.267	0.13	0.17	0.094	0.207	0.056	0.019	0.015	0.012
X21	0.354	0.47	0.307	0.082	0.171	0.193	0.029	-0.044	0.218	-0.018	0.062
X22	0.672	0.427	0.202	0.071	0.18	0.089	0.043	0.038	0.022	-0.013	-0.042
X23	0.801	0.177	0.114	0.184	0.229	0.036	0.208	0.055	0.049	0.009	0.002
X24	0.331	-0.006	0.009	0.804	0.185	0.055	0.167	0.081	-0.045	0.019	0.039
Eigenvalue	14.532	2.7154	1.6962	1.1413	0.9631	0.8573	0.7833	0.7322	0.6894	0.6648	0.6006
Variance explained (%)	45.411	8.4857	5.3007	3.5666	3.0096	2.679	2.4479	2.2881	2.1544	2.0776	1.8769
Cum. variance explained (%)	45.411	53.897	59.198	62.764	65.774	68.453	70.901	73.189	75.343	77.421	79.298

يتضح من النتائج المبينة بالجدول السابق أن النموذج الحالي للدراسة أنقسم إلى اثنا عشر نموذجاً فرعياً من PC1 إلى PC11 وكل نموذج من هذه النموذج يضم مجموعة من المتغيرات الرئيسية التي تؤثر على بناء رأي المراجع عن عدالة القوائم المالية فمثلاً المكون PC1 يحتوي على المتغيرات X22 & X23 بينما يحتوي المكون PC2 على المتغيرات من X17 إلى X21 وهكذا.....

وقد أسفر التفاعل بين النماذج المختلفة للمكونات الرئيسية عن نتائج التشغيل الآلية للتنبؤ برأي المراجع عن النتائج المختلفة لأخطاء التنبؤ برأي المراجع كما هو موضح بالجدول رقم (٣-٢) التالي:

جدول رقم (٣-٢): نتائج أخطاء التنبؤ برأي المراجع نتيجة استخدام شجرة القرارات

	تنبؤات نسب التقرير النظيف			تنبؤات نسب التقرير النظيف مع فقرة تفسيرية			تنبؤات نسب التقرير المتحفظ		
Model	**Accuracy (%)**	**Type 1 error (%)**	**Type 2 error (%)**	**Accuracy (%)**	**Type 1 error (%)**	**Type 2 error (%)**	**Accuracy (%)**	**Type 1 error (%)**	**Type 2 error (%)**
Training set	77.5	15.6	22.5	71.3	10.5	28.7	91.5	2.5	8.5
Ten-fold cross-validation	55.2	25.6	44.8	52.8	17.6	47.2	92.3	5.6	7.7
Train (80%)-Test (20%) split	51.6	24.8	48.4	53.4	15.3	46.6	90.7	4.3	9.3
الخطأ الأول Type 1 eror	خطأ التنبؤ بتقرير نظيف في حين أنه ربما يكون نظيف مع التفسير أو متحفظ			خطأ التنبؤ بتقرير نظيف مع التفسير في حين أنه ربما يكون نظيف أو متحفظ			خطأ التنبؤ بتقرير متحفظ في حين أنه ربما يكون نظيف أو نظيف مع التفسير		
الخطأ الثاني Type 2 eror	خطأ التنبؤ بتقرير نظيف مع التفسير أو متحفظ في حين أنه ربما يكون نظيف			خطأ التنبؤ بتقرير نظيف أو متحفظ في حين أنه ربما يكون نظيف مع التفسير			خطأ التنبؤ بتقرير نظيف أو نظيف مع التفسير في حين أنه ربما يكون متحفظ		

يوضح الجدول السابق ارتفاع مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبئ به حيث كانت الدقة تساوي ٧٧,٥٪، ٧١,٣٪، و٩١,٥٪ على التوالي بينما تبين وجود أخطاء تنبؤ لا يستهان بها حيث إنه في العمود الأول كانت أخطاء التنبؤ من النوع الأول ١٥,٦٪، ٢٥,٦٪، ٢٤,٨٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أنه من المفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية أو

متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٢٢,٥٪، ٤٤,٨٪، ٤٨,٤٪ على التوالي أي أن هناك احتمالية لأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية أو متحفظ في حين يفترض أن يكون التقرير نظيف.

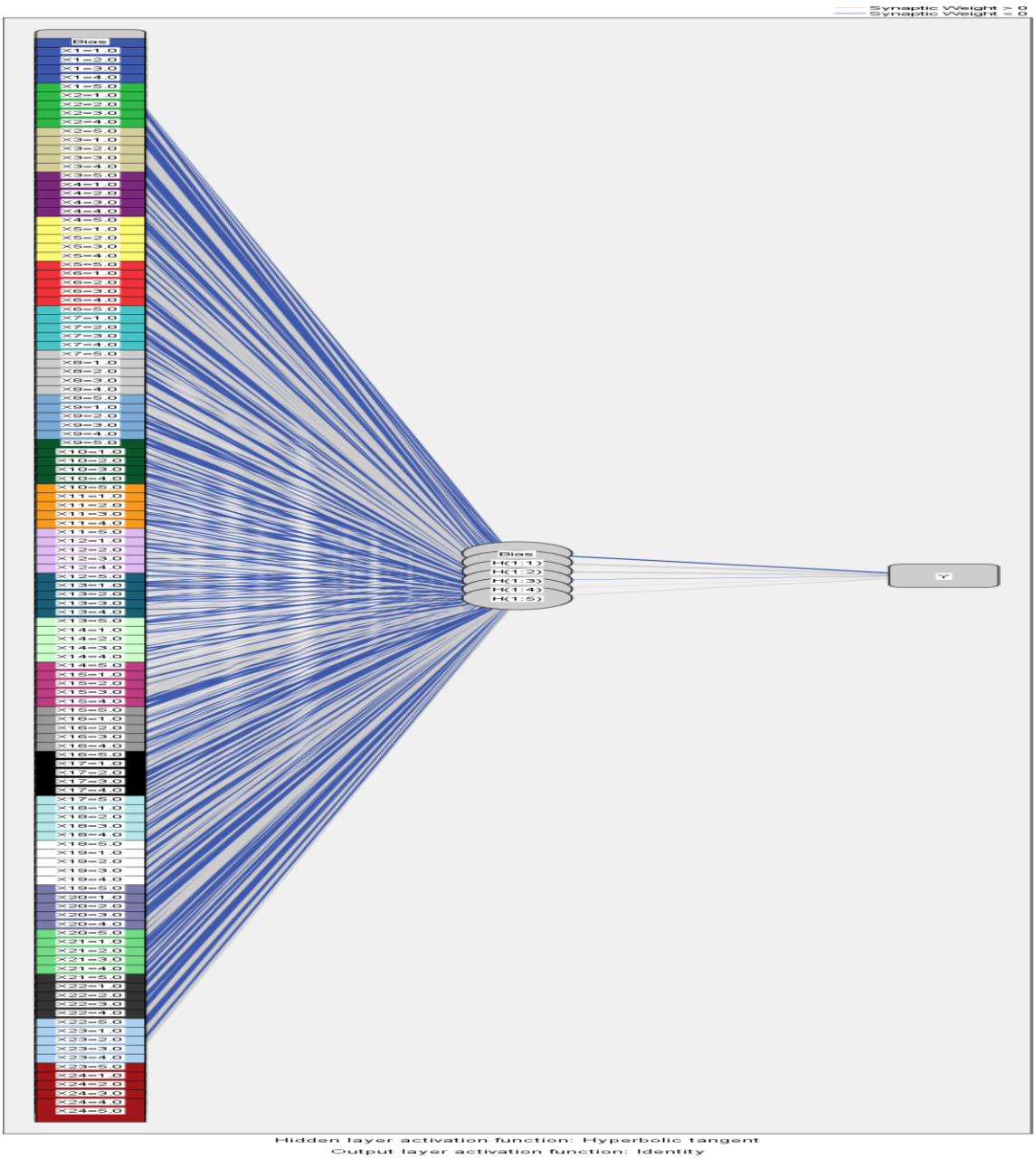
وفي العمود الثاني كانت أخطاء التنبؤ من النوع الأول ١٠,٥٪، ١٧,٦٪، ١٥,٣٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير نظيف أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٢٨,٧٪، ٤٧,٢٪، ٤٦,٦٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو متحفظ في حين يفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية.

أما العمود الثالث كانت أخطاء التنبؤ من النوع الأول ٢,٥٪، ٥,٦٪، ٤,٣٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير متحفظ مع أنه من المفترض أن يكون التقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية. بينما كانت أخطاء التنبؤ من النوع الثاني ٨,٥٪، ٧,٧٪، ٩,٣٪ على التوالي أي أن هناك احتمالية لأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير متحفظاً.

ويلاحظ مما سبق ارتفاع مستويات الخطأ في العمود الأول والثاني الخاصين بالرأي النظيف والرأي النظيف مع أهمية إضافة فقرة تفسيرية، بينما يتبين انخفاض مستويات الخطأ في إبداء الرأي المتحفظ، وبالتالي يتبين للباحثة أن استخدام شجرة القرارات في التنبؤ برأي المراجع وسيلة فعالة في اكتشاف أخطاء المراجعين في إبداء الرأي الفني المحايد بشأن عدالة القوائم المالية.

- **الشبكات العصبية للتنبؤ برأي المراجع**

تعتبر الشبكات العصبية أسلوب تقني يعمل باستخدام الدوال غير الخطية للتنبؤ بالأحداث المستقبلية بناء على البيانات التاريخية لمجموعة من العوامل المحيطة بالظاهرة محل التنبؤ، حيث إنها تعمل باستخدام أسلوب المحاكاة للوصول إلى أفضل السيناريوهات الممكنة للتنبؤ بشكل رأي المراجع الخارجي في التقارير، وقد أسفرت المحاكاة باستخدام التقنيات للشبكات العصبية عن التنبؤات التالية لعينة الدراسة.



شكل رقم (١): صور المحاكاة للقيم التنبؤية بناء على الشبكات العصبية

جدول رقم (٣-٣): نتائج أخطاء التنبؤ برأي المراجع نتيجة استخدام الشبكات العصبية

	تنبؤات نسب التقرير النظيف			تنبؤات نسب التقرير النظيف مع فقرة تفسيرية			تنبؤات نسب التقرير المتحفظ		
Model	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**
Training set	78.4	15.8	21.6	67.8	11.2	32.2	95.8	1.2	4.2
Ten-fold cross-validation	75.1	12.3	24.9	58.6	15.7	41.4	96.4	3.8	3.6
Train (80%)-Test (20%) split	67.5	14.4	32.5	64.3	16.1	35.7	96.8	2.1	3.2
الخطأ الأول Type 1 eror	خطأ التنبؤ بتقرير نظيف في حين أنه ربما يكون نظيف مع التفسير أو متحفظ			خطأ التنبؤ بتقرير نظيف مع التفسير في حين أنه ربما يكون نظيف أو متحفظ			خطأ التنبؤ بتقرير متحفظ في حين أنه ربما يكون نظيف أو نظيف مع التفسير		
الخطأ الثاني Type 2 eror	خطأ التنبؤ بتقرير نظيف مع التفسير أو متحفظ في حين أنه ربما يكون نظيف			خطأ التنبؤ بتقرير نظيف أو متحفظ في حين أنه ربما يكون نظيف مع التفسير			خطأ التنبؤ بتقرير نظيف أو نظيف مع التفسير في حين أنه ربما يكون متحفظ		

يوضح الجدول السابق ارتفاع مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبئ به حيث كانت الدقة تساوي ٧٨,٤٪، ٦٧,٨٪، ٩٥,٨٪ على التوالي بينما تبين وجود أخطاء تنبؤ لا يستهان بها حيث أنه في العمود الأول كانت أخطاء التنبؤ من النوع الأول ١٥,٨٪، ١٢,٣٪، ١٤,٤٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أنه من المفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٢١,٦٪، ٢٤,٩٪، ٣٢,٥٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية أو تقرير متحفظ في حين يفترض أن يكون التقرير نظيف.

وفي العمود الثاني كانت أخطاء التنبؤ من النوع الأول ١١,٢٪، ١٥,٧٪، ١٦,١٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير نظيف أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٣٢,٢٪، ٤١,٤٪، ٣٥,٧٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو متحفظ في حين يفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية.

وبالنسبة للعمود الثالث كانت أخطاء التنبؤ من النوع الأول ١,٢٪، ٣,٨٪، ٢,١٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير متحفظ مع أنه من المفترض أن يكون التقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية. بينما كانت أخطاء التنبؤ من النوع الثاني ٤,٢٪، ٣,٦٪، ٣,٢٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير متحفظاً.

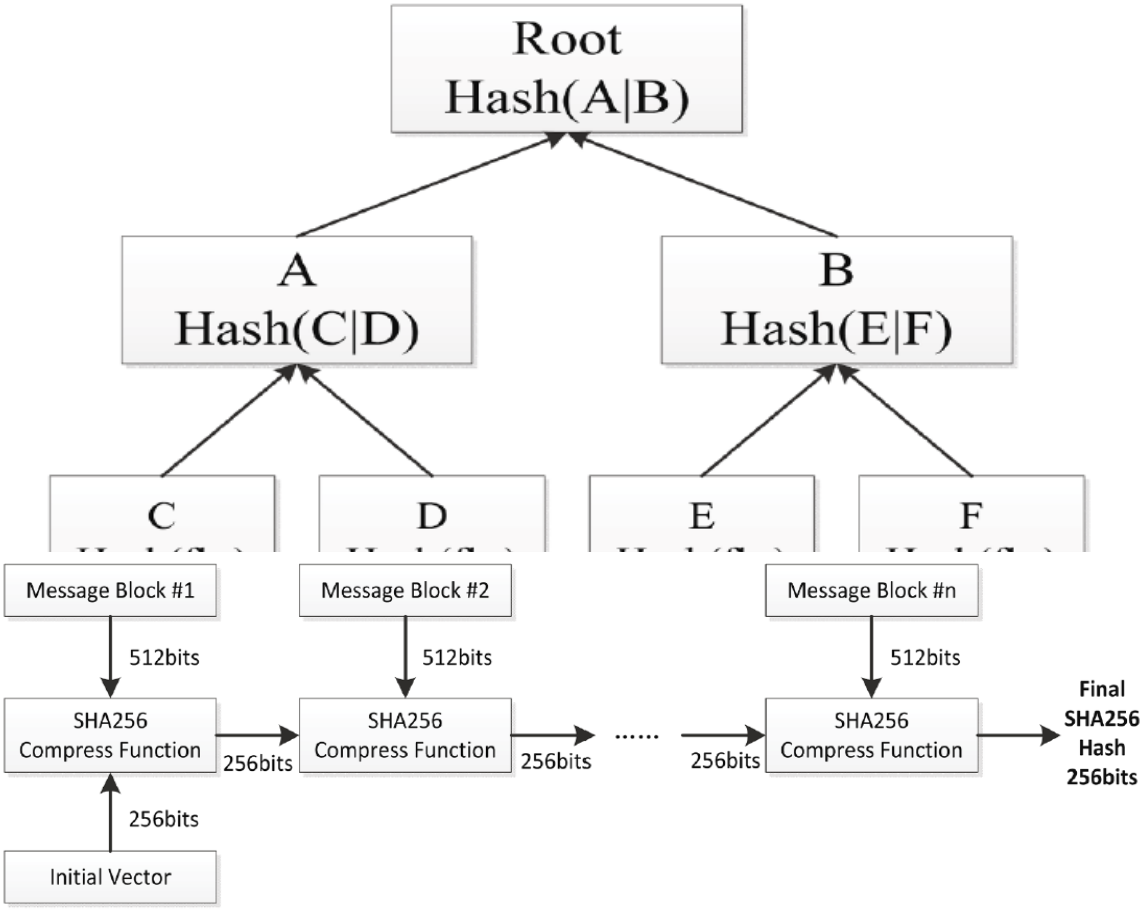
ويلاحظ مما سبق ارتفاع مستويات الخطأ في العمود الأول والثاني الخاصين بالرأي النظيف والرأي النظيف مع أهمية إضافة فقرة تفسيرية، بينما يتبين انخفاض مستويات الخطأ في إبداء الرأي المتحفظ. وبالتالي، يتبين للباحثة إن استخدام الشبكات العصبية في التنبؤ برأي المراجع قد اتفق مع استخدام شجرة القرارات في اكتشاف أخطاء المراجعين في إبداء الرأي الفني المحايد بشأن عدالة القوائم المالية.

- **نموذج تعلم الآلة للتنبؤ برأي المراجع**

يمكن للباحثة بناء الخوارزميات على برنامج التشغيل الالكتروني Python للحصول على نتائج التنبؤ برأي المراجع، وتجدر الإشارة إلى أن هذه اللغة البرمجية تساعد على ضم وتشغيل تقنيات نموذج تعلم الآلة، وبالتالي يمكن من خلالها بناء نظام ال Hash الخاص بالتشفير والتكويد على لغة البايثون للتعبير عن تقنيات نموذج تعلم الآلة، ثم توضيح آلية العمل عند إعطاء أمر التشغيل (Run) لتلك الدالة، وأخيراً توضيح الهيكل المستخرج للنموذج المعلوماتي بعد عملية التشغيل وهو ما يمكن استخدام نتائجه النهائية للتعبير عن أخطاء التنبؤ برأي المراجع، وذلك على النحو التالي:

١- بناء نظام ال Hash الخاص بالتشفير والتكويد

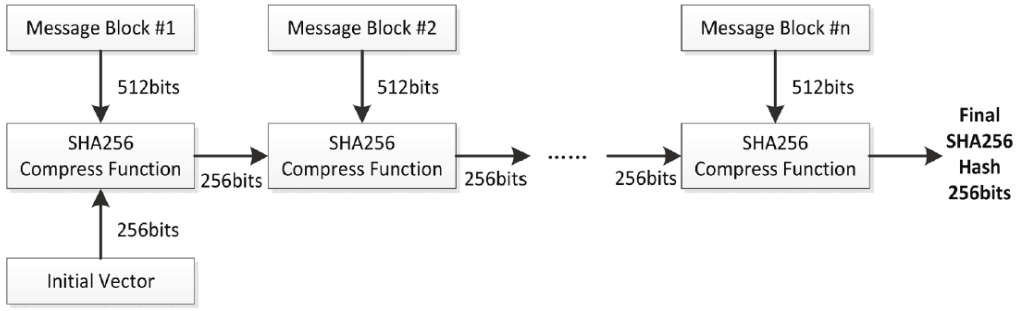
قامت الباحثة في هذه الخطوة بتكويد الدالة الرياضية مع تخزين المعلومات الخاصة بكل شركة على حدة على مستوى السلسلة الزمنية، وهو ما أدى إلى بناء شجرة التشفير والتكويد على الشكل التالي وفقاً لما هو مستخرج من البرنامج:



شكل رقم (٢): نظام التكويد والتشفير

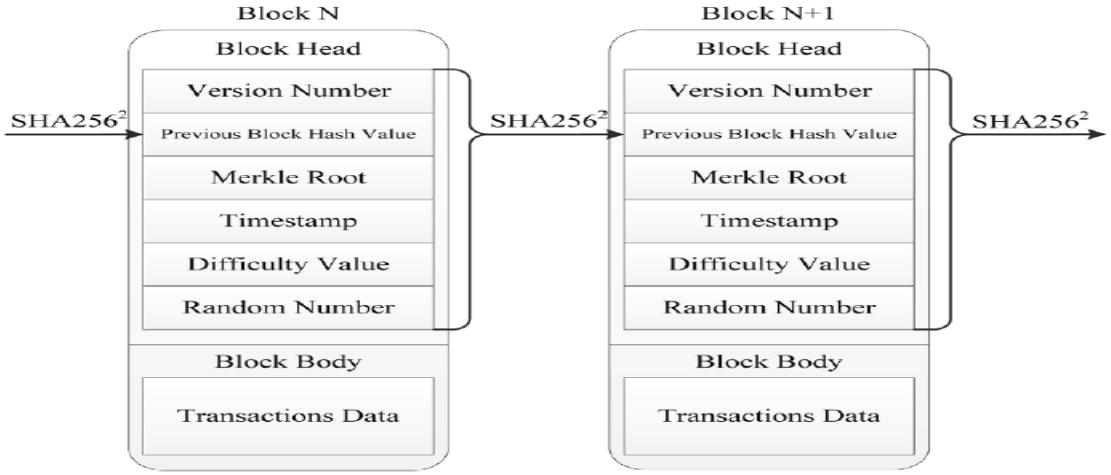
٢- إعطاء أمر التشغيل (Run) للدالة الرياضية:

يمكن للباحثة في هذه المرحلة إعطاء أمر التشغيل للدالة بعد إجراء عملية التكويد والشفرات بنظام الHash وذلك لغرض تحليل آلية عمل البرنامج لهذه الدالة الرياضية وهو ما أدى إلى وجود النوافذ التالية:



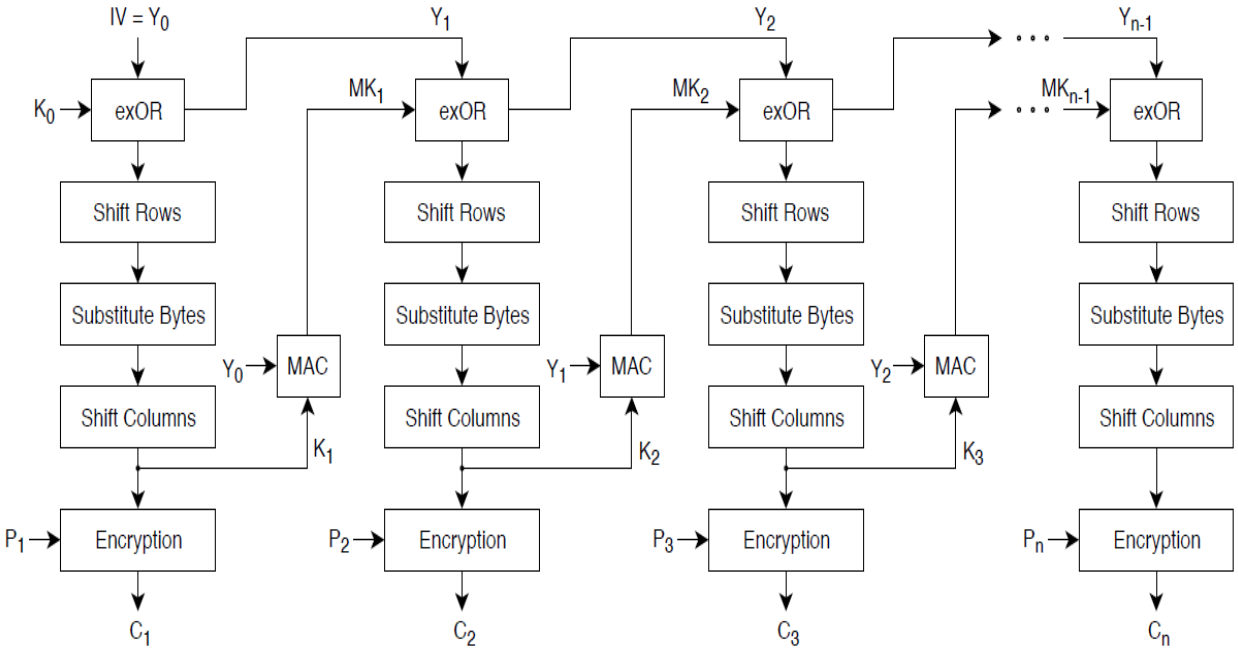
شكل رقم (٣): الرسائل التشغيلية للبرنامج بناء على البيانات المخزنة

وتشير تلك النافذة إلى أن الدالة الرياضية بدأت في العمل والبحث عن سلوك المعلومات التاريخية على مدار السلسة الزمنية المخزنة من خلال مجموعة من كتل المعلومات الموجودة ثم تبدأ في معالجة هذه المعلومات لتقدير السلوك التنبؤي المثالي للدالة على النحو التالي:



شكل رقم (٤): آلية عمل الدوال الرياضية لنموذج تعلم الآلة على البرنامج

تشير عملية المعالجة المعلوماتية السابقة إلى معالجة كافة المعلومات في حدود الوقت التنبؤي والأرقام العشوائية لتقديرات القيم العادلة للوصول إلى البيانات النهائية من حيث الترتيب والشكل اللازم لتشغيلها وفقاً للأحداث المستقبلية وتجميع كافة بيانات الشركات مع بعضها البعض وذلك على النحو التالي:



شكل رقم (٥): التقديرات العشوائية للقيم التنبؤية والتي يتم من خلالها اختيار القيم المثلى

يشير الشكل السابق إلى التدفق المعلوماتي للوصول إلى القيمة المتنبئ بها حيث تعبر قيمة C1,2….n إلى القيم المثالية لشكل تقرير المراجع للشركات التي تم تسجيل بياناتها على مدار السلسلة الزمنية. وقد أسفر التشغيل الآلي للنموذج عن النتائج التالية للتنبؤ برأي المراجع عن النتائج المختلفة لأخطاء التنبؤ كما هو موضح بالجدول رقم (٣-٤) التالي:

جدول رقم (٣-٤): نتائج أخطاء التنبؤ برأي المراجع نتيجة استخدام نموذج تعلم الآلة

	تنبؤات نسب التقرير النظيف			تنبؤات نسب التقرير النظيف مع فقرة تفسيرية			تنبؤات نسب التقرير المتحفظ		
Model	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**
Training set	74.4	11.4	25.6	60.8	12.7	39.2	98.5	2.3	1.5
Ten-fold cross-validation	73.5	14.6	26.5	87.9	14.1	12.1	93	2.6	7
Train (80%)-Test (20%) split	81.7	16.3	18.3	70.2	16.3	29.8	92.2	2.1	7.8
الخطأ الأول Type 1 eror	خطأ التنبؤ بتقرير نظيف في حين أنه ربما يكون نظيف مع التفسير أو متحفظ			خطأ التنبؤ بتقرير نظيف مع التفسير في حين أنه ربما يكون نظيف أو متحفظ			خطأ التنبؤ بتقرير متحفظ في حين أنه ربما يكون نظيف أو نظيف مع التفسير		
الخطأ الثاني Type 2 eror	خطأ التنبؤ بتقرير نظيف مع التفسير أو متحفظ في حين أنه ربما يكون نظيف			خطأ التنبؤ بتقرير نظيف أو متحفظ في حين أنه ربما يكون نظيف مع التفسير			خطأ التنبؤ بتقرير نظيف أو نظيف مع التفسير في حين أنه ربما يكون متحفظ		

يوضح الجدول السابق ارتفاع مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبئ به حيث كانت الدقة تساوي ٧٤,٤٪، ٦٠,٨٪، ٩٨,٥٪ على التوالي بينما تبين وجود أخطاء تنبؤ لا يستهان بها حيث إنه في العمود الأول كانت أخطاء التنبؤ من النوع الأول ١١,٤٪، ١٤,٦٪، ١٦,٣٪ على التوالي أي أن هناك احتمالية لأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أنه من المفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٢٥,٦٪، ٢٦,٥٪، ١٨,٣٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية أو متحفظ في حين يفترض أن يكون التقرير نظيف.

أما العمود الثاني كانت أخطاء التنبؤ من النوع الأول ١٢,٧٪، ١٤,١٪، ١٦,٣٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير نظيف أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٣٩,٢٪، ١٢,١٪، ٢٩,٨٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو متحفظ في حين يفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية.

وفي العمود الثالث كانت أخطاء التنبؤ من النوع الأول ٢,٣٪، ٢,٦٪، ٢,١٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير متحفظ مع أنه من المفترض أن يكون التقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية. بينما كانت أخطاء التنبؤ من النوع الثاني ١,٥٪، ٧٪، ٧,٨٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير متحفظاً.

ويلاحظ مما سبق ارتفاع مستويات الخطأ في العمود الأول والثاني الخاصين بالرأي النظيف والرأي النظيف مع أهمية إضافة فقرة تفسيرية، بينما يتبين انخفاض مستويات الخطأ في إبداء الرأي المتحفظ، ومن ثم يتبين للباحثة أن استخدام نموذج تعلم الآلة يتفق مع النموذجين السابقين من تقنيات البيانات الضخمة (شجرة القرارات، والشبكات العصبية) في التنبؤ برأي المراجع كوسيلة فعالة في اكتشاف أخطاء المراجعين في إبداء الرأي الفني المحايد بشأن عدالة القوائم المالية.

- **تقنية أقرب جار للتنبؤ برأي المراجع**

تعد تقنية الأقرب جار أحد أهم التقنيات المستخدمة من خلال التنقيب عن البيانات كما تعتبر هذه التقنية أحد طرق التصنيف التي لا تتطلب أي افتراضات بشأن شكل العلاقة بين دوال العضوية وأدوات التنبؤ، ومن ثم فهي طريقة لامعلمية لأنها لا تتضمن تقدير المعلمات في شكل الدالة الرئيسية للنموذج، وبالتالي فإن التساؤل الرئيسي لهذه التقنية يتعرف على كيفية قياس العلاقة بين السجلات المحاسبية والقيم المتنبئ بها، ومن ثم فإن الدالة الرئيسية لهذه التقنية تتمثل في:

$$d(x,y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

حيث أنه قد تم تعريف قيم X و Y فيما سبق بينما تشير قيمة n إلى عدد المشاهدات بحجم العينة، كما أن k تشير إلى خاصية المشاهدة محل الدراسة (وفي حالة الدراسة تعبر عن الشركة المأخوذ منها البيانات). وقد أسفرت نتائج تشغيل الدالة سالفة الذكر عن جدول التنبؤات التالي رقم (٣-٥) على النحو التالي:

جدول رقم (٣-٥): نتائج أخطاء التنبؤ برأي المراجع نتيجة استخدام الأقرب جار

	تنبؤات نسب التقرير النظيف			تنبؤات نسب التقرير النظيف مع فقرة تفسيرية			تنبؤات نسب التقرير المتحفظ		
Model	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**
Training set	100	0	0	100	0	0	100	0	0
Ten-fold cross-validation	75.4	15.6	24.6	69.9	18.5	30.1	96.8	3.4	3.2
Train (80%)-Test (20%) split	78	18.7	22	75.8	16.3	24.2	95.2	5.1	4.8
الخطأ الأول Type 1 eror	خطأ التنبؤ بتقرير نظيف في حين أنه ربما يكون نظيف مع التفسير أو متحفظ			خطأ التنبؤ بتقرير نظيف مع التفسير في حين أنه ربما يكون نظيف أو متحفظ			خطأ التنبؤ بتقرير متحفظ في حين أنه ربما يكون نظيف أو نظيف مع التفسير		
الخطأ الثاني Type 2 eror	خطأ التنبؤ بتقرير نظيف مع التفسير أو متحفظ في حين أنه ربما يكون نظيف			خطأ التنبؤ بتقرير نظيف أو متحفظ في حين أنه ربما يكون نظيف مع التفسير			خطأ التنبؤ بتقرير نظيف أو نظيف مع التفسير في حين أنه ربما يكون متحفظ		

يوضح الجدول السابق الاتفاق التام بين ما تم التنبؤ به والبيانات الفعلية المدرجة بعينة الدراسة حيث أن مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبئ به تساوي ١٠٠٪، ١٠٠٪، ١٠٠٪ على التوالي. كما تبين وجود أخطاء تنبؤ ضعيفة حيث إنها في العمود الأول كانت أخطاء التنبؤ من النوع الأول ٠٪، ١٥,٦٪، ١٨,٧٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أنه من المفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية أو تقرير متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٠٪، ٢٤,٦٪، ٢٢٪ على التوالي أي أن هناك احتمالية لأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية أو تقرير متحفظ في حين يفترض أن يكون التقرير نظيف.

وفي العمود الثاني كانت أخطاء التنبؤ من النوع الأول ٠٪، ١٨,٥٪، ١٦,٣٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير نظيفاً أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٠٪، ٣٠,١٪، ٢٤,٢٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو متحفظ في حين يفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية.

بينما في العمود الثالث كانت أخطاء التنبؤ من النوع الأول ٠٪، ٣,٤٪، ٥,١٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير متحفظ مع أنه من المفترض أن يكون التقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية. بينما كانت أخطاء التنبؤ من النوع الثاني ٠٪، ٣,٢٪، ٤,٨٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير متحفظاً.

ويلاحظ مما سبق انخفاض مستويات الخطأ في العمود الأول والثاني والثالث، ومن ثم يتبين للباحثة أن استخدام نموذج الأقرب جار يختلف مع النماذج السابقة من تقنيات البيانات الضخمة في التنبؤ برأي المراجع حيث إنه يتفق بشكل تام مع البيانات الفعلية ومن ثم فهي لا تعد وسيلة فعالة في التنبؤ برأي المراجع المستقبلي.

رابعاً: تحليل البيانات باستخدام الانحدار اللوجيستي (الأساليب التقليدية)

تعتمد طريقة الانحدار اللوجيستي على استخدام الانحدار الخطي الذي يعتمد على المعادلات الخطية من الدرجة الأولى وغالباً ما يتم استخدامه بطريقة Multinominal logistic حيث أن المتغير التابع متغير رتبي يأخذ القيم من ١ إلى ٦، وباستخدام نتائج تحليل المكونات الرئيسية الموجودة بالجدول رقم (٣-١) الخاص بمصفوفة العوامل المدورة وقد أسفرت نتائج التحليل الإحصائي عن الجدول (٣-٦) التالي:

جدول رقم (٣-٦): نتائج تشغيل نماذج الانحدار اللوجيستي للمكونات الرئيسية للأنواع الثلاثة من تقرير المراجع

نوع التقرير	Principal Components	Coef.	S. E	Z	P-value
التقرير النظيف	Intercept	3.672	4.281	2.8	0.01
	PC1	3.841	3.407	2.409	0.013
	PC2	2.266	2.484	2.714	0.023
	PC3	1.698	1.992	2.087	0.047
	PC4	3.224	1.259	2.553	0.047
	PC5	2.951	2.038	2.74	0.012
	PC6	3.546	2.476	2.5	0.006
	PC7	1.527	4.071	1.946	0.127
	PC8	2.474	4.553	1.374	0.311
	PC9	2.126	2.052	1.312	0.325
	PC10	2.46	4.555	2.375	0.001
	PC11	2.638	2.962	2.178	0.012
التقرير النظيف مع اضافة فقرة تفسيرية	Intercept	2.659	4.583	2.348	0.045
	PC1	1.587	2.468	2.647	0.048
	PC2	2.053	3.174	2.436	0.035
	PC3	2.872	2.232	2.458	0.037
	PC4	2.29	2.096	2.619	0.036
	PC5	2.057	4.399	2.507	0.029
	PC6	1.751	1.629	2.384	0.046

	PC7	3.867	3.282	1.619	0.743
	PC8	3.901	1.987	2.425	0.039
	PC9	3.857	4.571	2.389	0.028
	PC10	3.046	3.708	2.568	0.047
	PC11	4.188	2.792	1.308	0.725
التقرير المتحفظ	Intercept	2.237	4.419	2.766	0.005
	PC1	3.357	1.788	2.716	0.049
	PC2	3.648	1.93	1.497	0.648
	PC3	3.137	1.991	2.645	0.048
	PC4	1.296	2.705	2.717	0.03
	PC5	3.951	3.478	2.675	0.027
	PC6	1.597	3.313	2.02	0.001
	PC7	2.223	4.518	1.728	0.631
	PC8	1.634	4.385	2.612	0.006
	PC9	2.76	2.384	2.786	0.01
	PC10	3.051	2.781	1.974	0.523
	PC11	2.08	2.688	1.321	0.435

يتضح لدى الباحثة من خلال العرض السابق معنوية غالبية المعلمات أمام المكونات الرئيسية الإحدى عشر حيث أن قيم p-Value لغالبية العناصر بالمكونات الرئيسية كانت معنوية أقل من ٥٪، ومن ثم صلاحية المتغيرات المدرجة بالدراسة لاتخاذ القرار المتعلق برأي المراجع المتنبأ به (X1 : X24)، وبناء على ذلك يمكن استخدام الانحدار اللوجيستي للتنبؤ بشكل تقرير المراجع وقد أسفرت النتائج عن الجدول رقم (٣-٧) التالي:

جدول رقم (٣-٧): نتائج أخطاء التنبؤ برأي المراجع نتيجة استخدام الانحدار اللوجيستي

	تنبؤات نسب التقرير النظيف			تنبؤات نسب التقرير النظيف مع فقرة تفسيرية			تنبؤات نسب التقرير المتحفظ		
Model	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**	**Accuracy (%)**	**Type 1 eror (%)**	**Type 2 eror (%)**
Training set	65.1	25.9	34.9	58.8	19.2	41.2	92.7	5.4	7.3
Ten-fold cross-validation	52.9	27.1	47.1	60.6	12.5	39.4	91.3	4.1	8.7
Train (80%)-Test (20%) split	62.6	20.5	37.4	66.6	21.4	33.4	91.3	5.8	8.7
الخطأ الأول Type 1 eror	خطأ التنبؤ بتقرير نظيف في حين أنه ربما يكون نظيف مع التفسير أو متحفظ			خطأ التنبؤ بتقرير نظيف مع التفسير في حين أنه ربما يكون نظيف أو متحفظ			خطأ التنبؤ بتقرير متحفظ في حين أنه ربما يكون نظيف أو نظيف مع التفسير		
الخطأ الثاني Type 2 eror	خطأ التنبؤ بتقرير نظيف مع التفسير أو متحفظ في حين أنه ربما يكون نظيف			خطأ التنبؤ بتقرير نظيف أو متحفظ في حين أنه ربما يكون نظيف مع التفسير			خطأ التنبؤ بتقرير نظيف أو نظيف مع التفسير في حين أنه ربما يكون متحفظ		

يوضح الجدول السابق انخفاض مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبأ به حيث تساوي ٦٥,١٪، ٥٨,٨٪، ٩٢,٧٪ على التوالي. كما تبين ارتفاع مستوى أخطاء التنبؤ حيث أنها في العمود الأول كانت أخطاء التنبؤ من النوع الأول ٢٥,٩٪، ٢٧,١٪، ٢٠,٥٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أنه من المفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية أو تقرير متحفظ. بينما كانت أخطاء التنبؤ من النوع الثاني ٣٤,٩٪، ٤٧,١٪، ٣٧,٤٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية أو تقرير متحفظ في حين يفترض أن يكون التقرير نظيف.

وفي العمود الثاني كانت أخطاء التنبؤ من النوع الأول ١٩,٢٪، ١٢,٥٪، ٢١,٤٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير نظيفاً أو متحفظاً. بينما كانت أخطاء التنبؤ من النوع الثاني ٤١,٢٪، ٣٩,٤٪، ٣٣,٤٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو متحفظ في حين يفترض أن يكون التقرير نظيف مع أهمية إضافة فقرة تفسيرية.

بينما كانت أخطاء التنبؤ في العمود الثالث من النوع الأول ٥,٤٪، ٤,١٪، ٥,٨٪ على التوالي أي أن هناك احتمالية بأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير متحفظ مع أنه من المفترض أن يكون التقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية. بينما كانت أخطاء التنبؤ من النوع الثاني ٧,٣٪، ٨,٧٪، ٨,٧٪ على التوالي أي أن هناك احتمالية لأن المراجع الخارجي أخطأ في بعض المشاهدات المدرجة بعينة الدراسة وقد أبدى رأيه بتقرير نظيف أو نظيف مع أهمية إضافة فقرة تفسيرية في حين يفترض أن يكون التقرير متحفظاً.

ويلاحظ مما سبق ارتفاع مستويات الخطأ في العمود الأول والثاني، وبالتالي يتبين للباحثة أن استخدام الانحدار اللوجيستي يتسم بانخفاض مستويات الدقة في التنبؤ برأي المراجع بشكل عام ومن ثم فهو أقل دقة من تقنيات البيانات الضخمة سالفة الاستخدام ويختلف معها جوهرياً في التنبؤ برأي المراجع، ومن ثم فإن الأساليب التقليدية تنخفض في درجة دقة التنبؤ برأي المراجع. ومن ثم فهي لا تعد وسيلة فعالة في التنبؤ برأي المراجع المستقبلي.

خامساً: نتائج اختبارات الفروض الإحصائية

فيما يتعلق بنتائج اختبارات فروض الدراسة فلا يمكن الاستناد إلى الطرق الإحصائية المتعارف عليها من تحليل الانحدار المتعدد والتوصل إلى قيم المعنوية الإحصائية والتي يتم على أساسها قبول أو رفض الفرض الإحصائي للدراسة، ومن ثم فإن الباحثة ستستند في هذه الدراسة إلى قبول أو رفض الفرض بناء على التحليل المنطقي لمقارنة نتائج تشغيل التقنيات الرقمية للتنقيب عن البيانات مع الأساليب التقليدية للتنبؤ وذلك على النحو التالي:

- ينص الفرض الأول للدراسة على "توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام شجرة القرارات كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية" وبناء على النتائج سالفة الذكر يتبين انخفاض مستويات الخطأ وارتفاع مستوى دقة النتائج للتنبؤ برأي المراجع في مقابلة مستويات الدقة في التنبؤ باستخدام الأساليب التقليدية. وبالتالي، يمكن قبول الفرض الأول بوجود فروق جوهرية بين التنبؤ برأي المراجع باستخدام شجرة القرارات كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم يتفوق أسلوب شجرة القرارات على الأساليب التقليدية.

- ينص الفرض الثاني للدراسة على "توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام الشبكات العصبية كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية" وبناء على النتائج سالفة الذكر يتبين انخفاض مستويات الخطأ وارتفاع مستوى دقة النتائج للتنبؤ برأي المراجع في مقابلة مستويات الدقة في التنبؤ باستخدام الأساليب التقليدية. وبالتالي، يمكن قبول الفرض الثاني بوجود فروق جوهرية بين التنبؤ برأي المراجع باستخدام الشبكات العصبية كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم يتفوق أسلوب الشبكات العصبية على الأساليب التقليدية.
- ينص الفرض الثالث للدراسة على "توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام نموذج تعلم الآلة كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية" وبناء على النتائج سالفة الذكر يتبين انخفاض مستويات الخطأ وارتفاع مستوى دقة النتائج للتنبؤ برأي المراجع في مقابلة مستويات الدقة في التنبؤ باستخدام الأساليب التقليدية. وبالتالي، يمكن قبول الفرض الثالث بوجود فروق جوهرية بين التنبؤ برأي المراجع باستخدام نموذج تعلم الآلة كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم يتفوق أسلوب نموذج تعلم الآلة على الأساليب التقليدية.
- ينص الفرض الرابع للدراسة على "توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام تقنية الأقرب جار كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية" وبناء على النتائج سالفة الذكر يتبين تطابق النتائج للتنبؤ مع العينة الفعلية للدراسة حيث أن مستويات الدقة كانت ١٠٠ % ومن ثم فهي لم تقم بالتنبؤ وقد اعتمدت على محاكاة شكل العينة ولعل ذلك يرجع إلى ضعف العينة البالغة ١١٢ مشاهدة والمتعارف عليه عند استخدام التقنيات الرقمية ضرورة كبر حجم العينة وهو ما لم يتوافر بالبيئة المحاسبية المصرية نتيجة ضعف تطبيق التكنولوجيات الرقمية الحديثة. وبالتالي، يمكن رفض الفرض الرابع بوجود فروق جوهرية بين التنبؤ برأي المراجع باستخدام تقنية الأقرب جار كأحد تقنيات تنقيب البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم تتفوق الأساليب التقليدية على أسلوب تقنية الأقرب جار.

القسم الرابع: النتائج والتوصيات

في ضوء الدراسة النظرية والتطبيقية توصل البحث إلى مجموعة من النتائج والتوصيات، ويمكن عرضها على النحو التالي:

- **النتائج**

١- يتضح انخفاض مستويات الخطأ وارتفاع مستوى دقة النتائج للتنبؤ برأي المراجع في مقابلة مستويات الدقة في التنبؤ باستخدام الأساليب التقليدية حيث كانت نسب مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبأ به تساوي ٧٧,٥٪، ٧١,٣٪، و٩١,٥٪ على التوالي.

٢- تبين أن استخدام شجرة القرارات في التنبؤ برأي المراجع وسيلة فعالة في اكتشاف أخطاء المراجعين في إبداء الرأي الفني المحايد بشأن عدالة القوائم المالية.

٣- يؤدي ارتفاع مستويات الدقة بشكل عام لكافة الأشكال الثلاثة لرأي المراجع المتنبأ به حيث كانت الدقة تساوي ٧٨,٤٪، ٦٧,٨٪، ٩٥,٨٪ على التوالي إلى استنتاج إن استخدام الشبكات العصبية في التنبؤ برأي المراجع قد اتفق مع استخدام شجرة القرارات في اكتشاف أخطاء المراجعين في إبداء الرأي الفني المحايد بشأن عدالة القوائم المالية.

٤- قبول الفرض الثاني للدراسة "توجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام الشبكات العصبية كأحد تقنيات التنقيب عن البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية".

٥- يوجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام نموذج تعلم الآلة كأحد تقنيات التنقيب عن البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم يتفوق أسلوب نموذج تعلم الآلة على الأساليب التقليدية.

٦- يتبين تطابق النتائج للتنبؤ مع العينة الفعلية للدراسة الخاصة باستخدام أسلوب أقرب جار حيث أن مستويات الدقة كانت ١٠٠ % ومن ثم فهي لم تقم بالتنبؤ وقد اعتمدت على محاكاة شكل العينة ولعل ذلك يرجع إلى ضعف العينة البالغة ١١٢ مشاهدة والمتعارف عليه عند استخدام التقنيات الرقمية ضرورة كبر حجم العينة وهو ما لم يتوافر بالبيئة المحاسبية المصرية نتيجة ضعف تطبيق التكنولوجيات الرقمية الحديثة.

٧- رفض الفرض الرابع للدراسة الذي ينص على "يوجد فروق جوهرية بين التنبؤ برأي المراجع باستخدام تقنية الأقرب جار كأحد تقنيات التنقيب عن البيانات عن عدالة القوائم المالية وبين التنبؤ استخدام الأساليب التقليدية، ومن ثم تتفوق الأساليب التقليدية على أسلوب تقنية الأقرب جار.

- **التوصيات**

١- ضرورة قيام مكاتب المحاسبة والمراجعة المصرية بعمل دورات تدريبية على استخدام أساليب المراجعة الحديثة وخاصةً تقنيات التنقيب عن البيانات بالإضافة إلى الاستعانة بخبراء في مجال نظم المعلومات عند التعامل مع تلك التقنيات وذلك لتأهيل قدرات المراجعين في اكتشاف ممارسات الاحتيال المالي.

٢- تدريس المراجعة الالكترونية في أقسام المحاسبة بالجامعات المصرية بالإضافة إلى تدريس حالات عملية لشركات يطبق فيها تقنيات التنقيب عن البيانات حتى يتمكن الطلبة من التكيف مع التطور التقني في الشركات المصرية.

٣- مزيد من الاهتمام بالبحوث الموجهة نحو تكامل الأساليب الحديثة بما يخدم تحقيق الأهداف لتكوين معارف جديدة ومن ثم تعمل على خفض الوقت والتكلفة والجهد المبذول.

٤- توصي الباحثة بإجراء المزيد من الدراسات المستقبلية المتعلقة بأثر استخدام تقنيات التنقيب عن البيانات على كلاً من التنبؤ بالتعثر المالي للشركة في المستقبل وكفاءة الأداء المالي للشركات.

قائمة المراجع

- **المراجع العربية**

١. خميس، محمد محمد مصطفي جمعة، (٢٠٢٢)، "إطار مقترح لدور تقنية التنقيب عن العمليات في دعم مراحل نموذج DMAIC بمنهجية ستة سيجما لتحقيق استراتيجية التحسين المستمر لكايزن"، ***مجلة الإسكندرية للبحوث المحاسبية***، كلية التجارة، جامعة الإسكندرية، المجلد السادس، العدد الأول، ص ص: ٢٤١-٣٠٣.

٢. عبيد الله، فايزة، (٢٠١٦)، "التكامل بين التنقيب عن البيانات وممارسات المحاسبة الإدارية لتحسين الأداء المالي والتشغيلي للشركة"، ***مجلة الفكر المحاسبي***، كلية التجارة، جامعة عين شمس، المجلد العشرون، العدد الأول، ص ص: ٧٣-١٤٣.

- **المراجع الأجنبية**

1. Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. ***Journal of Data Science,*** 14(3), 553-569.
2. Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. ***International Journal of Accounting Information Systems***, 24, 32-58.
3. Abad, D., Sancgez-Ballesta, J. P., & Yague, J. (2017). Audit opinions and information asymmetry in the stock market. ***Accounting & Finance,*** 57(2), 565-595.
4. Awad, S. S. (2022). Using data mining tools to predivtion of going concern on auditor opinion-empirical study in Iraqi commercial. ***Academy of Accounting and Financial Studies Journal,*** 26(3).
5. Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. ***SpringerPlus,*** 5 (1), 1-16.
6. Changpetch, P., & Reid, M. (2021). Data mining techniques: Which one is your favorite?. ***Journal of Education for Business***, 96(3), 143-148.
7. Cressey, D.R. (1973). Other People's Money (Rev. Ed.), ***Montclair***.
8. Dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., & Scalabrin, E. E. (2019). Process mining techniques and applications–A systematic mapping study. ***Expert Systems with Applications***, 133, 260-295.
9. Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. ***Expert Systems with Applications***, 90, 374-393.

10. Ding, K., Lev, B., Peng, X., Sun, T., & Vasarhelyi, M.A. (2020). Machine learning improves accounting estimates: Evidence from insurance payments. ***Review of Accounting Studies***, 25(3), 1098- 1134.
11. Dahiya, N., Tyagi, L., & Sharma, N. (2021). Data mining techniques for Big data: Purpose, ***Techniques and Practical Implications***.
12. Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. ***Journal of Accounting Literature***.
13. Gaurab, T. (2015). Effective data mining for proper mining classification using neural networks. ***International Journal of Data Mining & Knowledge Management Process (IJDKP)*** Vol, 5, 112-125**.**
14. Gao, B. (2022). The use of machine learning combined with data mining technology in financial risk prevention. ***Computational Economics*** ,59(4), 1385-1405**.**
15. Geng, R. B., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. ***European Journal of Operational Research***, 241(1), 236–247.
16. Hunt, J. O., Rosser, D. M., & Rowe, S. P. (2021). Using machine learning to predict auditor switches: How the likelihood of switching affects audit quality among non-switching clients. ***Journal of Accounting and Public Policy***, 40(5), 106785.
17. Institute of Chartered Accountants of India (ICAI), (2016), Discovering fraud in forensic accounting using data mining techniques, ***International Journal of Computer Applications,*** 54(12), 37-47.
18. Jan C. L. (2018). An effective financial statement fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. ***Sustainability***, 10(2), 513.
19. Jie, S., Yujiao, Z., & Guowen, A. (2018). Research on multi class audit opinion prediction modeling based on ECOC-SVM[J], ***Friends of Accounting***, (24):87-92.
20. Kopun, D. (2018). A review of the research on data mining techniques in the detection of fraud in financial statements. ***Journal of Accounting and Management***, 8(1), 1-18.

21. Kopun, D. (2020). Application of Data Mining Techniques in the Detection of Financial Statement Fraud. ***Journal of Accounting and Management***, 10(2), 97-114.
22. Kim, Y. J., Baik, B., & Cho, S. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. ***Expert systemswith applications***, 62, 32-43.
23. Lin, C. C., Chiu, A. A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. ***Knowledge-Based Systems***, 89, 459-470.
24. Li Dan. (2017). Research on The Model of Audit Opinion Prediction Based on Grey System Theory and Neural Network [D]. ***Chongqing University of Technology***.
25. Ortmeier, C., Henningsen, N., Langer, A., Reiswich, A., Karl, A., & Herrmann, C. (2021). Framework for the integration of Process Mining into Life Cycle Assessment. ***Procedia CIRP***, 98, 163-168.
26. Oyedokun, P., & Emmanuel, G. (2015). Forensic investigation and forensic audit methodology in a computerized work environment, ***Available at SSRN 2593263***.
27. Papík, M., & Papíková, L. (2022). Detecting accounting fraud in companies reporting under US GAAP through data mining***. International Journal of Accounting Information Systems***, 45, 100559.
28. Popov, A. A., Masaev, S. N., Edimichev, D. A., & Pomolotova, O. V. (2020, November). Analytical treatment of transport logistics business processes by the Process Mining technology. ***In Journal of Physics: Conference Series*** (Vol. 1679, No. 2, p. 022095). IOP Publishing.
29. Sánchez-Serrano, J. R., Alaminos, D., García-Lagos, F., & Callejón-Gil, A. M. (2020). Predicting audit opinion in consolidated financial statements with artificial neural networks. ***Mathematics,*** 8(8), 1288.
30. Saglar, J., & Ilker, K. E. F. E. (2021). A review on data mining methods used in internal audit and external audit. ***EKEV Akademi Dergisi,*** (88), 259-274.
31. Salehi, M., Shiri, M. M., & Pasikhani, M. B. (2016). Predicting corporate financial distress using data mining techniques: An application in Tehran Stock Exchange. ***International Journal of Law and Management***.
32. Salehi, M., & Dehnavi, N. (2018). Audit report forecast: an application of nonlinear grey Bernoulli model. ***Grey Systems:theory and Application***.

33. Stanisic, N., Radojevic, T., & Stanic, N.(2019). Predicting the type of auditor opinion: Statistics, machine learning, or a combination of the two?. ***Machine learning, or a Combination of the Two***, 1-58.
34. Wu, H. P., & Li, L. (2021). The BP neural network Adam optimizer for predicting audit opinions of listed companies. ***IAENG International Journal of ComputerScience***,48 (2).
35. Werner, M., Wiese, M., & Maas, A. (2021). Embedding process mining into financial statement audits. ***International Journal of Accounting Information Systems***,41 : 100514**.**
36. Yaşar, A., Yakut, E., & Gutnu, M. M. (2015). Predicting qualified audit opinions using financial ratios: Evidence from the Istanbul Stock Exchange. ***International Journal of Business and Social Science,*** 6(8), 57-67..
37. Zeng, S., Li, Y., & Li, Y. (2022). Research on Audit Opinion Prediction of Listed Companies Based on Sparse Principal Component Analysis and Kernel Fuzzy Clustering Algorithm. ***Mathematical Problems in Engineering***, 2022.
38. Zeng, J., & Yang, Y. (2021). Study on Multi-category Audit Opinion Prediction--Based on 1D-DRCNN Model. ***Frontiers***, 2(11).
39. Zhiheng Z., Dan L., Yu L. (2017). Research on the Model of Audit Opinion Prediction Based on Integration of Neighborhood Rough Sets an Neural Network[J], ***Journal of Chongqing University of Technology (Social Science)***, 31(08):37-43.
40. Zhou, L. G., Lu, D., & Fujita, H. (2015). The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches**.** ***Knowledge-Based Systems***, 85, 52–61**.**
41. Zdolsek, D., Jagric, T., & Kolar, I. (2022). Auditor's going-concern opinion prediction: the case of Slivenia. ***Economic Research Ekonomska Istrazivanja***, 35(1), 106-121.

Reflections of using Data Mining Techniques in Forecasting the Opinion of The External Auditor And its Impact on the Fairness of The Financial Statements: An Applied Study

Dr. Heba Gamal Hashim

Abstract:

The research aimed to study the repercussions of using data mining techniques in predicting the auditor's opinion and its impact on the fairness of the financial statements. The external auditor in the fairness of the financial statements, and the study community is limited to all companies that have recently converted to digital systems, especially those registered in the Egyptian Stock market. Companies to apply digital systems, where their number reached 25 companies with a total of 112 views, and the research found that there are fundamental differences between prediction the opinion of the auditor using both neural networks, decision trees, and machine learning model as data mining techniques for the fairness of financial statements, and prediction using traditional methods, while it turned out that there were no significant differences between the prediction of the auditor's opinion using the nearest neighbor technique as data mining techniques for the fairness of financial statements and between forecasting results with the actual sample of the study, as the levels of accuracy were 100%, and therefore they did not make predictions.

Keywords: Data mining, Decision tree, Neural networks, Nearest neighbors, Machine learning, prediction the opinion of the auditor.