تحديد العتبة المثلى للمعاملات الشاذة باستخدام التعلم الآلي لتحسين أمن سلاسل الكتل

علي عطية[1] أ.د.ماهر عباس[2] د.وسيم رمضان[3]

(1) طالب دكتوراه في قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص.

(2) أستاذ دكتور في قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص.

(3) دكتور في كلية الهندسة الزراعية، جامعة حمص.

الملخص

إن التقارب بين اكتشاف الشذوذ وتحديد العتبة في تقنية البلوك تشين يخلق إطاراً أمنياً متطوراً يستفيد من التعلم الآلي لحماية الشبكات اللامركزية. يبدأ هذا النهج المتكامل بجمع ومعالجة بيانات معاملات البلوك تشين مسبقاً، حيث تحدد الخوارزميات المتخصصة الأنماط التي تنحرف عن السلوك المتوقع. تتطلب معاملات البلوك تشين مراقبة متقدمة لضمان الأمن والسلامة مع ظهور التعلم الآلي كأداة قوية لتحديد الأنشطة المشبوهة. من خلال تحليل الأنماط في بيانات المعاملات، يتم تحديد عتبات دقيقة للكشف عن الشذوذ. حيث يكمن مفتاح النجاح في ضبط هذه العتبات بناءً على المعاملات الطبيعية والشاذة، مما يسمح للنظام بالتمييز بين المعاملات الطبيعية والمعاملات الاحتيالية. يهدف هذا البحث إلى تحديد العتبة المثلى للمعاملات الشاذة في سلاسل الكتل باستخدام تقنيات التعلم الآلي، مما يسهم في تعزيز أمان الشبكات اللامركزية. تم تدريب خمس خوارزميات تصنيف، هي: شجرة القرار، الغابة العشوائية، الانحدار اللوجستي، أقرب جار، ومصنف الانتخاب، وذلك

باستخدام بيانات معاملات من مجموعة Metaverse Financial Transactions Dataset التي تحتوي على 78,600 سجل .أُجريت ثلاث تجارب باستخدام تقنيات مختلفة لموازنة البيانات، وقيّم الأداء بناءً على دقة التصنيف (Accuracy) ومعامل .F1-score أظهرت النتائج أن عتبة الخطورة المثلى هي فوق 85(High Risk) ، حيث تفوقت تجربةRandomOverSampler ، محققة دقة 99.99% و F1-score = 0.9 ، مع تحسين أداء خوارزميات الانحدار اللوجستي وأقرب جار بالنسبة لمقياس f1score بمقدار 13% و 39 %على التوالي .تدعم هذه النتائج استخدام تقنيات موازنة البيانات في تعزيز دقة اكتشاف الشذوذ، مما يمهد الطريق لتطوير أنظمة أكثر كفاءة لمكافحة الاحتيال في شبكات سلاسل الكتل.

الكلمات المفتاحية: سلاسل الكتل، كشف الشذوذ، التعلم الالي، عتبة المعاملات الشاذة، التصنيف.

Determining The Optimal Threshold for Anomalous Transactions Using Machine Learning to Improve Blockchain Security

Abstract

The convergence of anomaly detection and thresholding in blockchain technology creates a sophisticated security framework that leverages machine learning to protect decentralized networks. This integrated approach begins with the pre-collection and processing of blockchain transaction data, where specialized algorithms identify patterns that deviate from expected behavior. Blockchain transactions require advanced monitoring to ensure security and integrity, with machine learning emerging as a powerful tool to identify suspicious activities. By analyzing patterns in transaction data, precise thresholds for anomaly detection are determined. The key to success lies in adjusting these thresholds based on normal

and abnormal transactions, allowing the system to distinguish between normal and fraudulent transactions.

This research aims to identify the optimal threshold for anomalous transactions in blockchains using machine learning techniques, which contributes to enhancing the security of decentralized networks. Five classification algorithms, namely decision tree, random forest, logistic regression, nearest neighbor, and election classifier, were trained using transaction data from the Metaverse Financial Transactions Dataset containing 78,600 records.

Three experiments were conducted using different data-balancing techniques, and performance was evaluated based on classification accuracy (Accuracy) and F1-score. The results showed that the optimal risk threshold is above 85 (High Risk), where the RandomOverSampler experiment outperformed, achieving 99.99% accuracy and an F1-score of 0.9, while improving the performance of the logistic regression and nearest neighbor algorithms with respect to the F1-score metric by 13% and 39%, respectively. These results support the use of data balancing techniques to enhance the accuracy of anomaly detection, paving the way for the development of more efficient anti-fraud systems in blockchain networks

Keywords: Blockchains, Anomaly Detection, Machine Learning, Anomaly Transaction Threshold, Classification.

1. مقدمة

لقد أحدث ظهور تقنية البلوك تشين ثورة في المجال التقني، حيث وعدت بشفافية وأمان غير مسبوقين من خلال بنيتها اللامركزية. ومع ذلك، فإن هذه التكنولوجيا تقدم تحديات فريدة من نوعها لمنع الاحتيال. في حين تقدم تقنية البلوك تشين العديد من المزايا، فإن طبيعتها اللامركزية والمجهولة،

إلى جانب ثبات المعاملات، تخلق أرضاً خصبة للجهات الخبيثة. وتزداد هذه المشكلة خطورة مع ارتفاع خسائر الاحتيال في الأنظمة المالية الرقمية، والتي تُقدر بالمليارات سنوي[19] . إن طرق الكشف عن الاحتيال التقليدية، المصممة للأنظمة المركزية، غير مجهزة للتعامل مع التعقيدات المحددة لتقنية البلوك تشين. يكمن أحد التحديات الأساسية في تحديد العتبة المثلى التي تفصل بين المعاملات الطبيعية والمشبوهة بدقة.. وهذا أكثر تعقيداً مما يبدو، حيث تُظهِر معاملات blockchain اختلافات طبيعية يجب تمييزها عن السلوك الشاذ.

الطبيعة الديناميكية لكل من الاستخدام المشروع لـ blockchain والنشاط الاحتيالي تزيد من تعقيد هذه المهمة، حيث تظهر حالات استخدام مشروعة جديدة باستمرار، بينما يكيف الجهات الخبيثة استراتيجياتها باستمرار للتهرب من الكشف. وهذا يتطلب حلولاً تكيفية قادرة على التعلم والتطور جنباً إلى جنب مع التكنولوجيا.

إن التوفيق بين تقليل الإيجابيات الكاذبة (المعاملات المشروعة التي يتم وضع علامة عليها على أنها مشبوهة) والسلبيات الكاذبة (النشاط الاحتيالي المفقود) يتطلب دراسة متأنية، وندرة البيانات اللازمة لتدريب نماذج التعلم الآلي وعدم توازن البيانات يزيدان من تفاقم المشكلة.

جلبت التطورات الأخيرة في التعلم الآلي حلولاً واعدة لهذه المشكلة المعقدة. يمكن لهذه الأنظمة المتطورة تحليل الأنماط المعقدة عبر أبعاد متعددة من بيانات المعاملات، من القيم النقدية إلى الأنماط الزمنية والعلاقات الشبكية. مما يوفر مساراً محتملاً نحو اكتشاف الاحتيال بشكل أكثر قوة وتكيفاً في مشهد blockchain المتطور باستمرار. على الرغم من الجهود السابقة في استخدام التعلم الآلي للكشف عن الشذوذ، لا تزال هناك تحديات في تحديد العتبة المثلى للمعاملات الشاذة، حيث تختلف العتبات وفقاً لأنماط المعاملات والسياقات المختلفة. يهدف هذا البحث إلى معالجة هذه الفجوة من خلال تحديد العتبة المثلى للمعاملات الشاذة في سلاسل الكتل باستخدام تقنيات التعلم الآلي، مما يسهم في تعزيز أمان الشبكات اللامركزية. لتحقيق ذلك تم تدريب خمس خوارزميات

تصنيف، هي: شجرة القرار، الغابة العشوائية، الانحدار اللوجستي، أقرب جار، ومصنف الانتخاب وتم أُجراء ثلاث تجارب باستخدام تقنيات مختلفة لموازنة البيانات، وتقييّم الأداء بناءً على دقة التصنيف Accuracy ومعامل .F1-score

2. مشكلة البحث

تتمثل مشكلة البحث في تحديد عتبة المعاملات الشاذة في سلاسل الكتل باستخدام التعلم الآلي حول إنشاء حد معين يميز بدقة بين المعاملات الطبيعية والشاذة نظراً لكون اختيار العتبة الخاطئة قد يؤدي إلى تصنيف خاطئ للمعاملات، مما يعرقل عمليات كشف الاحتيال أو يزيد من حجب المعاملات السليمة عن طريق الخطأ. يشمل هذا التحدي على عدة عوامل، بداية بالصعوبة المتمثلة في تطوير نماذج التعلم الآلي التي يمكنها تحديد الشذوذ الحقيقي بشكل فعال مقابل الاختلافات الطبيعية في أنماط المعاملات، والتعقيد المنهجي لموازنة الحساسية والخصوصية في الكشف.

3. هدف البحث

يهدف هذا البحث إلى تحديد العتبة المثلى للمعاملات الشاذة في سلاسل الكتل باستخدام تقنيات التعلم الآلي، مما يسهم في تحسين كشف المعاملات الاحتيالية وتعزيز أمن الشبكات اللامركزية. لتحقيق ذلك، يتم تدريب نماذج تصنيف مختلفة على مجموعة بيانات سلاسل الكتل، مع دراسة تأثير تقنيات موازنة البيانات على دقة التصنيف. سيتم تحليل أداء النماذج وتحديد العتبة المثلى التي تضمن تحقيق أعلى كفاءة في كشف الشذوذ مع تقليل الأخطاء في التصنيف.

4. سلاسل الكتل

سلاسل الكتل هي دفتر رقمي لامركزي يخزن السجلات بشكل آمن عبر شبكة الاجهزة بطريقة شفافة وغير قابلة للتغيير ومقاومة للتلاعب. تتكون من كتل من البيانات مرتبطة ببعضها البعض في سلسلة زمنية. (مرجع)

5. التعلم الالي

هو فرع من فروع الذكاء الاصطناعي (AI) وعلوم الكمبيوتر الذي يركز على استخدام البيانات والخوارزميات لتقليد الطريقة التي يتعلم بها البشر، وتحسين دقتها تدريجياً. يعد التعلم الآلي مكوناً مهماً في المجال المتنامي لعلوم البيانات. من خلال استخدام الأساليب الإحصائية، يتم تدريب الخوارزميات على إجراء تصنيفات أو تنبؤات في مشاريع التنقيب عن البيانات. تؤدي هذه الأفكار بعد ذلك إلى اتخاذ القرار داخل التطبيقات والشركات، مما يؤثر بشكل مثالي على مقاييس النمو الرئيسية [12].

6. الشذوذ

يعرف الشذوذ بالشكل العام بأنه أنماط في البيانات لا تتوافق مع مفهوم محدد جيداً للسلوك الطبيعي.ويمكن تعريف أنظمة كشف الشذوذ على أنها أجهزة أو برامج تقوم بمراقبة البيانات المتعلقة بمجال معين لتحديد الحالات الشاذة والمختلفة عن السلوك الطبيعي [21]

6.1 الشذوذ في سلاسل الكتل

يشير الشذوذ إلى أي انحراف أو مخالفة أو قيمة شاذة عن نمط أو سلوك متوقع أو طبيعي. في سياق سلاسل الكتل، يمكن أن يشير الشذوذ إلى أنشطة غير عادية أو سلوك مريب أو تهديدات أمنية محتملة داخل شبكة سلاسل الكتل. يتضمن اكتشاف الشذوذ في سلاسل الكتل عملية تحديد مثل هذه الشذوذ. قد يكون اكتشاف الشذوذ في بيئة سلاسل الكتل أمراً صعباً بسبب عدة أسباب [6]: 1. الطبيعة الموزعة وغير القابلة للتغيير: سلاسل الكتل عبارة عن دفتر حسابات لامركزي وموزع، حيث يحتفظ كل مشارك بنسخة من سلاسل الكتل بالكامل. بمجرد تسجيل معاملة وإضافتها إلى كتلة، تصبح غير قابلة للتغيير عملياً. وهذا يعني أنه لا يمكن إزالة أي معاملة شاذة أو احتيالية بمجرد إضافتها إلى سلاسل الكتل بسهولة. يجب أن تأخذ تقنيات اكتشاف الشذوذ في الاعتبار هذه الخاصية الفريدة لسلسلة الكتل أثناء تحديد الشذوذ والتخفيف منه. 2. كمية هائلة من البيانات: تولد

شبكات سلسلة الكتل كمية هائلة من البيانات بسبب التسجيل المستمر للمعاملات. يمكن أن يكون تحليل ومعالجة هذا الحجم الكبير من البيانات في الوقت الفعلي مكثفاً من الناحية الحسابية. يجب أن تكون طرق اكتشاف الشذوذ قابلة للتطوير وفعالة بما يكفي للتعامل مع الإنتاجية العالية لمعاملات سلسلة الكتل. 3. أنماط المعاملات المعقدة: يمكن أن تظهر معاملات سلسلة الكتل أنماطاً وتبعيات معقدة. يمكن أن تظهر الشذوذ في أشكال مختلفة، مثل أحجام المعاملات غير الطبيعية، وأنواع المعاملات غير العادية، والتغيرات المفاجئة في سلوك الشبكة، أو الأنشطة الضارة. يجب أن تكون خوارزميات اكتشاف الشذوذ قادرة على التقاط وفهم أنماط المعاملات المعقدة هذه للكشف عن الشذوذ بشكل فعال.

6.1.1 أنواع الشذوذ في سلاسل الكتل

يمكن أن تحدث عدة أنواع من الشذوذ في سلاسل الكتل، والتي يمكن أن تكون مؤشراً على نشاط احتيالي أو ضار. فيما يلي بعض الأنواع الشائعة من الشذوذ في سلاسل الكتل [6] : 1. هجمات الإنفاق المزدوج: يحدث هجوم الإنفاق المزدوج عندما يحاول الفرد إنفاق نفس العملة المشفرة مرتين. يمكن القيام بذلك عن طريق إنشاء معاملة وهمية، والتي يتم بثها بعد ذلك إلى الشبكة. إذا تم قبول المعاملة المزيفة وإضافتها إلى سلاسل الكتل قبل المعاملة المشروعة، يمكن للفرد أن ينفق نفس العملة المشفرة مرتين فعلياً. 2. ازدحام الشبكة: يحدث ازدحام الشبكة عندما يتجاوز عدد المعاملات التي يتم بثها إلى الشبكة قدرتها على المعالجة. يمكن أن يؤدي هذا إلى تأخيرات في معالجة المعاملات وزيادة رسوم المعاملات وربما رفض المعاملات الصالحة. 3. المعاملات الاحتيالية: تحدث المعاملات الاحتيالية عندما يحاول الفرد التلاعب بالبلوك تشين عن طريق إنشاء معاملات وهمية أو تعديل المعاملات الموجودة. يمكن القيام بذلك في محاولة لسرقة العملة المشفرة أو الحصول على وصول غير مصرح به إلى الشبكة. 4. الأخطاء في العقود الذكية: العقود الذكية هي عقود ذاتية التنفيذ تعمل على سلاسل الكتل .إذا كانت هناك أخطاء في كود العقد الذكي، فقد يؤدي هذا

إلى سلوك غير متوقع، والذي يمكن استغلاله من قبل الجهات الخبيثة. 5. هجماتSybil: حدث هجمات Sybil عندما ينشئ فرد هويات مزيفة متعددة، أو Sybils على الشبكة. يمكن القيام بذلك في محاولة للتلاعب بآلية الإجماع والسيطرة على الشبكة.

7. الدراسة المرجعية

تتضمن أدبيات الدراسة مجموعة من الطرق لتحديد عتبة المعاملات الشاذة والطبيعية بدءاً من التحليل الاحصائي وهندسة الميزات حيث تعرف هندسة الميزات بأنها عملية إنشاء مجموعة ميزات باستخدام خصائص البيانات التي تعزز أداء خوارزميات تعلم الآلة [20] ووصولاً الى طرق التعلم الآلي والتعلم العميق وذلك من اجل تحديد العتبة الأمثل لفصل القيم الشاذة عن القيم الطبيعية في سلاسل الكتل.

تسعى الدراسة [14] إلى دمج تقنيات الذكاء الاصطناعي القابلة للتفسير وقواعد الشذوذ في مصنفات المجموعة القائمة على الشجرة للكشف عن معاملات البيتكوين الشاذة. يتم استخدام طريقة التفسير الإضافي لـ Shapley (SHAP) لقياس مساهمة كل ميزة. علاوة على ذلك، يتم تقديم قواعد لتفسير ما إذا كانت معاملة البيتكوين شاذة أم لا. ذلك باستخدام التمثيلات الشجرية لفهم سبب تصنيف حالات معينة على أنها شذوذ. لتوفير تفصيلاً خطوة بخطوة لعملية اتخاذ القرار وتسليط الضوء على الميزات والعتبات التي لعبت دوراً حاسماً في اتخاذ القرار، كما تم تقديم خوارزمية أخذ عينات ناقصة تسمى XGBCLUS، مصممة لموازنة بيانات المعاملات الشاذة وغير الشاذة. ومقارنة هذه الخوارزمية بتقنيات أخرى شائعة الاستخدام لتقليل أخذ العينات وزيادة أخذ العينات.

تم في البحث [15] تنفيذ وتقييم خوارزمية اكتشاف الشذوذ القائمة على K–means وهي خوارزمية تجميع شائعة يمكن تطبيقها على اكتشاف الشذوذ. يكتشف K–means الشذوذ من خلال عتبتين بعد إجراء التجميع. المسافة بين عقدة الهدف ومركز ثقل المجموعة التي تنتمي إليها العقدة.

يتم حساب الإحداثيات من كل كمية ميزة. وحجم المجموعة التي تنتمي إليها العقدة. إذا تجاوزت الحد الأقصى أو كانت أصغر من الحد الأقصى، تعتبر العقدة شاذة.

قدمت ورقة البحث [16] نظاماً موثوقاً به للكشف عن الشذوذ، تم استخدام Deep Autoencoder في تنفيذ نظام الكشف عن الشذوذ، على وجه التحديد، يتم البحث عن ضبط المعلمات المحددة جيداً، حتى عندما يؤدي هذا إلى زيادة تعقيد المصنف بشكل كبير، شريطة أن يكون المصنف قادراً على تصنيف البيانات المستهدفة بشكل صحيح. كلما كان النموذج أكثر تعقيداً، كلما كان نطاق التصنيف في مساحة البيانات المستهدفة أصغر، وانخفض احتمال تصنيف القيم المتطرفة بشكل صحيح. بشكل أساسي ومن أجل التنبؤ بما إذا كانت المعاملة الجديدة طبيعية أم احتيالية، يتم حساب خطأ إعادة بناء Deep Autoencoder من تفاصيل المعاملة نفسها. إذا كان الخطأ أكبر من عتبة محددة مسبقاً، فسوف يتم تصنيفه على أنه غير طبيعي، عن طريق استخدام عتبة مجموعة البيانات المثلى وهي طريقة لرفض عينات الفئة الإيجابية (العادية) القريبة من حافة فصل الفئة. قد تؤدي هذه العينات إلى تصنيف مضلل بحيث لا يتم أخذها في الاعتبار.

تقترح الدراسة [17] خوارزمية اكتشاف التطفل التعاوني القائمة على سمة التجميع لـ Blockchain والتي يمكنها التعرف بسرعة على سمات التجميع في معاملات بيانات .Blockchain يتم أيضاً إنشاء نموذج رياضي، بناءً عليه تم تصميم بروتوكول اكتشاف التطفل القائم على سمة التجميع للكشف بدقة عن عدد سمات التجميع. تتمتع الخوارزمية المقترحة بدقة التعرف الممتازة وتكلفة الوقت، يتم النظر في معامل الترجيح لخصائص البيانات وعلاقة مطابقة التشابه في النمط الطبيعي في هذا النهج. تتم مطابقة مواضع المجموعة لتقليل تكلفة اكتشاف بيانات التجميع، ويتم التعرف على خصائص التجميع بدقة.

في [2] تم تطوير مخطط للكشف عن الشذوذ الذي يميز معلمات blockchain على أنها طبيعية أو شاذة باستخدام التحليل الإحصائي وطرق التجميع الهرمي. تم استخدام الهيستوغرام وتوزيعات

الاحتمالات والرسوم البيانية الصندوقية للبيانات لتقدير العتبات للقيم المتطرفة التي قد تشير إلى الهجمات. تم استخدام العتبات التي تم الحصول عليها من المخططات الشجرية لتشكيل مجموعات ومجموعات فرعية بناءً على بنية البيانات الهرمية؛ تعتبر مؤشرات نقاط البيانات التي لا تقع ضمن العتبة شاذة وغير مدرجة في المجموعات.

يركز البحث [3] على اكتشاف الشذوذ داخل مجموعة بيانات معاملات البيتكوين. وتهدف إلى تحسين فهم سلوكيات الشذوذ داخل شبكات البلوك تشين واستكشاف كيفية تحديد هذه الشذوذ بشكل فعال. علاوة على ذلك، تحاول تعزيز الأساليب الحالية للكشف عن الشذوذ الثابت وتوفير تحليل نظري شامل لتقنيات الكشف عن الشذوذ الديناميكي. يتم تدريب النموذج باستخدام إما دالة الخسارة متوسط الخطأ التربيعي أو متوسط الخطأ المطلق بعد التدريب، يقوم النموذج بإجراء تنبؤات على مجموعة الاختبار. ثم يتم حساب خطأ كل معاملة من خلال مقارنة القيم المتوقعة والفعلية. يتم تصنيف المعاملات التي تتجاوز أخطاءها عتبة النسبة المئوية 95 على أنها شذوذ. تضمن هذه العتبة أن يركز النموذج على تحديد أكثر التناقضات تطرفًا، والتي تمثل المعاملات التي تنحرف بشكل كبير عن الأنماط المتوقعة التي يلتقطها النموذج. تم اختيار هذه العتبة لأنها توازن بين الحاجة إلى اكتشاف الشذوذ، والتي تعد نادرة بحكم التعريف، مع الحاجة إلى الحد من الإيجابيات الكاذبة.

تهدف الدراسة [12] إلى اكتشاف الجهات الخبيثة العاملة على شبكة Ethereum وتصنيف الهجمات بناءً على أفعالها. لتحقيق هدف البحث هذا، تم إنشاء مجموعة بيانات جديدة من خلال دمج البيانات حول الجهات الخبيثة المشاركة في أنشطة Ethereum غير المشروعة. تم استخراج الميزات الرئيسية من مجموعة البيانات هذه باستخدام تقنيات اختيار الميزات المتقدمة، بما في ذلك تحليل المكونات الأساسية ((PCA، وكسب المعلومات. تم تطبيق مصنفات التعلم الآلي مثل LGBM وXGBoost وRandom Forest وExtra Tree وBagging و K-Nearest Neighbors لتحديد وتصنيف الجهات الخبيثة بشكل فعال. تؤكد النتائج، التي حققت معدل دقة 98٪، فعاليةInformation Gain عند دمجها مع LGBM و .XGBoostوالجدير بالذكر أن

XGBoost يثبت كفاءته من خلال إكمال التحليل في 13.72 ثانية فقط. بالإضافة إلى تحديد الأنشطة الاحتيالية. يركز الحل المقترح في [7] على استخدام نهج ديناميكي حيث يتم استخدام سلوك التشغيل الطبيعي لسلسلة كتل Ethereum لتدريب خوارزميات التعلم الآلي وسيتم وضع علامة على أي انحراف على أنه شذوذ واكتشافه بواسطة النظام. تم استخدام أربع خوارزميات للتعلم الآلي بما في ذلك K-Nearest Neighbors (KNN) و Gaussian Naive Bayes (GaussianNB) و Random Forest و Stochastic Gradient Descent (SDG) لتدريب والتحقق من دقة الحل المقترح. أظهرت النتائج التجريبية أن خوارزمية الغابة العشوائية قدمت أفضل دقة بنسبة 99.84٪ مقارنة بخوارزميات التعلم الآلي الأخرى.

يوضح الجدول 1 مقارنة الدراسات المرجعية، من الواضح ان الدراسة الحالية تعالج مشكلة البيانات الغير المتوازنة وتعتمد على تحديد عتبة الشذوذ بناء على خصائص المعاملة وسلوك المستخدم ووصلت دقة كشف الشذوذ الى 99.99 %:

جدول 1 مقارنة الدراسات المرجعية

الدراسة	استخدام تقنيات موازنة	دراسة عتبة شذوذ	تحديد عتبة الشذوذ بناءعلى خصائص المعاملة وسلوك المستخدم	طريقة تحديد عتبة الشذوذ	النتائج
[14]	✓	✓	✖	أهمية الميزة باستخدام شجرة القرار	دقة كشف شذوذ تصل الى 0.97
[15]	✖	✓	✖	K-means	معدل fp يصل الى 0.03
[16]	✖	✓	✖	ODT	دقة كشف شذوذ تصل الى 0.99
[17]	✖	✓	✖	التجميع	FPR=0.92
[2]	✖	✓	✖	التجميع والطرق الإحصائية	------

[3]	✖	✓	✖	LSTML Model Traning	(MSE) مساوية ل 85255
الدراسة الحالية	✓	✓	✓	اعتماد عتبة رقمية للشذوذ بناء على خصائص المعاملة وسلوك المستخدم	دقة كشف شذوذ تصل الى 99.99

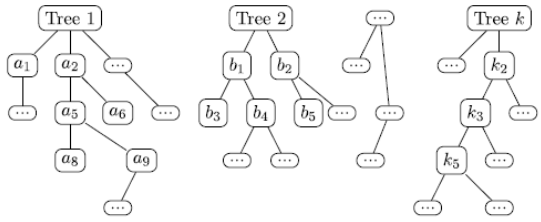
8. الخوارزميات المستخدمة

8.1 شجرة القرار

تعد خوارزمية شجرة القرار خوارزمية مستخدمة على نطاق واسع للتصنيف، والتي تستخدم قيم السمات لتقسيم مساحة القرار إلى مساحات فرعية أصغر بطريقة تكرارية، يمكن تمثيل عمليات اتخاذ القرار مثل كشف الشذوذ بيانياً كشجرة [5].

8.2 الغابة العشوائية

هي مجموعة من أشجار القرار التي تم إنشاؤها من مجموعة البيانات يتم تجميع كل شجرة على حدة لإنتاج تنبؤ فريد بالإجماع على كشف الشذوذ. يتم تمثيل الفكرة الأساسية للغابات العشوائية بشكل تخطيطي في الشكل 1:



شكل1 الفكرة الأساسية للغابة العشوائية

8.3 خوارزمية اقرب جار

تجد الخوارزمية أكثر الملاحظات تشابهاً مع تلك التي يجب أن تتنبأ بها والتي تستمد منها حدساً جيداً للإجابة المحتملة عن طريق حساب متوسط القيم المجاورة، أو عن طريق اختيار فئة الإجابة الأكثر شيوعاً فيما بينها. يمكن لـ KNN التعامل بسهولة مع مئات التسميات [9]. عادة، يعمل KNN على معرفة جيران الملاحظة بعد استخدام مقياس المسافة مثل الإقليدية (الخيار الأكثر شيوعاً)[8] . على سبيل المثال في حال كانت نقطة بيانات معينة يجب التنبؤ بها بأنها شاذة او طبيعية فيتم النظر الى أقرب جيرانها لتصنيف هذه النقطة.

8.4 الانحدار اللوجستي

تقدر الانحدارات اللوجستية احتمالية وقوع حدث ما، مثل التصويت أو عدم التصويت، استنادًا إلى مجموعة بيانات معينة من المتغيرات المستقلة. غالباً ما يستخدم هذا النوع من النماذج الإحصائية (المعروف أيضاً باسم نموذج اللوغاريتم) للتصنيف والتحليلات التنبؤية. نظراً لأن النتيجة هي احتمالية، فإن المتغير التابع محصور بين 0 و1. في الانحدار اللوجستي [18].

8.5 مصنف الانتخاب voting

تم دمج المصنفات الثلاثة الأفضل وهي KNN,RF,DT للحصول على مصنف انتخابي بين هذه الخوارزميات بحيث يتم تصنيف البيانات الشاذة والطبيعة بناء على غالبية الأصوات بدلاً من الاعتماد على خوارزمية واحدة.

9. معايير الأداء

تم الاعتماد على المعايير التالية نظراً لكون المشكلة هي مشكلة تصنيف البيانات على انها طبيعية او شاذة وذلك في مرحلة اختبار الخوارزميات.

- **دقة التصنيف accuracy**

دقة التصنيف هي ما نعنيه عادةً عندما نستخدم مصطلح الدقة. إنها نسبة عدد التنبؤات الصحيحة إلى العدد الإجمالي لعينات الإدخال [10]. وتعطى الدقة بالمعادلة 1:

$$\text{Accuracy}= \frac{\text{number of correct predictions}}{\text{total number of predictions}} \ldots(1)$$

- **مصفوفة الارتباك Confusion Matrix**

مصفوفة الارتباك كما يوحي الاسم تعطي مصفوفة كإخراج وتصف الأداء الكامل للنموذج. يوضح الشكل 2 مصفوفة الارتباك:

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

شكل 2 مصفوفة الارتباك [11]

يمكن حساب دقة المصفوفة بأخذ متوسط القيم الموجودة عبر "القطر الرئيسي" كما هو موضح بالمعادلة 2:

$$\text{Accuracy} = \frac{\text{TruePositive+TrueNegative}}{\text{TotalSample}} \ldots(2)$$

- **الدقة Precision**

هي عدد النتائج الإيجابية الصحيحة مقسوماً على عدد النتائج الإيجابية التي تنبأ بها المصنف كما هو موضح في المعادلة 3، بعبارات أبسط، الدقة هي النسبة بين الإيجابيات الحقيقية وجميع الإيجابيات.

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TotalPredicted Positive}} \ldots(3)$$

- **الاسترجاع Recall**

هو عدد النتائج الإيجابية الصحيحة مقسوماً على عدد جميع العينات ذات الصلة (جميع العينات التي كان ينبغي تحديدها على أنها إيجابية). الاسترجاع هو مقياس النموذج الذي يحدد الإيجابيات الحقيقية بشكل صحيح. رياضياً موضح بالمعادلة 4:

$$Recall=\frac{TruePositive}{TruePositive+FasleNegative} \quad ...(4)$$

- **F1 Score**

F1 Score هو المتوسط التوافقي بين الدقة والاسترجاع. نطاق نقاط F1 هو [0، 1]. يخبر بمدى دقة المصنف(عدد الحالات التي يصنفها بشكل صحيح). رياضياً f1score موضح بالمعادلة 5.

$$F1score=\frac{2*Precision*Recall}{Precision+Recall} \quad ...(5)$$

10. الاعداد التجريبي

يجدر الذكر بأنه تم استخدام جهاز حاسوب محمول بالمواصفات التالية كما يوضح الجدول 2.

جدول 2 مواصفات الحاسوب المستخدم

المواصفة	القيمة
تردد المعالج	2.5 HZ
الذاكرة العشوائية RAM	8 GB
عدد الأنوية	3
جيل الحاسوب	7

.11 التطبيق العملي

11.1 مجموعة البيانات المستخدمة

تم استخدام مجموعة بيانات Metaverse Financial Transactions Dataset توفر مجموعة البيانات هذه المعاملات القائمة على تقنية سلاسل الكتل، بهدف توفير مجموعة غنية ومتنوعة وواقعية من البيانات لتطوير واختبار نماذج الكشف عن الشذوذ، وتحليل الاحتيال[4].

تم تصميم مجموعة البيانات هذه لمجموعة واسعة من الاستخدامات، بما في ذلك على سبيل المثال لا الحصر:

- الكشف عن الشذوذ وتحليل الاحتيال في معاملات سلاسل الكتل.
- البحث في إدارة الأصول الرقمية الآمنة والشفافة.
- تطوير واختبار الخوارزميات لتقييم المخاطر والتحقق من المستخدم.

تتضمن مجموعة البيانات 78600 سجل، يمثل كل منها معاملة بالميزات التالية الموضحة بالجدول 3.

جدول 3 ميزات مجموعة البيانات

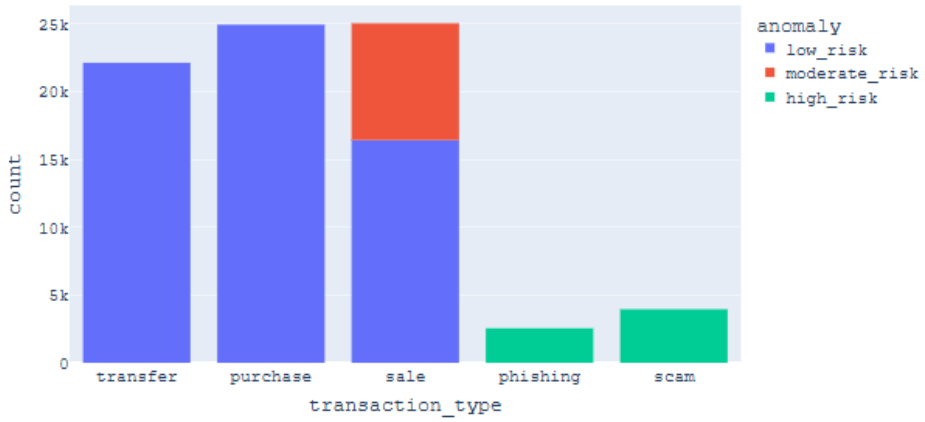
الميزة	الشرح
Timestamp	تاريخ ووقت المعاملة.
Hour of Day	جزء الساعة من الطابع الزمني للمعاملة.
Sending Address	عنوان المرسل.
Receiving Address	عنوان المستقبل.
Amount	مبلغ المعاملة.
Transaction Type	تصنيف المعاملة (على سبيل المثال، التحويل، البيع، الشراء، الاحتيال، التصيد الاحتيالي).
Location Region	المنطقة الجغرافية المحاكاة للمعاملة.
PrefixIP	بادئة عنوان IP المحاكاة للمعاملة.
Login Frequency	تردد جلسات تسجيل الدخول للمستخدم، ويختلف حسب الفئة العمرية.

Session Duration	مدة جلسات النشاط بالدقائق.
Purchase Pattern	النمط السلوكي للمشتريات (على سبيل المثال، مركّزة، عشوائية، عالية القيمة).
Age Group	تصنيف المستخدمين إلى جدد، وراسخين، وقدامى بناءً على تاريخ نشاطهم.
Risk Score	درجة المخاطرة المحسوبة بناءً على خصائص المعاملة وسلوك المستخدم.
Anomaly	تقييم مستوى المخاطرة (على سبيل المثال، مخاطرة عالية، مخاطرة معتدلة، مخاطرة منخفضة).

من ثم تم التحقق من عدم وجود القيم الفارغة والمكررة في مجموعة البيانات.

11.1.1 تحليل البيانات

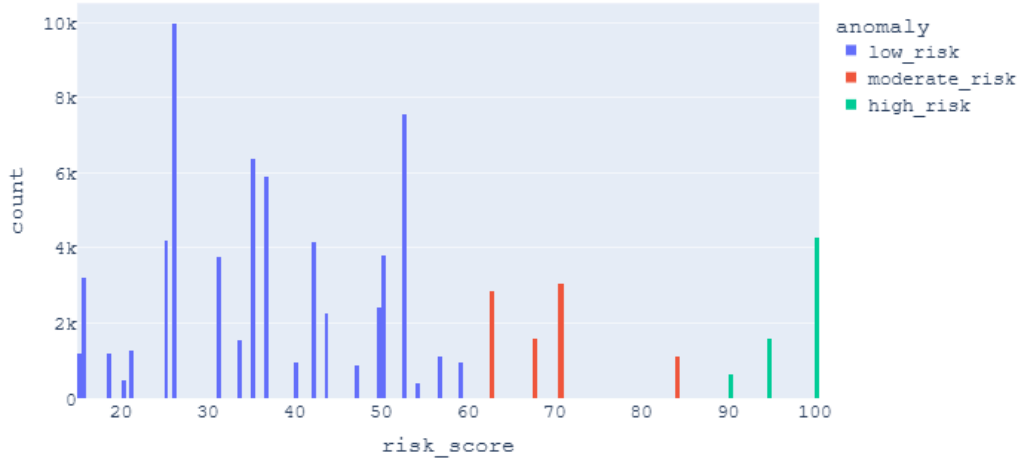
يوضح الشكل 3 أنواع خطورة المعاملات وفق الشذوذ، التصيد والاحتيال من أنواع المعاملات عالية المخاطر، حيث أن حوالي 50% من معاملات البيع تكون ذات مخاطرة متوسطة



شكل 3 انواع خطورة المعاملات وفق الشذوذ

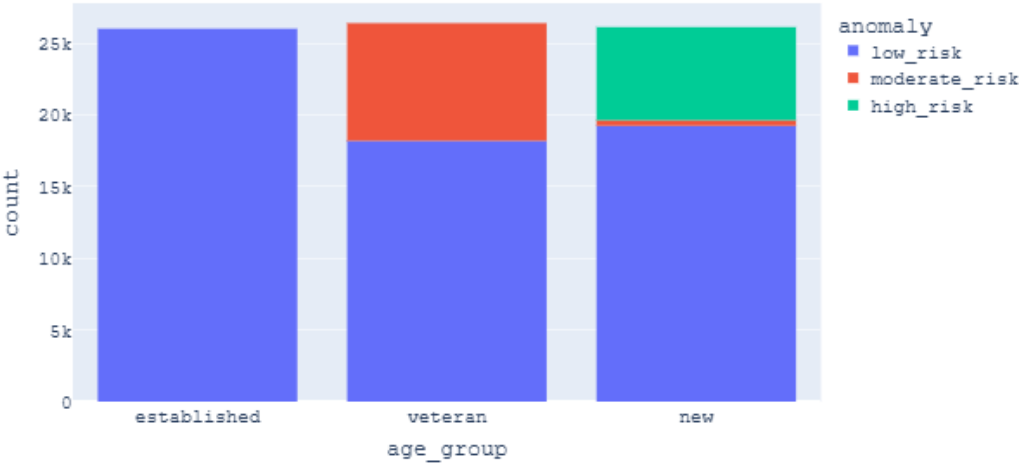
يوضح الشكل 4 درجة خطورة المعاملات (وهي قيمة بين 0 و 100 تشير القيم الدنيا الى ان الخطورة قليلة والقيم العالية على ان الخطورة اكبر وهي قيمة محسوبة بناءً على خصائص المعاملة وسلوك المستخدم وفق ما ورد بوصف مجموعة البيانات بالجدول 3)، أقل من 60 هي معاملات

ذات خطورة منخفضة، وبين 60 و 85 هي معاملات ذات خطورة متوسطة، وأعلى من 85 هي معاملات ذات خطورة عالية.



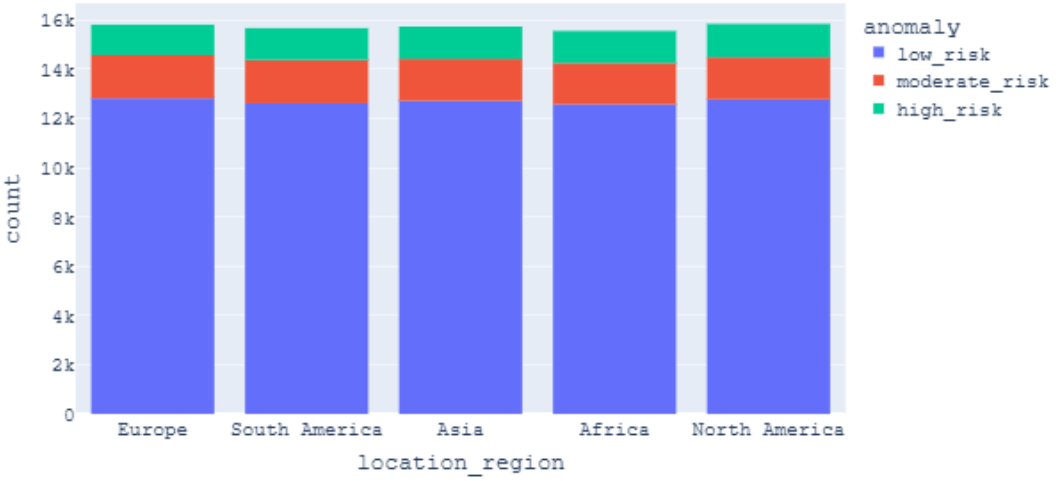
شكل 4 درجة خطورة المعاملات

يوضح الشكل 5 خطورة المعاملة وفق الفئة العمرية، هناك مخاطر عالية للمستخدمين الجدد لإجراء معاملات احتيالية، ومخاطر معتدلة للمستخدمين المخضرمين في حين أن المخاطر منخفضة للمستخدمين الحاليين.



شكل 5 خطورة المعاملة وفق الفئة العمرية

يوضح الشكل 6 خطورة المعاملة وفق المنطقة، يوجد عدد متساوٍ تقريبًا من نقاط البيانات من كل منطقة موقع بها نفس القدر من المخاطر.



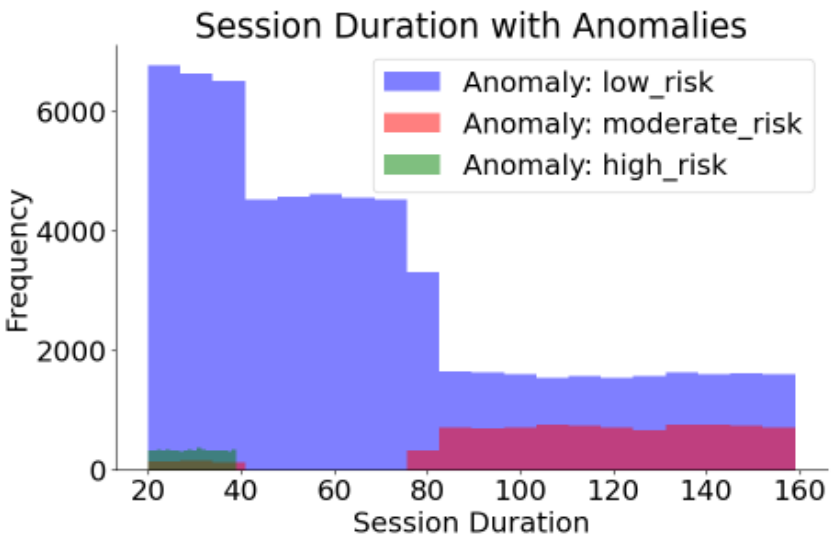
شكل 6 خطورة المعاملة وفق المنطقة

يوضح الشكل 7 خطورة المعاملة وفق مدة الجلسة:

• المعاملات عالية المخاطر لها مدة جلسة قصيرة تتراوح بين 20 إلى 40 دقيقة.

• المعاملات متوسطة المخاطر لها مدة جلسة إما قصيرة (20–40 دقيقة) أو طويلة جدًا (75–160 دقيقة).

• المعاملات منخفضة المخاطر لها علاقة عكسية بمدة الجلسة، فكلما زادت المدة انخفض عدد المعاملات منخفضة المخاطر.



شكل 7 خطورة المعاملة وفق مدة الجلسة

ليتم بعدها إنشاء مجموعة ميزات جديدة لإعطاء وزن للقيم التي تؤثر في التنبؤ بنتيجة المخاطر (risk score) واضافتها الى مجموعة البيانات:

1. تم إنشاء ميزة ثنائية لمعاملات البريد العشوائي والتصيد الاحتيالي (spam, phishing) من الميزة transaction_type وهي:
   - is_phishing_transaction لتحديد إذا ماكانت المعاملة تصيد أم لا.
   - Is_spam_transaction لتحديد إذا ماكانت المعاملة spam أم لا.
2. إنشاء ميزة ثنائية لمعاملات age_group لتحديد المعاملات المرتبطة بالفئات العمرية وهي:
   - is_new_transaction لتحديد إذا ماكانت المعاملة من قبل مستخدمين جدد أم لا.
   - is_veteran_transaction لتحديد إذا ماكانت المعاملة من قبل مستخدمين قدامى أم لا.
3. حساب القيمة العشرية لعنوان IPv4 من الميزة ip_prefix:

- تم تحويل كل ثمانية بتات إلى قيمتها العشرية ثم إضافة كل قيمة عشرية لتحقيق المكافئ العشري لعنوان IP.

يظهر الشكل 8 إضافة مجموعة الميزات الجديدة الى مجموعة البيانات.

age_group	risk_score	anomaly	is_spam_transaction	is_phishing_transaction	is_new_transaction	is_veteran_transaction	ip_prefix_numeric
established	18.75	low_risk	0	0	0	0	3221225472
established	25.00	low_risk	0	0	0	0	2885681152
established	31.25	low_risk	0	0	0	0	3232235520
veteran	36.75	low_risk	0	0	0	1	2885681152
veteran	62.50	moderate_risk	0	0	0	1	2886729728

شكل 8 إضافة مجموعة الميزات الجديدة الى مجموعة البيانات

ثم تطبيق عملية ترميز البيانات على الميزات الفئوية وهي:

- transaction_type
- ip_prefix_numeric
- sending_address
- receiving_address
- age_group

يوضح الشكل 9 مجموعة البيانات بعد عملية الترميز:

age_group	risk_score	anomaly	is_spam_transaction	is_phishing_transaction	is_new_transaction	is_veteran_transaction	ip_prefix_numeric
0	18.75	low_risk	0	0	0	0	3
0	25.00	low_risk	0	0	0	0	1
0	31.25	low_risk	0	0	0	0	4
2	36.75	low_risk	0	0	0	1	1
2	62.50	moderate_risk	0	0	0	1	2

شكل 9 مجموعة البيانات بعد عملية الترميز

وبعدها تمت عملية حذف الميزات التالية:

- age_group: كونه تم الحصول على الميزات الجديدة الخاصة بالفئات العمرية.
- location_region: يوجد عدد متساوٍ تقريباً من نقاط البيانات من كل منطقة موقع بها نفس القدر من المخاطر، وبالتالي يتم إسقاطها لأنها لن تؤثر على التنبؤ (موضح بالشكل 9).

وبعدها عملية تطبيع البيانات للميزات الرقمية باستخدام StandardScaler وهي الميزات التالية:

- Amount
- session_duration

لتصبح بعدها البيانات جاهزة لإجراء التجارب.**11.2 العتبة الموزونة (Threshold Balancing Mechanism – TBM)**

شرح الطريقة (TBM)

تقوم TBMعلى الجمع بين تقنيات موازنة البيانات وتحديد العتبة المثلى للمخاطر في تصنيف المعاملات في سلاسل الكتل على النحو التالي:

- **موازنة البيانات**
  - في البداية، يتم تطبيق تقنيات موازنة البيانات مثل RandomOverSampler، بهدف التعامل مع مشكلة عدم التوازن في البيانات، حيث تميل المعاملات ذات الخطورة العالية (high risk) إلى أن تكون أقل بكثير مقارنة بالمعاملات ذات الخطورة المتوسطة أو المنخفضة.
  - هذه الموازنة تعمل على زيادة تكرار المعاملات ذات الخطورة العالية بحيث تكون متوازنة مع المعاملات ذات الخطورة المتوسطة والمنخفضة، مما يتيح

لنموذج التعلم الآلي فهماً أفضل للأنماط الشاذة المرتبطة بالمعاملات عالية الخطورة.

- **تحديد العتبة المثلى (Threshold)**
  - بمجرد موازنة البيانات، يتم تحديد عتبة الشذوذ بناءً على درجة الخطورة .يتم اعتبار المعاملات التي تقع فوق العتبة المحددة (مثلاً المعاملات التي تتجاوز درجة خطورة 85) كمعاملات شاذة.
  - يتم تحديد هذه العتبة بناءً على أداء المصنفات المختلفة وتجربة أفضل نتائج النماذج عبر التجارب، حيث تكون العتبة التي تُحسن دقة التصنيف و الـ F1 Scoreهي الأنسب.
- **مزايا الآلية**
  - تحسين دقة المصنفات: بتطبيق موازنة البيانات، يمكن تقليل التأثير السلبي للبيانات غير المتوازنة، مما يحسن قدرة النموذج على التعرف على الشذوذ بشكل دقيق.
  - اكتشاف شذوذ أفضل: تحديد العتبة المثلى يساعد على تحسين الكشف عن الشذوذ في المعاملات ذات الخطورة العالية، مما يزيد من دقة تصنيف المعاملات الشاذة.
  - توفير وقت وموارد: من خلال تحسين الأداء باستخدام RandomOverSampler، يمكن تقليل الحاجة إلى معالجات إضافية ومعقدة، مما يحسن الكفاءة العامة للنظام.
- **تطبيق الآلية**
  - المرحلة الأولى: تطبيق RandomOverSampler على البيانات لتوازن توزيع المعاملات.

- المرحلة الثانية: تحديد العتبة المثلى بناءً على التجارب والمقارنات بين دقة المصنفات في حالات مختلفة.
- المرحلة الثالثة: استخدام هذه العتبة لتصنيف المعاملات كـ شاذة أو طبيعية بناءً على درجة الخطورة.

11.3 التجارب

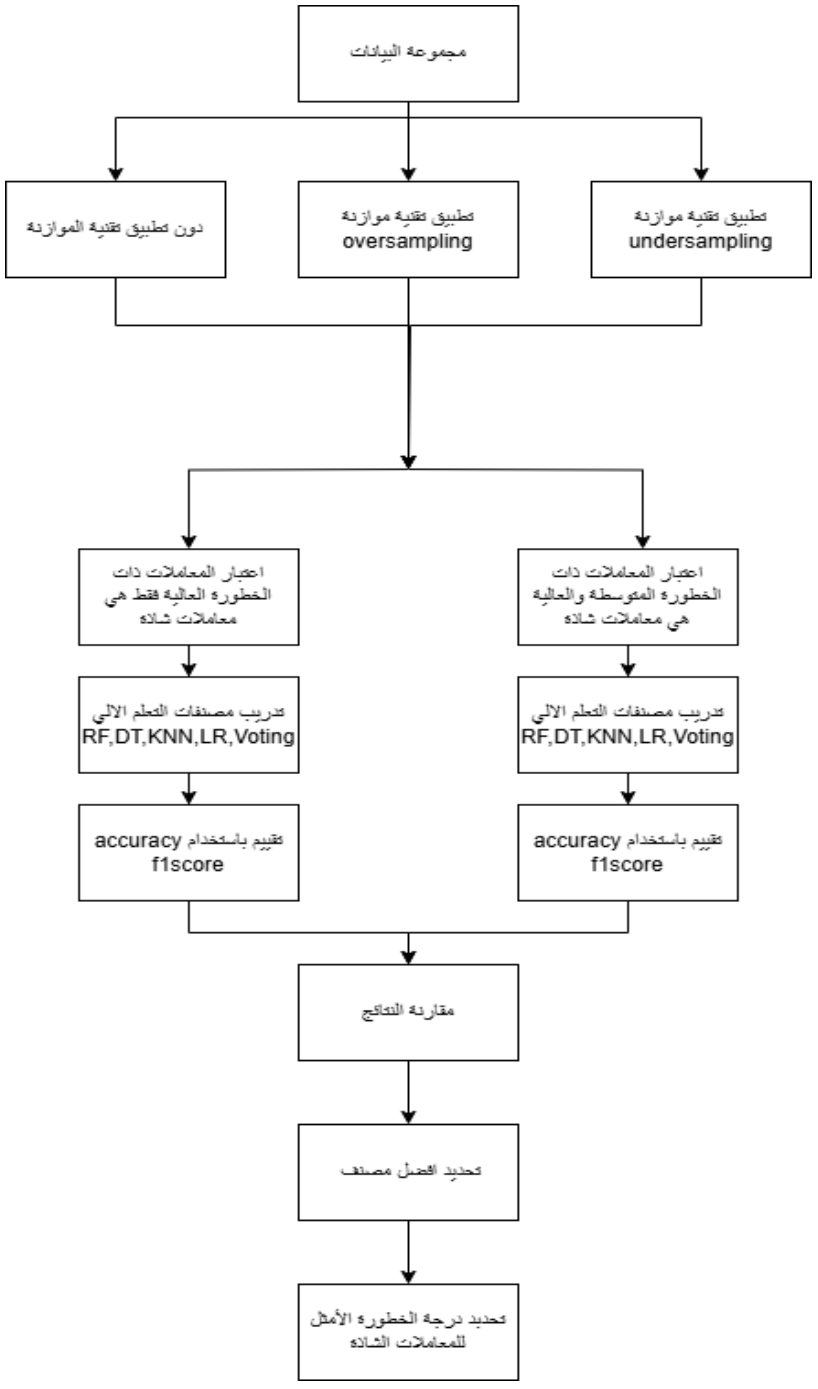
تم إجراء ثلاث تجارب: في التجربة الأولى لم يتم تطبيق أي تقنية موازنة وفي التجربة الثانية تم تطبيق تقنية موازنة البيانات RandomUnderSampler وفي التجربة الثالثة تم تطبيق تقنية موازنة البيانات RandomOverSampler

وذلك من اجل تحديد عتبة الشذوذ الأمثل الشذوذ في سلاسل الكتل.

وفي كل تجربة:

- يتم اعتبار المعاملات التي درجة خطورتها عالية كمعاملات شاذة وتدريب نماذج التعلم الآلي وتقييم اداءها.
- يتم اعتبار المعاملات التي درجة خطورتها متوسطة وعالية كمعاملات شاذة وتدريب نماذج التعلم الآلي وتقييم اداءها.

وبعدها يتم مقارنة أداء كافة المصنفات وتحديد درجة الخطورة الأمثل بناء على نتائج المصنفات، يوضح الشكل 10 تدفق العمل في البحث.



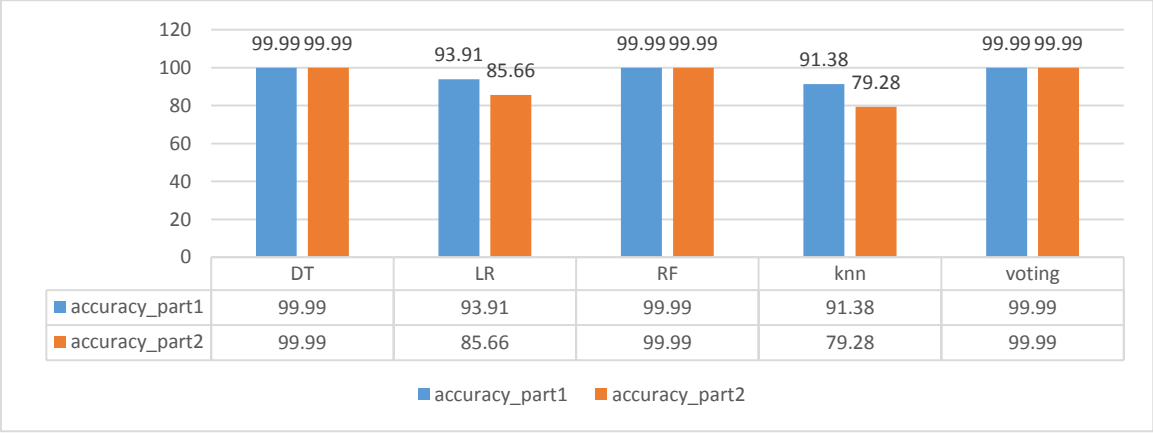
شكل 10 مخطط العمل في البحث

11.3.1 التجربة الأولى

تم استخدام مجموعة البيانات دون تقنيات موازنة

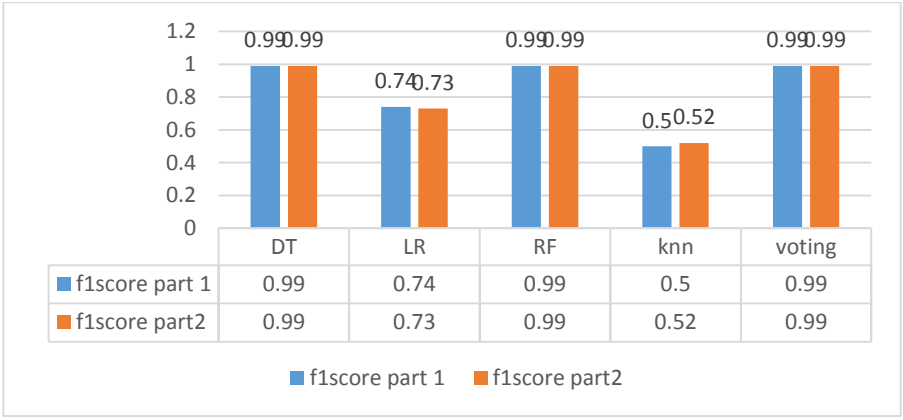
- **على اعتبار ان المعاملات الشاذة هي فقط high risk (في المخطط تمت تسميتها part 1).**
- **على اعتبار ان المعاملات الشاذة هي high risk, mid risk (في المخطط تمت تسميتها part 2).**

يوضح الشكل 11 تجميع دقة المصنفات في التجربة الأولى، **نتائج المصنفات في part 1 هي الأفضل لكن لايمكن الاعتماد على هذا المعيار لوحدة لعدم توازن البيانات**



شكل 11 تجميع دقة المصنفات في التجربة الاولى

يوضح الشكل 12 تجميع f1score للمصنفات في التجربة الأولى **نتائج المصنفات متقاربة في part 1 و part 2.**



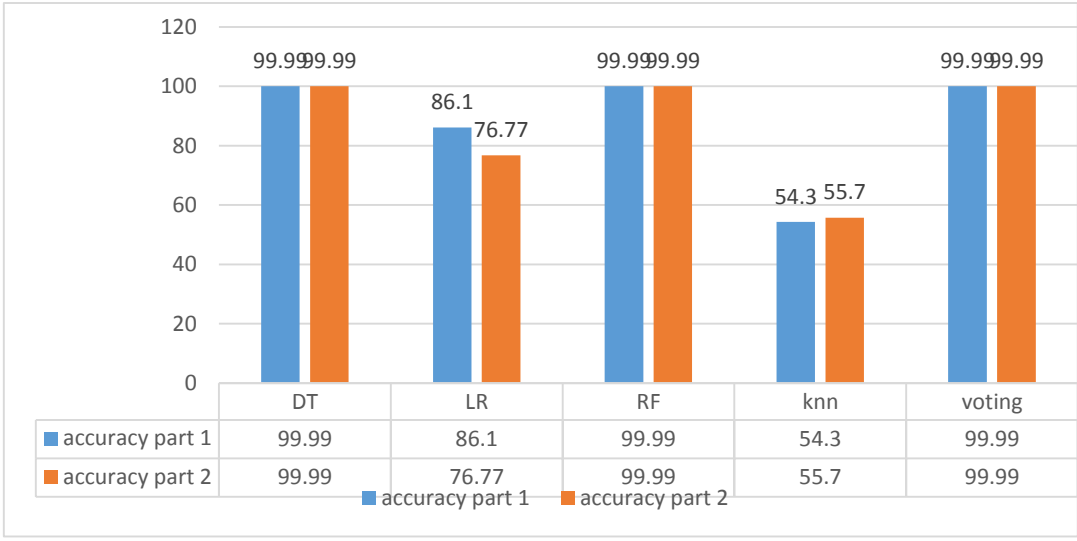
شكل 12 تجميع f1score المصنفات في التجربة الأولى

11.3.2 التجربة الثانية

تم استخدام تقنية موازنة البيانات RandomUnderSampler

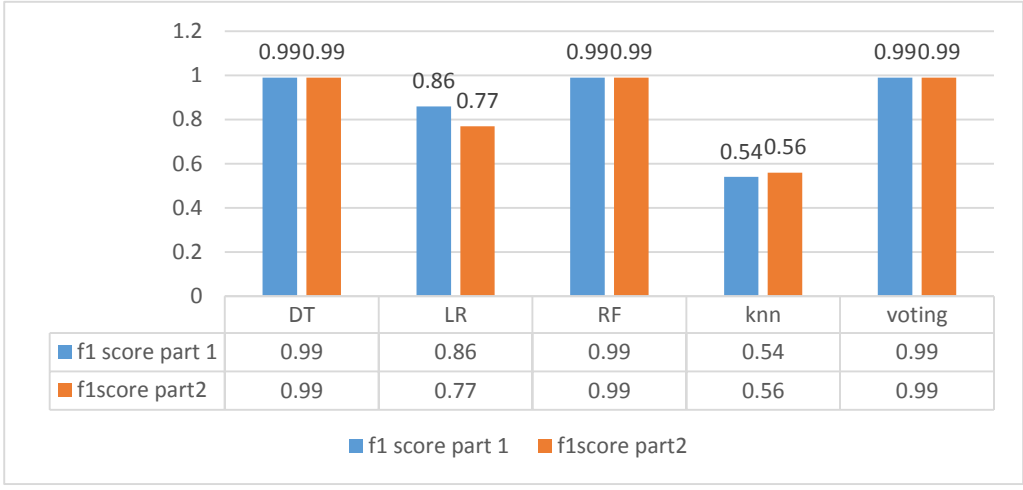
- **على اعتبار ان المعاملات الشاذة هي فقط high risk (في المخطط تمت تسميتها part 1)**
- **على اعتبار ان المعاملات الشاذة هي high risk, mid risk (في المخطط تمت تسميتها part 2)**

يوضح الشكل 13 تجميع دقة مصنفات التجربة الثانية **نتائج المصنفات في part 1 هي الأفضل وتم تحسين الأداء عن التجربة الأولى بالنسبة لخوارزميتي LR , KNN لكن لايمكن الاعتماد على هذا المعيار لوحدة لعدم توازن البيانات.**



شكل 13 تجميع دقة مصنفات التجربة الثانية

يوضح الشكل 14 تجميع f1score للمصنفات في التجربة الثانية، **نتائج المصنفات في part 1 هي الأفضل وتم تحسين الأداء عن التجربة الأولى بالنسبة لخوارزميتي LR , KNN.**



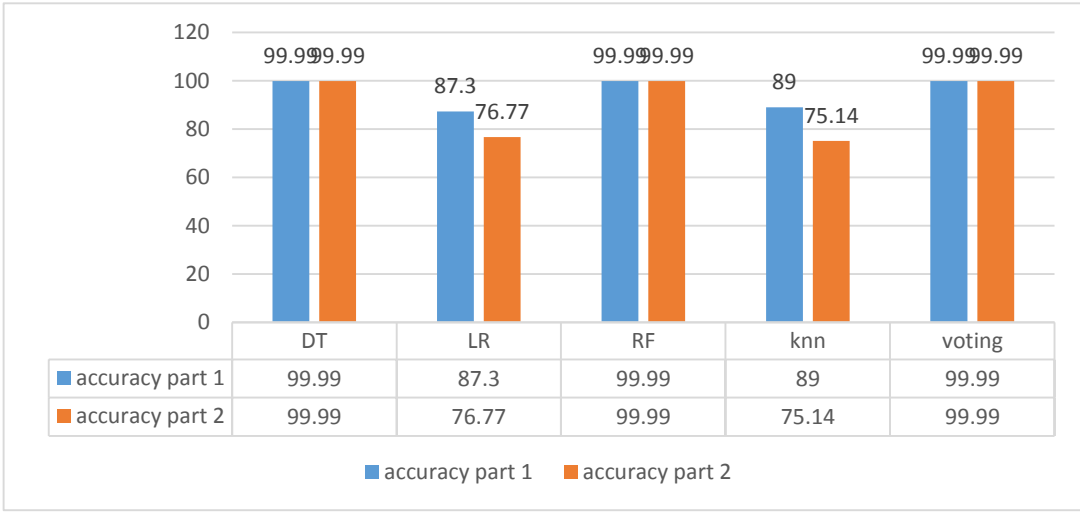
شكل 14 تجميع f1score المصنفات في التجربة الثانية

11.3.3 التجربة الثالثة

تم استخدام تقنية موازنة البيانات RandomOverSampler

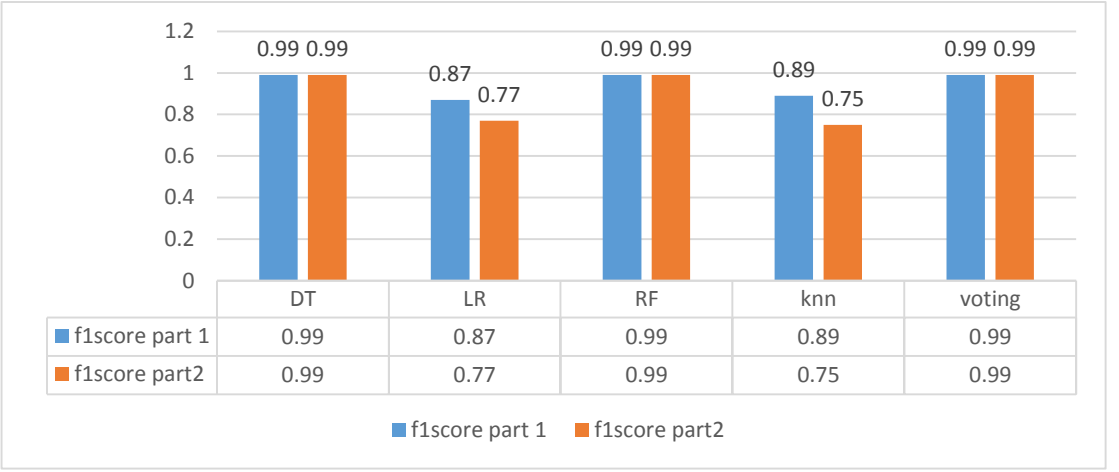
- **على اعتبار ان المعاملات الشاذة هي فقط high risk (في المخطط تمت تسميتها part 1)**
- **على اعتبار ان المعاملات الشاذة هي high risk, mid risk (في المخطط تمت تسميتها part 2)**

يوضح الشكل 15 تجميع دقة مصنفات التجربة الثالثة، **نتائج المصنفات في part 1 هي الأفضل وتم تحسين الأداء عن التجارب السابقة بالنسبة لخوارزميتي KNN , LR**



شكل 15 تجميع دقة مصنفات التجربة الثالثة

يوضح الشكل 16 تجميع f1score لمصنفات التجربة الثالثة، **نتائج المصنفات في part 1 هي الأفضل وتم تحسين الأداء عن التجارب السابقة بالنسبة لخوارزميتي KNN , LR**



شكل 16 تجميع f1score المصنفات في التجربة الثالثة

وبناء على ما سبق تم استنتاج ما يلي:

1. في كافة التجارب السابقة كان أداء كافة المصنفات أفضل على اعتبار ان المعاملات الشاذة هي فقط high risk .
2. في كافة التجارب السابقة كانت مصنفات DT,RF,Voting ذات دقة عالية وصلت لـ 99%.
3. في تجربة موازنة البيانات باستخدام تقنية RandomOverSampler حققت الخوارزميات اعلى نتائج مقارنة ببقية التجارب:

- تم تحسين أداء خوارزمية LR بالنسبة لمعيار f1score بمقدار 0.13 عن التجربة الاولى.
- تم تحسين أداء خوارزمية KNN بالنسبة لمعيار f1score بمقدار 0.39 عن التجربة الأولى.
- تم تحسين أداء خوارزمية LR بالنسبة لمعيار f1score بمقدار 0.01 عن التجربة الثانية.

- تم تحسين أداء خوارزمية KNN بالنسبة لمعيار f1score بمقدار 0.35 عن التجربة الثانية.

4. المعاملات التي درجة خطورتها أعلى من 85 هي المعاملات التي سيتم اعتبارها شاذة وكل ما دون درجة الخطوة هذه يعتبر معاملة طبيعية

12. الخلاصة والعمل المستقبلي

تم في هذا البحث تحديد عتبة للمعاملات الشاذة في سلاسل الكتل باستخدام التعلم الآلي عن طريق تدريب خمس مصنفات وهي شجرة القرار والغابة العشوائية والانحدار اللوجستي وأقرب جار ومصنف الانتخاب بين أفضل ثلاثة خوارزميات على مجموعة البيانات قبل تطبيق تقنيات موازنة عليها وبعد تطبيق تقنيات موازنة البيانات عليها، ومقارنة أداء الخوارزميات باستخدام معيار الدقة accuracy و f1score ، وقد اثبتت النتائج ان أفضل عتبة خطورة للمعاملات الشاذة هي فوق 85 high) risk) وذلك في كافة التجارب المطبقة وخصوصاً في موازنة البيانات باستخدام تقنية RandomOverSampler حيث وصلت الخوارزميات الى اعلى النتائج مقارنة ببقية التجارب فحققت مصنفات شجرة القرار والغابة العشوائية والانتخاب على دقة 99.99 و f1score مساوياً 0.99 وتم تحسين أداء مصنفي الانحدار اللوجستي واقرب جار بالنسبة لمقياس f1score بمقدار يقارب 13 % و 39 % على التوالي.

بالنسبة للعمل المستقبلي من الممكن تطبيق تقنيات التعلم العميق او ادخال عملية هندسة الميزات المتطورة وايضاً دراسة استهلاك الموارد مثل الذاكرة ونسبة استهلاك المعالج والتخزين من قبل هذه الخوارزميات لتحقيق أفضل موازنة بين الأداء واستهلاك الموارد.

13. جدول الاختصارات

AI	**Artificial Intelligence**
DT	**Decision Tree**
KNN	**K-Nearest Neighbor**
LR	**Logistic Regression**
MI	**Machine learning**
NN	**Neural network**
RF	**Random forest**
SGD	**Stochastic Gradient Descent**

المراجع

[1] Gadekallu, T. R., Huynh-The, T., Wang, W., Yenduri, G., Ranaweera, P., Pham, Q. V., ... & Liyanage, M. (2022). Blockchain for the metaverse: A review. arXiv preprint arXiv:2203.09738.

[2] Nechiti, Alexandru-Tudor. Anomaly Detection in Blockchain Networks. BS thesis. University of Twente, 2023

[3] Pendino, Stephanie R. BLOCKCHAIN NETWORK BEHAVIOR-BASED ANOMALY DETECTION. Diss. Monterey, CA; Naval Postgraduate School, 2019.

[4] https://www.kaggle.com/datasets/faizaniftikharjanjua/metaverse-financial-transactions-dataset

[5] L. Rokach, O. Maimon, Decision trees, in: O. Maimon, L. Rokach (Eds.), "Data Mining and Knowledge Discovery Handbook" , Springer, Boston, MA, pp. 165–192, 2005.

[6]Siddamsetti, S., Tejaswi, C., & Maddula, P. (2024). Anomaly detection in blockchain using machine learning. Journal of Electrical Systems, 20(3), 619-634.

[7] Anthony, Njoku ThankGod, et al. "Anomaly detection system for Ethereum blockchain using machine learning." Proceedings of the 19th International Conference on Manufacturing Research. Vol. 25. IOS Press, 2022.

[8]Introduction to Algorithms for Data Mining and Machine Learning Xin-She Yang Middlesex University School of Science and Technology London, United Kingdom.

[9] John Wiley & Sons, "Machine Learning For Dummies" Published by: New Jersey Media and software compilation 2016

[10] https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

[11] https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

[12] Ehsan et al., "Enhanced Anomaly Detection in Ethereum: Unveiling and Classifying Threats With Machine Learning," in IEEE Access, vol. 12, pp. 176440-176456, 2024, doi: 10.1109/ACCESS.2024.3504300

[13] Mehra, Sidharth & Hasanuzzaman, Mohammed, "Detection of Offensive Language in Social Media Posts", 2020.

[14] Mohammad Hasan, Mohammad Shahriar Rahman, Helge Janicke, Iqbal H. Sarker, Detecting anomalies in blockchain transactions using machine learning classifiers and

explainability analysis, Blockchain: Research and Applications, Volume 5, Issue 3, 2024, 100207

[15] Shin Morishima, Scalable anomaly detection in blockchain using graphics processing unit, Computers & Electrical Engineering, Volume 92, 2021.

[16] Demertzis, K., Iliadis, L., Tziritas, N., & Kikiras, P. (2020). Anomaly detection via blockchained deep learning smart contracts in industry 4.0. Neural Computing and Applications. doi:10.1007/s00521-020-05189-8

[17] Liang, Wei, et al. "Data fusion approach for collaborative anomaly intrusion detection in blockchain-based systems." IEEE Internet of Things Journal 9.16 (2021): 14741-14751

[18] https://www.ibm.com/think/topics/logistic-regression

[19] Ashfaq T, Khalid R, Yahaya AS, Aslam S, Azar AT, Alsafari S, Hameed IA. A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism. Sensors (Basel). 2022 Sep 21;22(19):7162. doi: 10.3390/s22197162. PMID: 36236255; PMCID: PMC9572131.

[20] Ali Yassin, Dr. Kamal Al-Salloum, Dr. Wassim Ramadan Selecting the Optimal Combination of Hyperparameter Tuning and Feature Selection to Improve the Performance of Anomaly Detection Systems Eng. Mechanical, Homs University Journal. Volume 44, Issue 5, 2022

[21] Eng. Ali Yassin, Dr. Kamal Al-Salloum, Dr. Wassim Ramadan, Selecting the Optimal Classification Threshold Dynamically in Early Anomaly Detection Systems Based on Deep Learning, Homs University Journal, Volume 44, Issue 8, 2022.