تقييم استهلاك الموارد لخوارزميات التعلم الآلي في كشف الشذوذ في شبكات 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

يعد اكتشاف الشذوذ في سلاسل الكتل إجراءً أمنياً بالغ الأهمية، يركز على تحديد الأنماط غير العادية أو غير المتوقعة في شبكات سلاسل الكتل والمعاملات. من خلال الكشف بشكل استباقي عن هذه الانحرافات، يهدف إلى حماية أنظمة سلاسل الكتل من الهجمات الضارة والاحتيال ومشاكل الأداء، وضمان سلامة وموثوقية التكنولوجيا.

يستعرض هذا البحث تحليلاً شاملاً لاستهلاك الموارد اللازمة لتشغيل خمسة من خوارزميات التعلم الآلي: شجرة القرار، الغابة العشوائية، أقرب جار، الانحدار اللوجستي، ومصنف الانتخاب بين أفضل ثلاثة خوارزميات. تم تقييم أداء هذه الخوارزميات من حيث استهلاك الذاكرة، نسبة استخدام المعالج، ومتطلبات التخزين، إضافةً إلى دقتها في كشف الشذوذ باستخدام معايير مثل الدقة و .F1-score كما تم تصنيف استهلاك الموارد لهذه الخوارزميات لثلاثة تصنيفات بحيث يتم تصنيف الخوارزمية

الى تصنيف مرتفع لاستهلاك الموارد او تصنيف متوسط لاستهلاك الموارد أو تصنيف منخفض لاستهلاك الموارد.

أظهرت النتائج أن الغابة العشوائية كانت الأفضل بين الخوارزميات ذات المتطلبات العالية، محققة دقة 99% و F1-score قدره 0.99. بينما تفوقت شجرة القرار بين الخوارزميات ذات المتطلبات المتوسطة بدقة 94.6% و F1-score بلغ 0.95. أما في فئة الخوارزميات ذات الموارد المنخفضة، فقد كانت خوارزمية الانحدار اللوجستي الأكثر كفاءة بدقة 50.22 و f1score 0.5.

الكلمات المفتاحية: سلاسل الكتل، كشف الشذوذ، التعلم الالي، تحليل الموارد، التصنيف.

Evaluation of Resource Consumption of Machine Learning Algorithms in Anomaly Detection in Blockchain Networks

Ali Atia [(1)], Prof. Maher Abbas [(2)], Dr. Wassim Ramadan [(3)]

(1) PhD student in the Department of Networks and Computer Systems Engineering, Faculty of Information Engineering, University of Homs.

(2) Professor in the Department of Networks and Computer Systems Engineering, Faculty of Information Engineering, University of Homs.

(3) PhD in the Faculty of Agricultural Engineering, University of Homs.

Abstract

Blockchain anomaly detection is a critical security measure that focuses on identifying unusual or unexpected patterns in blockchain networks and

transactions. By proactively detecting these anomalies, it aims to protect blockchain systems from malicious attacks, fraud, and performance issues, and ensure the safety and reliability of the technology.

This paper presents a comprehensive analysis of the resource consumption required to run five machine learning algorithms: decision tree, random forest, nearest neighbor, logistic regression, and selection classifier among the top three algorithms. The performance of these algorithms was evaluated in terms of memory consumption, processor utilization, and storage requirements, in addition to their accuracy in anomaly detection using criteria such as accuracy and F1-score.

The resource consumption of these algorithms was also classified into three categories, so that the algorithm is classified into a high resource consumption category, a medium resource consumption category, or a low resource consumption category. The results showed that random forest was the best among the algorithms with high requirements, achieving an accuracy of 99% and an F1-score of 0.99. While decision tree outperformed among the algorithms with medium requirements with an accuracy of 94.6% and an F1-score of 0.95. As for the category of algorithms with low resources, the logistic regression algorithm was the most efficient with an accuracy of 50.22 and an F1-score of 0.5.

Keywords: Blockchain, Anomaly Detection, Machine Learning, Resource Analysis, Classification.

1. مقدمة

يعد دمج اكتشاف الشذوذ في أنظمة سلاسل الكتل وسيلة واعدة لتعزيز الأمان والثقة. من خلال تحديد الأنماط والسلوكيات غير العادية، يمكن لهذه الخوارزميات المساعدة في اكتشاف الأنشطة الضارة ومنع الاحتيال والحفاظ على سلامة شبكة سلاسل الكتل .ومع ذلك، تأتي هذه الفوائد مع تحذير بالغ الأهمية في استهلاك الموارد. تتوزع بيئات سلاسل الكتل بطبيعتها وغالباً ما تكون مقيدة بالموارد. على عكس الأنظمة المركزية حيث قد تكون الطاقة الحسابية الوفيرة متاحة بسهولة، تعتمد سلاسل الكتل غالباً على عقد لامركزية بمستويات متفاوتة من قدرات المعالجة مما يفرض ضرورة لمعرفة متطلبات الموارد اللازمة لكل خوارزمية لمعرفة قدرة العقد في سلاسل الكتل على استخدامها، على الرغم من تعدد الدراسات التي تناولت كشف الشذوذ في سلاسل الكتل باستخدام خوارزميات التعلم الآلي، إلا أن معظمها ركز على تحسين دقة الكشف دون تحليل تأثير هذه الخوارزميات على استهلاك الموارد. يهدف هذا البحث إلى سد هذه الفجوة من خلال تقديم تحليل شامل لمتطلبات الموارد، مما يساعد في اختيار الخوارزمية الأكثر كفاءة وفقاً لقدرات العقد داخل شبكات سلاسل الكتل.

2. مشكلة البحث

مع اكتساب تقنية سلاسل الكتل اعتماداً واسع، أصبحت الحاجة إلى تدابير أمنية قوية أمراً بالغ الأهمية. يعالج اكتشاف الشذوذ باستخدام التعلم الالي في سلاسل الكتل هذه الحاجة من خلال تحديد الانحرافات عن السلوك الطبيعي الذي قد يشير إلى هجمات ضارة أو ثغرات أو أنشطة احتيالية.

تعتمد العديد من الدراسات على خوارزميات التعلم الآلي لكشف الشذوذ، لكنها غالباً ما تهمل تحليل متطلبات الموارد اللازمة لتشغيل هذه الخوارزميات في بيئات سلاسل الكتل، التي تتميز بكونها لامركزية ومحدودة الموارد. هذه الفجوة البحثية تثير تساؤلاً حول مدى قدرة العقد المختلفة في الشبكة على تشغيل خوارزميات كشف الشذوذ بكفاءة دون التأثير على الأداء العام.

3. هدف البحث

يهدف البحث بشكل رئيسي الى دراسة وتحليل متطلبات الموارد المطلوبة لاستخدام مصنفات التعلم الالي لكشف الشذوذ في سلاسل الكتل عن طريق تقسيم الخوارزميات الى ثلاثة أنواع: مصنفات ذو متطلبات عالية ومصنفات ذو متطلبات متوسطة ومصنفات ذو متطلبات ضعيفة مع مراعاة دقة كشف الشذوذ المترافقة مع استخدام الموارد.

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 [3]، ويمكن لجميع عقد الشبكة الوصول الى سلسلة الكتل وتتبع جميع المعاملات التي يتم اجراؤها [30].

1.4 نواع سلاسل الكتل

لقد تطورت تقنية سلاسل الكتل بشكل كبير في السنوات القليلة الماضية وبناءً على سماتها المختلفة، يمكن تقسيمها إلى أنواع متعددة [3].

- **سلاسل الكتل العامة**

سلاسل الكتل العامة مفتوحة للعامة ويمكن لأي فرد أن يشارك في عملية صنع القرار من خلال أن يصبح عقدة، ولكن قد يستفيد المستخدمون أو لا يستفيدون من مشاركتهم في عملية صنع القرار.

- **سلاسل الكتل الخاصة**

هذه الأنواع من سلاسل الكتل ليست مفتوحة للعامة ومفتوحة فقط لمجموعة من الأشخاص أو المنظمات ويتم مشاركة السجل مع الأعضاء المشاركين فقط.

- **سلاسل الكتل شبه الخاصة**

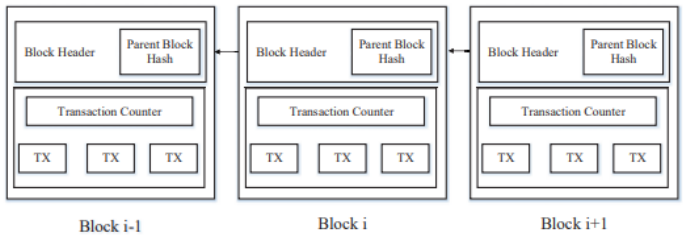
في سلسلة الكتل شبه الخاصة، يكون جزء من سلسلة الكتل خاصاً ويخضع لسيطرة مجموعة أو منظمات، والباقي مفتوح للجمهور ليتمكن أي شخص من المشاركة فيه.

- **دفتر الأستاذ المرخص**

في هذا النوع من سلسلة الكتل، يكون المشاركون معروفين وموثوقين بالفعل. في دفتر الأستاذ المرخص، يتم استخدام بروتوكول اتفاق للحفاظ على نسخة مشتركة من الحقيقة بدلاً من آلية الإجماع.

2.4 بنية سلاسل الكتل

نشأت فكرة سلسلة الكتل من ورقة بيضاء كتبها ناكاموتو ساتوشي في عام 2008 [17]. تمتلك سلسلة الكتل، والتي تسمى أيضاً دفتر الأستاذ الموزع، كتلاً متتالية، ترتبط ببعضها البعض من خلال قيمة التجزئة لرأس الكتلة السابقة. وبصرف النظر عن التجزئة، يتم أيضاً تضمين الطابع الزمني والرقم العشوائي وبيانات المعاملات في كتلة[1] كما هو موضح في الشكل 1.



شكل 1 بنية سلاسل الكتل المصدر [2]

تتكون سلاسل الكتل مما يلي [2]:

- **الكتلة:**

تتكون الكتلة من رأس الكتلة وجسم الكتلة. وعلى وجه الخصوص، يتضمن رأس الكتلة: (أ) إصدار الكتلة: يشير إلى مجموعة قواعد التحقق من صحة الكتلة التي يجب اتباعها. (ب) تجزئة جذر شجرة ميركل: قيمة تجزئة جميع المعاملات في الكتلة. (ج) الطابع الزمني. (د) nBits: عتبة الهدف لتجزئة الكتلة الصالحة. (هـ) Nonce: حقل مكون من 4 بايت، يبدأ عادةً بالرقم 0 ويزداد مع كل حساب تجزئة (و) تجزئة الكتلة الأصلية: قيمة تجزئة 256 بت تشير إلى الكتلة السابقة، يتكون جسم الكتلة من عداد المعاملات. يعتمد الحد الأقصى لعدد المعاملات التي يمكن أن تحتويها الكتلة على حجم الكتلة وحجم كل معاملة. تستخدم سلاسل الكتل آلية تشفير غير متماثلة للتحقق من صحة مصادقة المعاملات [13].

- **المناقلات:** تمثل المعاملة تفاعلًا بين الأطراف. في العملات المشفرة cryptocurrency، على سبيل المثال، تمثل المعاملة نقل العملة المشفرة بين مستخدمي شبكة سلاسل الكتل.

3.4 مزايا تقنية البلوك تشين

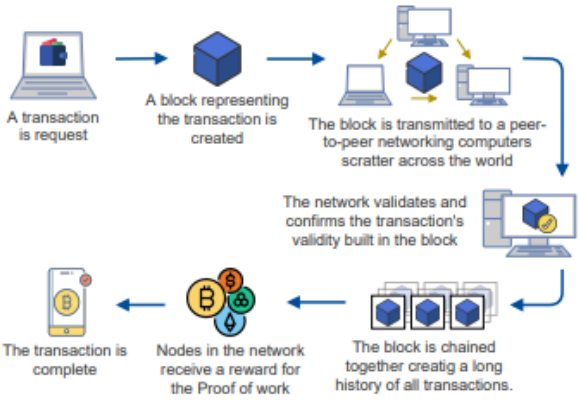
أبرز مزايا تقنية سلاسل الكتل وفق :[3]

- واحدة من أكبر مزايا تقنية البلوك تشين هي التوزيع الذي يسمح بمشاركة قاعدة البيانات دون وجود هيئة أو كيان مركزي. نظراً للطبيعة اللامركزية لتقنية البلوك تشين، يكاد يكون من المستحيل تعديل البيانات مقارنة بقاعدة البيانات التقليدية.
- يتم تمكين المستخدمين من التحكم في معلوماتهم ومعاملاتهم.
- توفر تقنية البلوك تشين الشفافية والثبات للمعاملات حيث لا يمكن تغيير أو حذف جميع المعاملات.

4.4 الية عمل سلاسل الكتل

تبدأ العملية بطلب المعاملة ومن ثم يتم إنشاء كتلة تمثل المعاملة ثم يتم نقل الكتلة إلى العقد nodes المتصلة بشبكة نظير إلى نظير وبعدا تتحقق الشبكة من صحة المعاملة المضمنة في الكتلة وتؤكدها

لتكتمل المعاملة ثم تتلقى العقد في الشبكة مكافأة لإثبات العمل يتم ربط الكتلة معًا لإنشاء تاريخ طويل لجميع المعاملات. يوضح الشكل 2 معاملة تتم معالجتها بواسطة تقنية سلاسل الكتل.



شكل 2 الية عمل سلاسل الكتل المصدر [1]

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].

تمر عملية التعلم الآلي في عدد من الخطوات وهي: استخراج البيانات – تحليل البيانات – إعداد البيانات – تدريب النموذج – تقييم النموذج [14].

15. أنواع التعلم الآلي

بناءً على أساليب وطريقة التعلم، ينقسم التعلم الآلي إلى أربعة أنواع أساسية، وهي[13]:

- **التعلم الآلي الخاضع للإشراف**

كما يوحي اسمه، يعتمد التعلم الآلي الخاضع للإشراف على الإشراف. هذا يعني أنه في أسلوب التعلم الخاضع للإشراف، يتم تدريب الآلات باستخدام مجموعة البيانات "المعنونة" (labels) ، ومن ثم بناءً على التدريب، تتنبأ الآلة بالخرج الذي يكون إما تنبؤ أو تصنيف (classification, regression)، تم في هذا البحث الاعتماد على التعلم الخاضع للإشراف وتحديداً التصنيف وهو ماسيتم التركيز عليه.

- **التعلم بدون إشراف**

يختلف التعلم غير الخاضع للإشراف عن أسلوب التعلم الخاضع للإشراف؛ كما يوحي اسمها، ليس هناك حاجة للإشراف. وهذا يعني في التعلم الآلي غير الخاضع للإشراف، أن الآلة يتم تدريبها باستخدام مجموعة البيانات غير المعنونة.

- **التعلم شبه الخاضع للإشراف**

التعلم شبه الخاضع للإشراف هو نوع من خوارزمية التعلم الآلي التي تقع بين التعلم الآلي الخاضع للإشراف والتعلم غير الخاضع للإشراف. فهو يمثل الأرضية الوسيطة بين خوارزميات التعلم الخاضع للإشراف (مع بيانات التدريب المعنونة) والتعلم غير الخاضع للإشراف (بدون بيانات تدريب معنونة) ويستخدم مزيجاً من مجموعات البيانات المصنفة وغير المصنفة خلال فترة التدريب.

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 [31].

1.6 تحديات كشف الشذوذ

على مستوى مجرد، يتم تعريف الشذوذ على أنه نمط لا يتوافق مع السلوك الطبيعي المتوقع. وبالتالي، فإن أسلوب اكتشاف الشذوذ المباشر هو تحديد منطقة تمثل السلوك الطبيعي وإعلان أي ملاحظة في البيانات لا تنتمي إلى هذه المنطقة الطبيعية على أنها شذوذ. لكن هناك عدة عوامل تجعل هذا النهج البسيط ظاهرياً صعباً للغاية [16]:

- من الصعب للغاية تحديد منطقة طبيعية تشمل كل سلوك طبيعي محتمل. بالإضافة إلى ذلك، غالباً ما تكون الحدود بين السلوك الطبيعي والشاذ غير دقيقة.
- عندما تكون الحالات الشاذة نتيجة لأفعال خبيثة، غالباً ما يتكيف الأعداء الخبثاء لجعل الملاحظات الشاذة تبدو طبيعية، مما يجعل مهمة تعريف السلوك الطبيعي أكثر صعوبة.
- عادة ما يكون توافر البيانات المصنفة للتدريب / التحقق من صحة النماذج المستخدمة بواسطة تقنيات اكتشاف العيوب مشكلة رئيسية.

نظراً للتحديات المذكورة أعلاه، فإن مشكلة اكتشاف الشذوذ، في شكلها الأكثر عمومية، ليس من السهل حلها.

2.6 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 [17]: 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.

1.2.6 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 [17] : 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. هجمات :Sybil تحدث 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 ادبيات الدراسة مجموعة من الخوارزميات لمعالجة الشذوذ في مجموعات البيانات ولكون مجموعات بيانات سلاسل الكتل لها وضع خاص فقد تنوعت هذه الاساليب فمنها التعلم الآلي والتعلم العميق. سيتم الاعتماد في هذا البحث على ذكر الدراسات السابقة على طريقة تصنيفها حسب نوع خوارزمية الكشف عن الشذوذ في سلاسل الكتل حيث يفيد هذا التصنيف في النقاط التالية:

1. التعرف على التقنيات المستخدمة حالياً لكشف الشذوذ في سلاسل الكتل.
2. التعرف على الطرائق الأفضل من خلال عمليات المقارنة بين طرائق كشف الشذوذ.
3. تحديد أسباب اختيار خوارزمية عن أخرى في اكتشاف الشذوذ الشبكي.

حيث ركزت الدراسات على استخدام طرائق التعلم الآلي بإشراف أكثر من التعلم الآلي دون اشراف نظراً لاعتبار مشكلة الشذوذ مشكلة تصنيف حيث تصنف البيانات في سلاسل الكتل على أنها إما بيانات طبيعية أو بيانات غير طبيعية. بالمقابل تناولت العديد من الدراسات استخدام التعلم العميق نظراً لفعالية التعلم العميق والقدرة على اكتشاف أنماط شذوذ جديدة وقلة تأثرها بمشكلة عدم توازن البيانات.

فكانت خوارزمية الغابة العشوائية وشجرة القرار وخوارزمية أقرب جار وLogistic regressionn هي الأكثر استخداماً مقارنة بغيرها.

1.7 طرائق التعلم بإشراف

تبحث [22] الدراسة في تطبيق نماذج التعلم الآلي للكشف عن الشذوذ وتحليل الاحتيال في معاملات سلاسل الكتل داخل Open Metaverse، وسط التعقيد المتزايد للمعاملات الرقمية. باستخدام مجموعة بيانات مكونة من 78600 معاملة تعكس مجموعة واسعة من سلوكيات المستخدم وأنواع المعاملات، تم تقييم فعالية العديد من النماذج التنبؤية، بما في ذلك RandomForest و LinearRegression و SVR و DecisionTree و KNeighbors و GradientBoosting و AdaBoost و Bagging و XGB و LightGBM، بناءً على متوسط خطأ التربيع المتبادل للتحقق MSE كشف التحليل أن أساليب المجموعة، وخاصة RandomForest و Bagging، أظهرت أداءً متفوقًا مع متوسط خطأ تربيعي للتحقق المتبادل −0.00445 و−0.00415 على التوالي، مما يسلط الضوء على قوتها في مجموعة بيانات المعاملات المعقدة. وعلى النقيض من ذلك، كان LinearRegression و SVR من بين الأقل فعالية، حيث بلغ متوسط أخطاء التباين المتوسط − 224.67 و−468.57، مما يشير إلى عدم التوافق المحتمل مع خصائص مجموعة البيانات. ويؤكد هذا البحث على أهمية اختيار استراتيجيات التعلم الآلي المناسبة في سياق معاملات سلاسل الكتل داخل Open Metaverse، مما يبرز الحاجة إلى مناهج متقدمة وقابلة للتكيف.

يقدم البحث [23] نهجاً هندسياً آلياً للميزات لمعاملات سلاسل الكتل .يتضمن ذلك عملية استخراج الميزات، حيث تعرف بأنها عملية إنشاء مجموعة ميزات باستخدام خصائص البيانات التي تعزز أداء خوارزميات تعلم الآلة [32]. حصلت هذه الميزات على دقة اختبار عالية بشكل ملحوظ لتصنيف معاملات البيتكوين الشاذة (88.0٪) ومعاملات Ethereum EOA (97.86٪ . تم تقييم الميزات المصممة باستخدام نماذج التعلم الآلي المختلفة LR, RF, k-NN, DT , NB , XG ، والتي أظهرت أداء تصنيف عالياً في التجارب حيث كانت RF الأفضل . تم تحليل أهمية الميزة باستخدام تقنية .ShAP عكست درجة أهمية الميزة أن الميزات المصممة هي في الواقع الأكثر أهمية لتصنيف معاملات سلاسل الكتل.

تم اقتراح نموذجاً في [24] للكشف عن الاحتيال يعتمد على التعلم الآلي في سلاسل الكتل. تم استخدام خوارزميتان للتعلم الآلي – XGboost والغابة العشوائية تستخدمان لتصنيف المعاملات. تدرب تقنيات التعلم الآلي مجموعة البيانات بناءً على أنماط المعاملات الاحتيالية والمتكاملة وتتنبأ بالمعاملات الواردة الجديدة حيث وصلت دقة النماذج الى 90%.

تستخدم الدراسة [25] تقنيات طريقة التصفية مثل المعلومات المتبادلة وتحليل التباين وإزالة الميزة المتكررة لتحديد المجموعة المثالية من الميزات تقوم الدراسة بفحص وتصنيف أفضل 10 مجموعات من الميزات باستخدام مصنف الغابة العشوائية لأهمية الميزةRF، بناءً على مجموعة البيانات التي تم إنتاجها بواسطة أفضل نهج للتصفية (مما يعطي دقة أعلى). بعد ذلك، يتم استخدام منهجية المجموعة لإنشاء النموذج النهائي، باستخدام مجموعة البيانات النهائية المكونة من 10 ميزات. الغرض من هذا النهج هو تعزيز مستوى اكتشاف الشذوذ في شبكة سلاسل الكتل .لتحديد فعالية النموذج المقترح، يتم إجراء التجارب ومقارنته بالمصنفات الفردية XGBوشجرة القرار والانحدار اللوجستي والغابة العشوائية وأقرب جار تكشف نتائج الدراسة أن نهج التصويت الجماعي يحقق معدل دقة 96.78٪، متجاوزاً دقة نماذج التصنيف الفردية التي تستخدم الميزات المثلى. بالإضافة إلى ذلك، تشير نتائج الدراسة إلى أن اختيار الميزات وكميتها يؤثران بشكل كبير على ناتج النموذج.

تم في الدراسة [26] استخدام تقنيات التعلم الآلي لتقديم نهج يمكنه اكتشاف الحسابات الاحتيالية على .Ethereum تم استخدام خوارزميات K-Nearest Neighbor وRandom Forest وXGBoost على مجموعة بيانات مجمعة من 4681 حالة إلى جانب 2179 حسابًا احتياليًا مرتبطًا و2502 حسابًا عاديًا. حققت تقنيات XGBoost وRF وKNN متوسط دقة 96.80٪ و94.8٪ و8٪ و87.85٪ ومتوسط 0.995 AUC و0.99 و0.93 على التوالي.

يركز الحل المقترح في [27]على استخدام نهج ديناميكي حيث يتم استخدام السلوك التشغيلي الطبيعي لسلاسل كتل Ethereum لتدريب خوارزميات ML وسيتم وضع علامة على أي انحراف على أنه

شذوذ وسيتم اكتشافه بواسطة النظام. تم استخدام أربع خوارزميات ML بما في ذلك K-Nearest Neighbors (KNN)و Gaussian Naive Bayes (GaussianNB) و Random Forestو 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.

2.7 طرائق التعلم دون اشراف

قام الباحثون في [19] باستخدام خوارزمية Isolation Forest للكشف عن الشذوذ داخل مجموعة بيانات معاملات سلاسل الكتل التي تضمنت ميزات مختلفة مثل مبالغ المعاملات ودرجات المخاطرة ومدة الجلسة. من إجمالي 78600 معاملة تم تحليلها، حددت الخوارزمية 3930 معاملة على أنها شذوذ، تمثل حوالي 5٪ من مجموعة البيانات. كشف التحليل أن أكثر أنواع المعاملات شيوعاً المرتبطة بالشذوذ كانت "البيع" و "الاحتيال"، مما يشير إلى ميل أعلى لهذه الفئات لإظهار سلوك غير منتظم وتشير الأنماط الجغرافية والسلوكية التي أبرزتها هذه الدراسة إلى أن التركيز على المخاطر وأنواع المعاملات الخاصة بالمنطقة يمكن أن يعزز الأمن العام لشبكات سلاسل الكتل . ومن خلال تحديد المناطق وأنواع المعاملات الأكثر عرضة للشذوذ، يمكن لأصحاب المصلحة تنفيذ استراتيجيات مراقبة وتدخل أكثر فعالية

الغرض الرئيسي من هذه الدراسة في [20] هو تقديم طريقة جديدة لاكتشاف الشذوذ في Bitcoin بكفاءة أكثر ملاءمة. في هذه الدراسة، تم استخدام نهج الشذوذ الجماعي. بدلاً من اكتشاف شذوذ العناوين والمحافظ الفردية، تم فحص شذوذ المستخدمين. بالإضافة إلى استخدام طريقة اكتشاف الشذوذ الجماعي، تم استخدام خوارزمية Trimmed_Kmeans للتجميع. تظهر نتائج هذه الدراسة أن الشذوذ أكثر وضوحًا بين المستخدمين الذين لديهم محافظ متعددة. كشفت الطريقة المقترحة عن 14 مستخدمًا ارتكبوا عمليات احتيال، بما في ذلك 26 عنوانًا في 9 حالات، في حين كشفت

الأعمال السابقة عن 7 عناوين كحد أقصى في 5 حالات احتيال. بالإضافة إلى تقليل تكلفة المعالجة لاستخراج الميزات، فإن النهج المقترح يكشف عن المزيد من المستخدمين غير الطبيعيين والسلوكيات الشاذة.

يناقش البحث [21] تحديد الشذوذ مع الإشارة بشكل خاص إلى شبكة معاملات البيتكوين .في هذه الحالة، يعد سلوك الشذوذ وكيلًا للنشاط المقلق، وبالتالي فإن الهدف هو العثور على الشذوذ في مجموعة البيانات من حيث نسبتها المئوية. لتحقيق ذلك، تم استخدام طريقة اختيار الميزة وهي اختيار الميزة الأمامية المتسلسلة جنباً إلى جنب مع ثلاث تقنيات ML، وتجميع kmeans، وغابة العزل، وآلة الدعم المتجه (SVM) وتم الحصول على أعلى دقة بنسبة 98.2٪ في SVM مقارنة بجميع الطرق الأخرى

بعدما تم عرض التوجهات المستخدمة لكشف الشذوذ في سلاسل الكتل باستخدام التعلم الآلي بنوعيه التعلم بإشراف وتعلم بدون إشراف. نلاحظ تفوق خوارزميات التعلم الآلي لاكتشاف الشذوذ في سلاسل الكتل وهي الغابة العشوائية وشجرة القرار وخوارزمية أقرب جار والانحدار اللوجستي على باقي خوارزميات التعلم الآلي وكثرة استخدامهم في هذا المجال ولذلك سيتم اختيارهم في هذا البحث كمايتم استخدام خوارزمية الانتخاب (الجميع بين أفضل ثلاثة مصنفات بينهم).

يوضح الجدول 1 مقارنة الدراسات المرجعية التي استخدمت التعلم بإشراف لكشف الشذوذ في سلاسل الكتل، حيث لم تتم دراسة متطلبات موارد خوارزميات التعلم الآلي لاكتشاف الشذوذ في سلاسل الكتل من قبل الأبحاث السابقة لذلك سيتم الاكتفاء بعرض معايير أداء الخوارزميات في كشف الشذوذ دون معايير قياس موارد أداء الخوارزميات

جدول 1 مقارنة الدراسات المرجعية

الدراسة	الخوارزميات المستخدمة	أبرز النتائج
[22]	RandomForest و LinearRegression و SVR و DecisionTree و KNeighbors و GradientBoosting و AdaBoost و Bagging و XGB و LightGBM	RandomForest و Bagging، أظهرت أداءً متفوقًا مع متوسط خطأ تربيعي للتحقق المتبادل – 0.00445 و−0.00415 على التوالي، مما يسلط الضوء على قوتها في مجموعة بيانات المعاملات المعقدة
[23]	LR, RF, k-NN, DT , NB , XG	RF الأفضل لتصنيف معاملات البيتكوين الشاذة (88.0٪) ومعاملات Ethereum EOA (97.86٪) .
[24]	XGboost والغابة العشوائية	وصلت دقة النماذج الى 90%.
[25]	XGB شجرة القرار والانحدار اللوجستي والغابة العشوائية وأقرب جار	تكشف نتائج الدراسة أن نهج التصويت الجماعي يحقق معدل دقة 96.78٪
[26]	K-Nearest Neighbor و Random Forest و XGBoost	حققت تقنيات XGBoost و RF و KNN متوسط دقة 96.80٪ و 94.8٪ و 8٪ و 87.85٪
[27]	K-Nearest Neighbors (KNN) و Gaussian Naive Bayes (GaussianNB)و Random Forest و Stochastic Gradient Descent (SDG)	أظهرت النتائج التجريبية أن خوارزمية الغابة العشوائية قدمت أفضل دقة بنسبة 99.84٪
الدراسة الحالية	شجرة القرار، الغابة العشوائية، أقرب جار، الانحدار اللوجستي، ومصنف الانتخاب	تقييم أداء الخوارزميات من حيث استهلاك الذاكرة، نسبة استخدام المعالج، ومتطلبات التخزين، إضافةً إلى دقتها في كشف الشذوذ، وتصنيف استهلاك الموارد لهذه الخوارزميات لثلاثة تصنيفات. الغابة العشوائية كانت الأفضل بين الخوارزميات ذات المتطلبات العالية، محققة دقة 99% و F1-score قدره 0.99. بينما تفوقت شجرة القرار بين الخوارزميات ذات المتطلبات المتوسطة بدقة 94.6% و F1-score بلغ 0.95. أما في فئة الخوارزميات

		ذات الموارد المنخفضة، فقد كانت خوارزمية الانحدار اللوجستي الأكثر كفاءة بدقة 50.22 و f1score 0.5

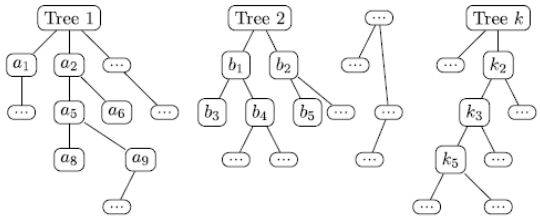
8. الخوارزميات المستخدمة

1.8 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].

2.8 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 (3):



شكل3 الفكرة الأساسية للغابة العشوائية

3.8 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.

4.8 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 [29].

5.8 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 تقسيم معايير الأداء الى معايير أداء الخوارزميات ومعايير قياس موارد أداء الخوارزميات وذلك في مرحلة اختبار الخوارزميات.

1.9 معايير قياس الموارد

تم اعتماد على الذاكرة العشوائية RAM و نسبة استخدام المعالج CPU ومساحة التخزين اللازم لتشغيل مصنف التعلم الآلي وذلك بالاعتماد على الأدوات المستخدمة في البحث.

2.9 معايير أداء الخوارزميات

تم الاعتماد على المعايير التالية نظراً لكون المشكلة هي مشكلة تصنيف البيانات على انها طبيعية او شاذة.

- **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:

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

- **Confusion Matrix**

مصفوفة الارتباك كما يوحي الاسم تعطي مصفوفة كإخراج وتصف الأداء الكامل للنموذج. يوضح الشكل 4 مصفوفة الارتباك:

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

شكل 4 مصفوفة الارتباك المصدر [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.

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

- **Recall الاسترجاع**

هو عدد النتائج الإيجابية الصحيحة مقسوماً على عدد جميع العينات ذات الصلة (جميع العينات التي كان ينبغي تحديدها على أنها إيجابية) [60]. الاسترجاع هو مقياس النموذج الذي يحدد الإيجابيات الحقيقية بشكل صحيح. رياضياً موضح بالمعادلة 4:

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

- **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.

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

10. الاعداد التجريبي

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

جدول 2 مواصفات الحاسوب المستخدم

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

11. التطبيق العملي

تم استخدام لغة البرمجة بايثون و Juypter كمحرر للتعليمات البرمجية وذلك من اجل تحليل بيانات سلاسل الكتل وتقييم استهلاك الموارد لخوارزميات التعلم الآلي في كشف الشذوذ في شبكات سلاسل الكتل واقتراح تصنيف استهلاك الموارد لكل خوارزمية بحيث يتم تصنيف الخوارزمية الى تصنيف مرتفع لاستهلاك الموارد او تصنيف متوسط لاستهلاك الموارد أو تصنيف منخفض لاستهلاك الموارد بناء على اجمالي النقاط التي حصلت عليها في استهلاك المعالج والذاكرة والتخزين لكشف الشذوذ.

1.11 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 [18].

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 بالجدول3.

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).

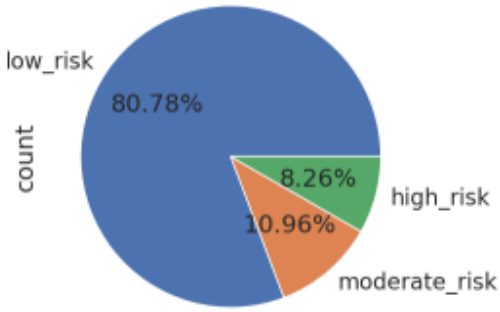
2.11 تحليل البيانات ومعالجتها

في البداية تم قراءة مجموعة البيانات كما هو موضح في الشكل 5.

timestamp	hour_of_day	sending_address	receiving_address	amount
2022-04-11 12:47:27	12	0x9d32d0bf2c00f41ce7ca01b66e174cc4dcb0c1da	0x39f82e1c09bc6d7baccc1e79e5621ff812f50572	796.949206
2022-06-14 19:12:46	19	0xd6e251c23cbf52dbd4721079147873e655d8096f	0x51e8fbe24f124e0e30a614e14401b9bbfed5384c	0.010000
2022-01-18 16:26:59	16	0x2e0925b922fed01f6a85d213aa2718f54b8ca305	0x52c7911879f783d590af45bda0c0ef2b8536706f	778.197390
2022-06-15 09:20:04	9	0x93efefc25fcaf31d7695f28018d7a11ece55457f	0x8ac3b7bd531b3a833032f07d4e47c7af6ea7baca	300.838358
2022-02-18 14:35:30	14	0xad3b8de45d63f5cce28aef9a82cf30c397c6ceb9	0x6fdc047c2391615b3facd79b4588c7e9106e49f2	775.569344

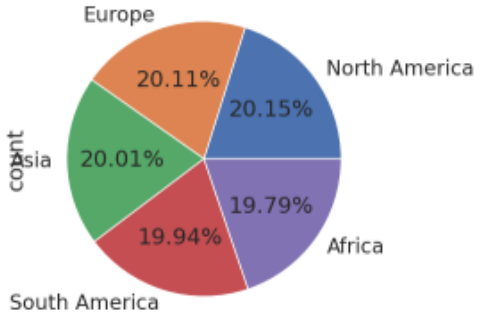
شكل 5 عرض اول خمسة اسطر من مجموعة البيانات

ليتم بعدها التأكد من خلو البيانات من القيم الفارغة او المكررة، يوضح الشكل 6 النسبة المئوية لتوزيع المعاملات ذات الخطورة العالية والمنخفضة، تشكل المعاملات ذات الخطورة المنخفضة low risk نسبة 80% من بين اجمالي المعاملات في مجموعة البيانات.



شكل 6 النسبة المئوية لتوزيع المعاملات ذات الخطورة العالية والمنخفضة

يوضح الشكل 7 مصدر المعاملات بين القارات من الواضح ان التوزيع متقارب جدا (موزع بالتساوي) بالتالي لا يرتبط مصدر المعاملات بنسبة خطورة المعاملة.

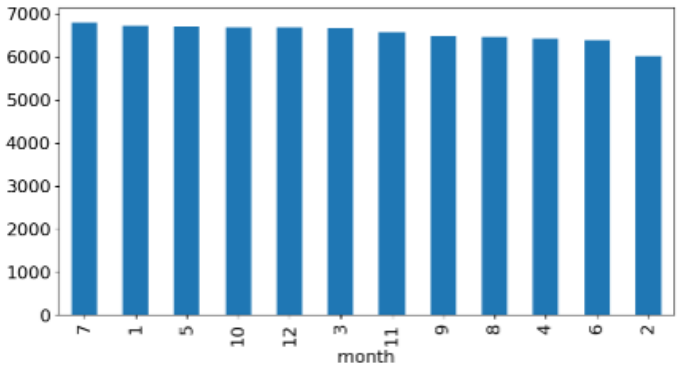


شكل 7 مصدر المعاملات

يوضح الشكل 8 نوع المعاملات في مجموعة البيانات من الواضح ان المعاملات المرتبطة بالبيع او الشراء او نقل الأموال هي الأكثر.

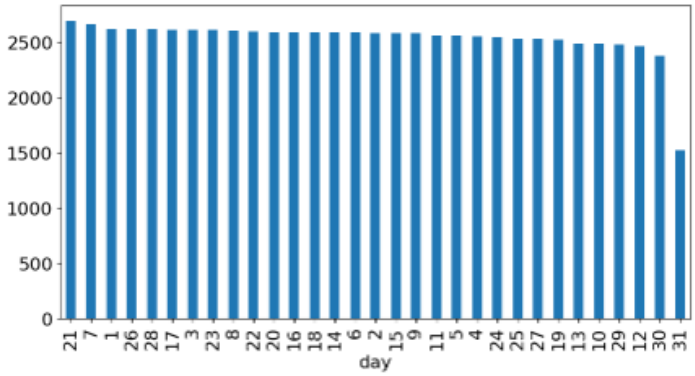
شكل 8 نوع المعاملة

من ثم تم إنشاء مجموعة جديدة من الميزات وهي السنة والشهر واليوم والدقيقة والساعة والثانية من العمود time stamp ليتم اضافتها الى مجموعة البيانات. يوضح الشكل 9 توزع المعاملات وفق الشهر، كافة الأشهر يتم فيها معاملات بنسب متقاربة.



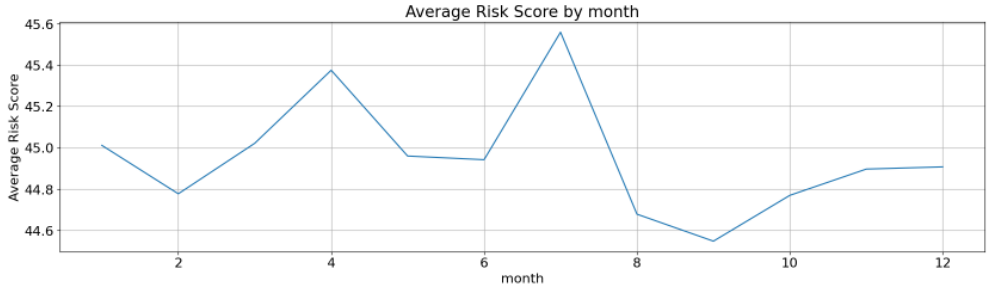
شكل 9 توزع المعاملات وفق الشهر

يوضح الشكل 10 توزع المعاملات وفق اليوم، كافة الايام يتم فيها معاملات بنسب متقاربة.



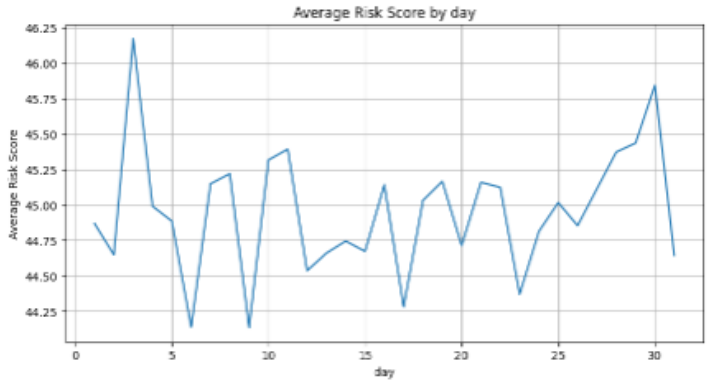
شكل 10 توزع المعاملات وفق اليوم

يوضح الشكل 11 متوسط الخطورة في المعاملات نسبة للشهر، اعلى مستوى خطورة هي في الشهر السابع والرابع.



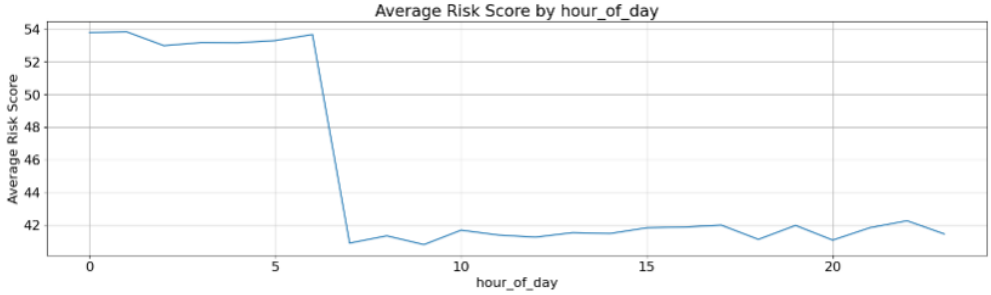
شكل 11 متوسط الخطورة في المعاملات نسبة للشهر

يوضح الشكل 12 متوسط الخطورة في المعاملات نسبة لليوم، اعلى مستوى خطورة هي في بداية أيام الشهر واخرها.



شكل 12 متوسط الخطورة في المعاملات نسبة لليوم

يوضح الشكل 13 متوسط الخطورة في المعاملات نسبة للساعة في المتوسط، هناك مخاطر عالية جدًا في ساعات الصباح الباكر (من الساعة 12 صباحاً إلى الساعة 7 صباحاً) مقارنة بالساعات المتبقية.



شكل 13 متوسط الخطورة في المعاملات نسبة للساعة

تم بعد ذلك معالجة البيانات وفق مايلي:

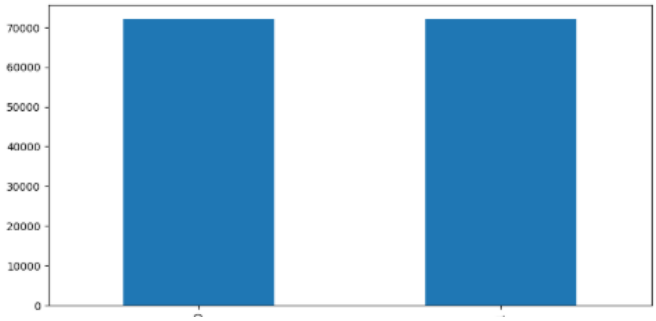
- حذف الميزة time stamp نظراً لانه تم استخلاص الميزات السابقة منه.
- حذف الميزات التالية عن طريق الاعتماد على هندسة الميزات والمعرفة بالمجال transaction_type, location_region, ip_prefix, login_frequency, session_duration, purchase_pattern, age_group, risk_score

فالميزات السابقة هي ميزات تم الحصول عليها من خصائص المستخدمين في شبكة سلاسل الكتل والمعاملات التي تم اجراءها فيها، لكن هذه الميزات لايتم اضافتها الى معلومات الكتلة في الشبكة مما يتعذر استخدامها لكشف الشذوذ في تطبيق سلاسل الكتل.

بعد ذلك تم اعتماد على المعاملات ذات الخطورة العالية high risk كمعاملات شاذة والباقي معاملات طبيعية كون نسبة الخطورة فيها متوسطة فما دون. ليكون عدد المعاملات الطبيعية 72105والمعاملات الشاذة 6495.

ليتم بعد ذلك تعيين العناوين إلى فهارس فريدة ثم ترميز العناوين الفريدة باستخدام LabelEncoder حيث يعدّ LabelEncoder أسلوب ترميز شائع للتعامل مع المتغيرات الفئوية. في هذه التقنية، يتم تعيين عدد صحيح فريد لكل عنوان عقدة. بعدها تم تقييس الميزة amount ,ويعد تقييس الميزة طريقة مستخدمة لتطبيع نطاق المتغيرات المستقلة أو ميزات البيانات. في معالجة البيانات، يُعرف أيضاً باسم تطبيع البيانات ويتم إجراؤه عموماً أثناء خطوة المعالجة المسبقة للبيانات. حيث تساعد عملية التطبيع، على سبيل المثال – في تمحور الميزات حول 0 أو في النطاق (0،1) حسب تقنية القياس وقد تم استخدام standard scaler.

ونظراً لكون البيانات غير متوازنة تم تطبيق تقنية موازنة البيانات RandomOverSampler وهي تقنية تستخدم لإجراء أخذ عينات عشوائية زائدة من الفئات الأقلية في مجموعات البيانات غير المتوازنة. وتعمل عن طريق تحديد عينات عشوائية من الفئة الأقلية مع الاستبدال لموازنة الفئات. يوضح الشكل 14 توزع البيانات بعد عملية الموازنة حيث أصبح كل فئة تحتوي على 72105 عينة.

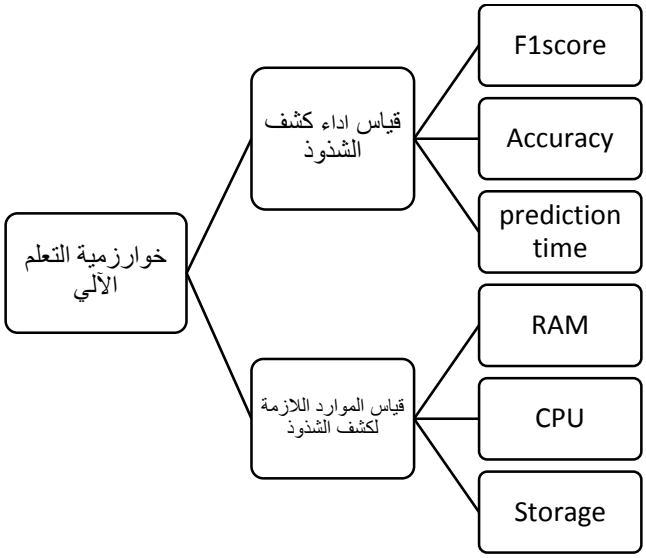


شكل 14 توزع البيانات بعد عملية الموازنة

ليتم بعدها الانتقال الي مرحلة تقسيم البيانات حيث تم تقسيم البيانات الى 80% للتدريب و 20% للاختبار.

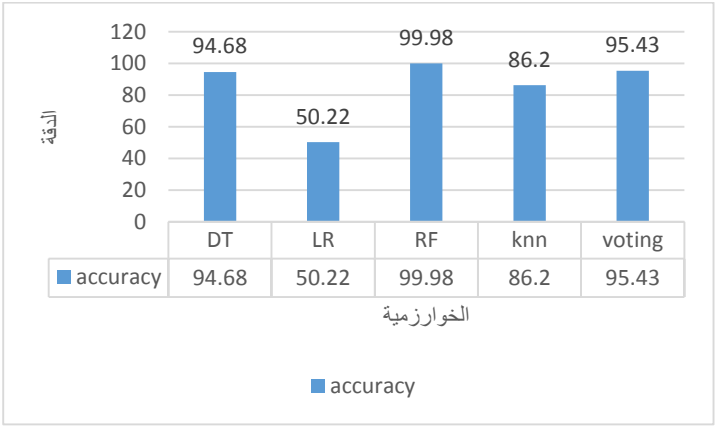
3.11 التجارب

تم استخدام نماذج التعلم الالي وهي DT,RF,KNN,LR لاكتشاف الشذوذ في سلاسل الكتل وتم قياس استهلاك الموارد لكل خوارزمية. يوضح الشكل 15 مخطط العمل في التجارب.



شكل 15 مخطط العمل في التجارب

يوضح الشكل 16 تجميع تائج الدقة لكافة المصنفات، من الواضح تفوق خوارزمية الغابة العشوائية وذلك يوافق الدراسات السابقة:

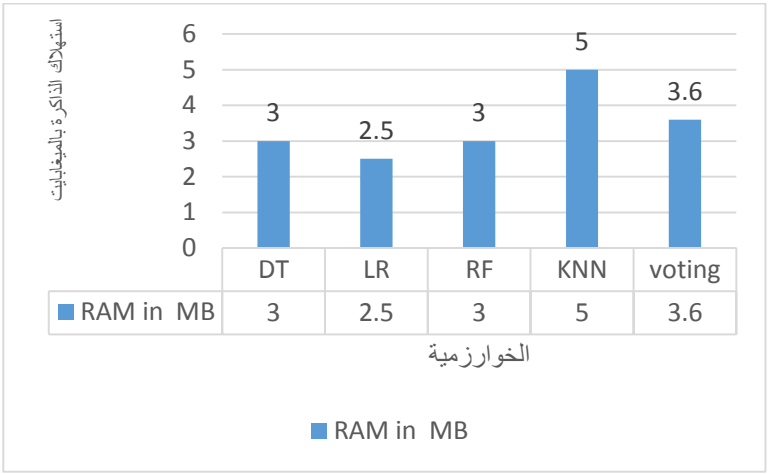


شكل 16 تجميع نتائج الدقة لكافة المصنفات

يوضح الشكل 17 تجميع f1score الدقة لكافة المصنفات، من الواضح تفوق خوارزمية الغابة العشوائية وذلك يوافق الدراسات السابقة:

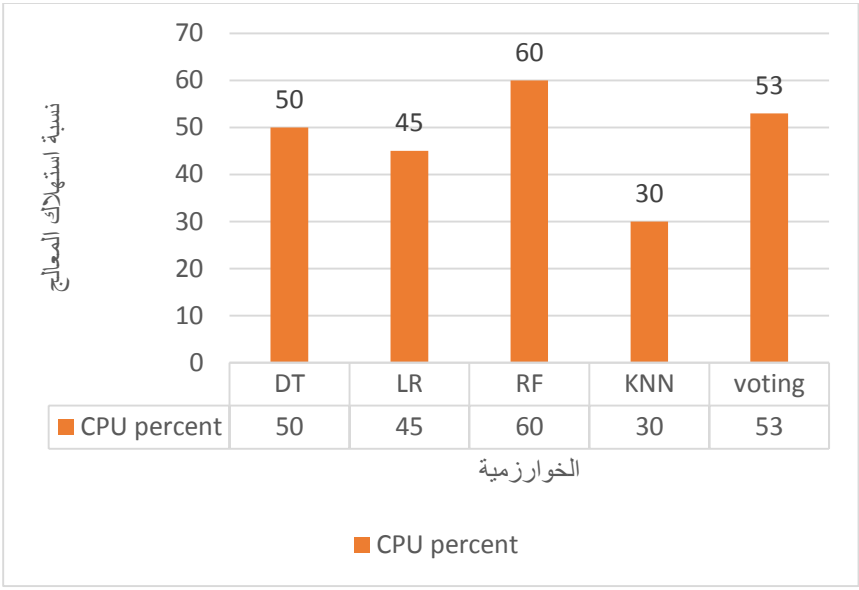
شكل 17 تجميع f1score الدقة لكافة المصنفات

يوضح الشكل 18 استهلاك الذاكرة للخوارزميات اثناء التجارب، من الواضح تقارب استهلاك الذاكرة للخوارزميات.



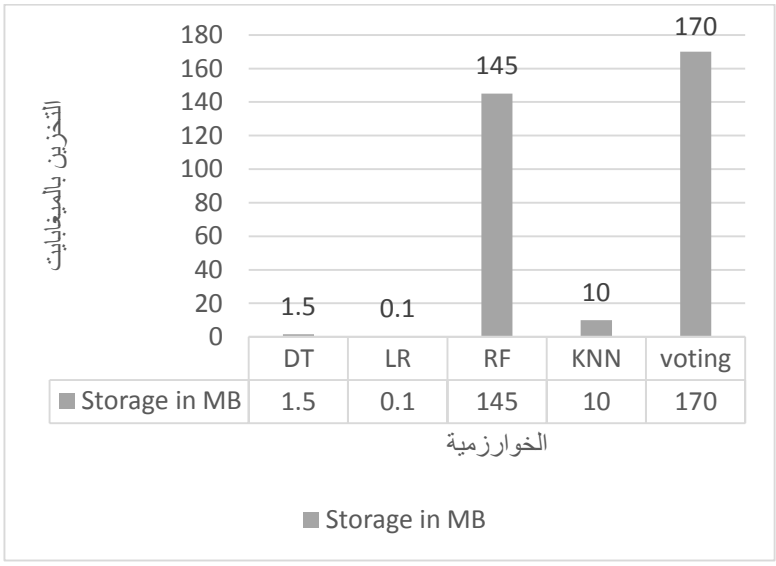
شكل 18استهلاك الذاكرة للخوارزميات اثناء التجارب

يوضح الشكل 19 نسبة استهلاك المعالج اثناء التجارب، من الواضح ان خوارزمية الغابة العشوائية والانتخاب هي من أكثر الخوارزميات استهلاكاً للمعالج.



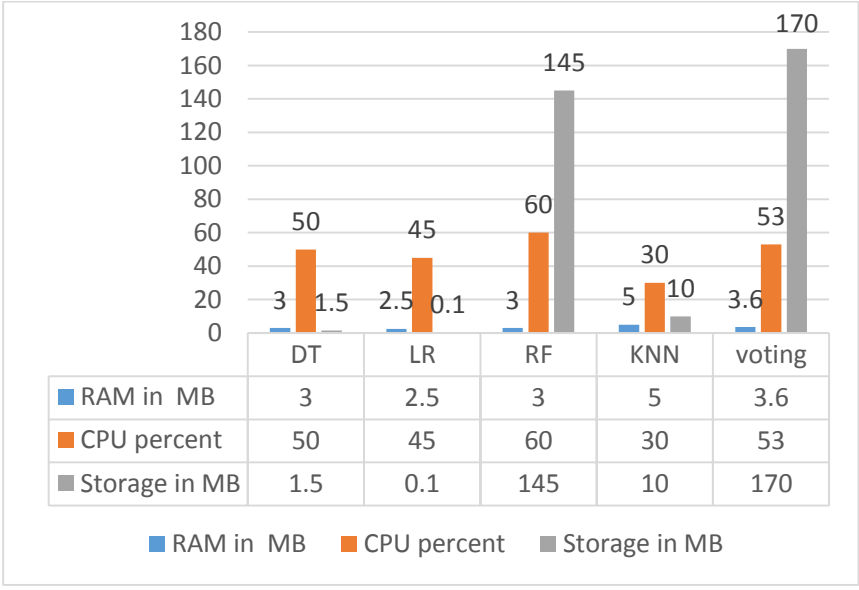
شكل 19 نسبة استهلاك المعالج اثناء التجارب

يوضح الشكل 20 مساحة التخزين اللازمة للخوارزميات، من الواضح ان خوارزمية الغابة العشوائية والانتخاب هي من أكثر الخوارزميات استهلاكا لمساحة التخزين.



شكل 20 مساحة التخزين اللازمة للخوارزميات

يوضح الشكل 21 افة الموارد التي تم استهلاكها اثناء التجارب، من الواضح ان خوارزمية الغابة العشوائية والانتخاب هي من أكثر الخوارزميات تطلباً للموارد.



	DT	LR	RF	KNN	voting
RAM in MB	3	2.5	3	5	3.6
CPU percent	50	45	60	30	53
Storage in MB	1.5	0.1	145	10	170

شكل 21 كافة الموارد التي تم استهلاكها اثناء التجارب

من النتائج السابقة يمكن التوصل الى التالي:

- **تصنيف استهلاك الموارد لكل خوارزمية**

يوضح الجدول 4 تصنيف استهلاك الموارد لكل خوارزمية بحيث يتم تصنيف الخوارزمية الى تصنيف مرتفع لاستهلاك الموارد او تصنيف متوسط لاستهلاك الموارد أو تصنيف منخفض لاستهلاك الموارد بناء على اجمالي النقاط التي حصلت عليها.

جدول 4 تصنيف استهلاك الموارد لكل خوارزمية

الخوارزمية	استهلاك المعالج (CPU %)	استهلاك الذاكرة (RAM MB)	حجم التخزين (MB)	إجمالي النقاط	التصنيف النهائي
الغابة العشوائية(RF)	60%	3 MB	145 MB	8	مرتفع (High Resource)
مصنف الانتخاب (Voting Classifier)	53%	3.6 MB	170 MB	8	مرتفع (High Resource)
شجرة القرار(DT)	50%	3 MB	1.5 MB	5	متوسط (Mid Resource)
أقرب جار(KNN)	30%	5 MB	10 MB	5	متوسط (Mid Resource)
الانحدار اللوجستي (LR)	45%	2.5 MB	0.1 MB	3	منخفض (Low Resource)

- **طريقة حساب التصنيف:**

1. **استهلاك المعالج:(CPU %)**
   - أكثر من 50% → 3 نقاط (مرتفع)
   - بين 40% و 50% → 2 نقاط (متوسط)
   - أقل من 40% → 1 نقطة (منخفض)
2. **استهلاك الذاكرة:(RAM MB)**

- **تم ملاحظة أن جميع النماذج لا تستهلك كم كبير من الذاكرة وللحفاظ على تناسق التصنيف يمكن اقتراح**
- أكثر من 4MB → 3 نقاط (مرتفع)
- بين 3MB و 4MB → 2 نقاط (متوسط)
- أقل من 3MB → 1 نقطة (منخفض)

3. **حجم التخزين المطلوب:(MB)**

- أكثر من 100MB → 3 نقاط (مرتفع)
- بين 50MB و100MB → 2 نقاط (متوسط)
- أقل من 50MB → 1 نقطة (منخفض)

التصنيف النهائي:

- **High Resource (مرتفع)** : مجموع النقاط ≤ 7
- **Mid Resource (متوسط):** مجموع النقاط بين 4 و6
- **Low Resource (منخفض):** مجموع النقاط ≥ 3

وبناء على كافة التجارب والنتائج وتصنيف النماذج تبين مايلي:

- من بين الخوارزميات ذات المتطلبات العالية كانت الغابة العشوائية افضل بدقة وصلت الى 99 % و f1score 0.99.
- من بين الخوارزميات ذات المتطلبات المتوسطة كانت شجرة القرار افضل بدقة وصلت الى 94.6% و f1score 0.95.

- من بين الخوارزميات ذات المتطلبات المنخفضة وصلت دقة LR الى 50.22 و f1score 0.5
- بغض النظر عن استهلاك الموارد كانت خوارزمية الغابة العشوائية هي اعلى نتائج من بين الخوارزميات افضل بدقة وصلت الى 99 % و f1score 0.99 في حين كانت خوارزمية LR هي الأسوء بدقة 50.22 و f1score 0.5.

12. الخلاصة والعمل المستقبلي

تم في هذا البحث دراسة وتحليل الموارد المطلوبة لتشغيل خوارزميات التعلم الالي لاكتشاف الشذوذ في سلاسل الكتل من اجل ذلك تم استخدام 5 خوارزميات تعلم الي وقياس أداء الخوارزميات لاكتشاف الشذوذ وفق معيار الدقة و f1score وتم قياس استهلاك هذه الخوارزميات للموارد وفق استهلاك الذاكرة ونسبة استهلاك المعالج والتخزين وتم بعدها تقسيم الخوارزميات بالنسبة لاستهلاك الموارد لثلاثة تصنيفات وهي تصنيف موارد منخفض وتصنيف موارد متوسط وتصنيف موارد عالي وتم تصنيف خوارزمية شجرة القرار وخوارزمية الانتخاب كموارد عالية وخوارزمية DT كموارد متوسطة وخوارزميتي LR و KNN كموارد منخفضة، فمن بين الخوارزميات ذات المتطلبات العالية كانت الغابة العشوائية افضل بدقة وصلت الى 99 % و f1score 0.99 ومن بين الخوارزميات ذات المتطلبات المتوسطة كانت شجرة القرار افضل بدقة وصلت الى 94.6 % و f1score 0.95 ومن بين الخوارزميات ذات المتطلبات الضعيفة كانت LR افضل بدقة 50.22 و f1score 0.5

للعمل المستقبلي من الممكن استخدام التعلم العميق للتحقق من القدرة على اكتشاف الشذوذ في سلاسل الكتل وقياس نسبة استهلاك الموارد لمعرفة إمكانية تضمينها ضمن عقد سلاسل الكتل.

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] Zheng, Zibin, et al. "An overview of blockchain technology: Architecture, consensus, and future trends." 2017 IEEE international congress on big data (BigData congress). Ieee, 2017.

[3] Sarmah, Simanta Shekhar. "Understanding blockchain technology." Computer Science and Engineering 8.2 (2018): 23-29.

[4] Bashir, Imran. Mastering blockchain. Packt Publishing Ltd, 2017.

[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] Song, Y. Y., & Lu, Y, "Decision tree methods: applications for classification and prediction". Shanghai archives of psychiatry, 27(2), 130–135. 2015.

[7] Hastie T, Tibshirani R, Friedman J. "The Elements of Statistical Learning", Springer, . pp: 269-272, 2001

[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] https://www.ibm.com/cloud/learn/machine-learning

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

[14]https://towardsdatascience.com/steps-of-a-machine-learning-process-7acc43973385

[15]Monowar H. Bhuyan,Dhruba K. Bhattacharyya,Jugal K. Kalita "Network Traffic Anomaly Detection and Prevention Concepts ،Techniques ،and Tools", 2017.

[16]Jose, Shijoe & Malathi, D. & Reddy, Bharath & Jayaseeli J D, Dorathi. "A Survey on Anomaly Based Host Intrusion Detection System". Journal of Physics, 2018.

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

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

[19] Siddique, Quba. "Anomaly Detection in Blockchain Transactions within the Metaverse Using Anomaly Detection Techniques." Journal of Current Research in Blockchain 1.2 (2024): 155-165.

[20] Shayegan, Mohammad Javad, et al. "A collective anomaly detection technique to detect crypto wallet frauds on bitcoin network." Symmetry 14.2 (2022): 328.

[21] Singh, P., Agrawal, D., & Pandey, S. (2023). Anomaly detection and analysis in blockchain systems.

[22] Airlangga, G. (2024). Anomaly Detection in Blockchain Transactions: A Machine Learning Approach within the Open Metaverse. Jurnal Informatika Ekonomi Bisnis, 308-312.

[23] Samantha Jeyakumar , Eugene Yugarajah , Andrew Charles , Punit Rathore , et al. Feature Engineering for Anomaly Detection and Classification of Blockchain Transactions. TechRxiv. March 27, 2023.

[24] 9.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. 2022; 22(19):7162. https://doi.org/10.3390/s22197162

[25] 12. Hisham, S., Makhtar, M., & Aziz, A. A. (2023). An interpretable ensemble model framework for real-time anomaly detection

and prediction of Ethereum blockchain transactions. International Journal of Advanced Technology and Engineering Exploration, 10(103), 676.

[26] Sallam, Amer, et al. "Fraudulent account detection in the Ethereum's network using various machine learning techniques." International Journal of Software Engineering and Computer Systems 8.2 (2022): 43-50.

[27] 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.

[28] Airlangga, Gregorius. "Deep Learning for Anomaly Detection and Fraud Analysis in Blockchain Transactions of the Open Metaverse." Jurnal Informatika Ekonomi Bisnis (2024): 324-329.

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

[30] Eng. Mohammad khalil, Dr. Mohammed Alchaita, Dr. Mohammad Assoura, Study of the blockchain and its potential use in verifying ownership of the public key in sending encrypted messages,Homs University Journal, Volume 43, Issue 10, 2021

[31] 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.

[32] 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