

نموذج تعلم آلي ذكي لكشف الاحتيال في بطاقة الائتمان

- تطبيق عملي على منصة تفاعلية -

A Smart Machine Learning Model for Credit Card Fraud Detection - A Practical Application on an Interactive Platform -

فراحي فضيلة

جامعة الجزائر 3 (الجزائر)، frahifad14@gmail.com

تاريخ النشر: 2025/05/25

تاريخ القبول: 2025/04/29

تاريخ الاستلام: 2025/03/12

ملخص:

تقدم هذه الدراسة نموذج تعلم آلي ذكي يعتمد على خوارزمية الغابة العشوائية، أحد أهم خوارزميات تعلم الآلة بالإشراف، وهي فرع من فروع الذكاء الاصطناعي، حيث تُستخدم الخوارزمية في الكثير من مجالات المالية والاقتصاد. يتم التركيز في هذه الورقة على تخريج النموذج كمنصة تفاعلية في البيئات العملية على شكل نظام لكشف الاحتيال في بطاقات الائتمان، لتسهيل استخدامه في المؤسسات المالية.

بالإضافة إلى دوره الهام في هذه المؤسسات الحيوية، يمكن لهذا النموذج أن يكون أداة فعالة للأجهزة الأمنية والمخابراتية، في كشف المحتالين والمجرمين الذين يستغلون الأنظمة المالية لأغراض غير مشروعة. توصلت النتائج إلى تحديد أهم المتغيرات كمميزات أساسية، تساعد في تحليل البيانات واكتشاف الأنماط الاحتيالية، مما يساهم في تعزيز الأمن المالي وحماية الاقتصاد الوطني من التهديدات الإجرامية.

كلمات مفتاحية: خوارزمية الغابة العشوائية، كشف الاحتيال، بطاقة الائتمان.

تصنيف JEL: C45، G28.

Abstract:

This study presents an intelligent machine learning model based on the Random Forest algorithm, one of the most significant supervised learning algorithms and a key branch of artificial intelligence. The algorithm is widely used in various financial and economic applications. This paper focuses on deploying the model as an interactive platform in real-world environments, serving as a system for credit card fraud detection to facilitate its use in financial institutions.

In addition to its crucial role in these vital institutions, the model can also serve as an effective tool for security and intelligence agencies in identifying fraudsters and criminals who exploit financial systems for illicit purposes. The findings highlight the key variables as features importance, that contribute to data analysis and fraud pattern detection, thereby enhancing financial security and protecting the national economy from criminal threats.

Keywords: Random Forest Algorithm, Credit Card, Fraud Detection.

JEL Classification: C45، G28.

1. مقدمة:

يشهد العالم تطورات متسارعة في مجال التكنولوجيا المالية، ما أدى إلى تزايد التحديات المرتبطة بالأمان المالي، خصوصاً فيما يتعلق بالاحتيال في بطاقات الائتمان. يعتبر الاحتيال المالي أحد أكبر التهديدات التي تواجه المؤسسات المالية، مما يتطلب حلولاً ذكية تعتمد على الذكاء الاصطناعي لتحديد ومنع المعاملات الاحتيالية بفعالية. ورغم أن الكشف عن الاحتيال المالي موضوع ناشئ ذو أهمية كبيرة في العصر الحديث، إلا أنه لم يتم التطرق إليه بشكل يسمح بتحديد المتغيرات المؤثرة في معظم الدراسات التطبيقية المنجزة. وقد تم تحليل أهم المقالات من الجانب النظري حول هذا الموضوع نُشرت بين عامي 1997 و2008، حيث صنفت الاحتيال المالي إلى أربع فئات وهي الاحتيال المصرفي، الاحتيال في التأمين، والاحتيال في الأوراق المالية والسلع، والاحتيال الذي يتصل بصفة غير مباشرة بالمال¹. وبعض من مقالات أخرى تطبيقية تناولت تحليل البيانات من حيث التصنيف، والانحدار، والتجميع، والتنبؤ، والكشف عن القيم الشاذة لتحديد وجود احتيال من عدمه. تمثل هذه الورقة محاولة تطبيقية تهدف إلى استعراض نموذج تعلم آلي ذكي يعتمد على خوارزمية الغابة العشوائية يتميز دوره في كشف الاحتيال في بطاقات الائتمان ثم استخراجها على شكل تطبيق عملي في منصة تفاعلية على واجهة محلية للويب. لا يقتصر دور هذا النموذج على المؤسسات المالية فحسب، بل يمتد لي شمل الأجهزة الأمنية والمخابراتية مثل الشرطة والدرك، حيث يمكن توظيفه في كشف المحتالين والمجرمين الذين يستخدمون المعاملات المالية غير المشروعة لإخفاء أذيتهم الإجرامية. من خلال تحليل البيانات المالية واكتشاف الأنماط غير الاعتيادية، يمكن للنموذج تقديم مؤشرات مبكرة عن السلوكيات المشبوهة، مما يساعد الجهات المختصة في تعقب العمليات الاحتيالية ومنع الجرائم المالية قبل وقوعها. كما يمكن دمج هذا النموذج مع أنظمة تحليل البيانات الجنائية لتعزيز الأمن القومي، والتصدي لعمليات غسل الأموال، والحد من التهديدات المالية التي قد تؤثر على الاستقرار المالي والقومي للدول.

1.1 إشكالية البحث:

تسعى هذه الدراسة إلى معالجة مشكلة تزايد عمليات الاحتيال المالي في بطاقات الائتمان، من خلال تطوير نموذج تعلم آلي ذكي قادر على التمييز بدقة بين المعاملات الاحتيالية وغير الاحتيالية، مع إمكانية نشره في منصة تفاعلية تُسهّل استخدامه من قبل المؤسسات المالية. وتكمن الإشكالية في كيفية اختيار الخوارزمية الأنسب، وتحديد المتغيرات الأكثر تأثيراً، وتصميم تطبيق تفاعلي عملي يساهم في اتخاذ قرارات فورية وفعّالة، وعليه يتم طرح الإشكالية كما يلي:

✓ كيف يمكن تطوير نموذج تعلم آلي ذكي باستخدام خوارزمية الغابة العشوائية، يتميز بالدقة والفعالية في كشف الاحتيال في بطاقات الائتمان، مع إمكانية تطبيقه عملياً من خلال منصة تفاعلية تُخدم المؤسسات المالية والأمنية؟

2.1 أسئلة البحث:

- يمكن لنموذج الغابة العشوائية تحسين دقة كشف الاحتيال، والتقليل من نسبة المعاملات الاحتيالية قبل تنفيذها؟
- ما هي المتغيرات الأكثر أهمية التي يعتمد عليها في تحديد معاملات الاحتيال المالي؟
- كيف يمكن بناء تطبيق تفاعلي يجعل النموذج أكثر قابلية للاستخدام ويساهم في اتخاذ قرارات سريعة.

3.1 فرضيات البحث:

- يمكن لنموذج الغابة العشوائية تحسين دقة كشف الاحتيال، والتقليل من نسبة المعاملات الاحتيالية قبل تنفيذها.

- يمكن تحديد المتغيرات الأكثر أهمية التي يعتمد عليها نموذج الغابة العشوائية من أجل كشف معاملات الاحتيال المالي.
- يمكن بناء منصة تفاعلية فعالة تسهل استخدام نموذج كشف الاحتيال من قبل المؤسسات المالية وتساهم في اتخاذ قرارات سريعة ودقيقة.

4.1 منهجية الدراسة:

نظرا لطبيعة الموضوع ومن أجل تحقيق الهدف، اعتمدت الدراسة على المنهج الوصفي لوصف الظاهرة المدروسة، وعلى المنهج التجريبي، للتفاعل العملي مع الظاهرة المدروسة.

5.1 حدود الدراسة:

- الحدود الموضوعية: تناولت الدراسة توظيف خوارزمية الغابة العشوائية في بناء نموذج ومنصة تفاعلية لتحديد المتغيرات التي تساهم في التعرف على معاملة احتيالية من عدمها في المجال المالي، مما يتيح للمؤسسات توجيه جهودها نحو تطوير أدواتها.
- الحدود البشرية: 3471 ملاحظة مع 9 متغيرات، مستخرجة من حاملي بطاقات الائتمان ¹ NerdWallet
- الحدود المكانية: تغطي هذه الدراسة بيانات معاملات مالية إلكترونية خاصة بحاملي بطاقات ائتمان ضمن محافظ رقمية تتبع لمؤسسات مالية مرخصة في كل من الولايات المتحدة الأمريكية، المملكة المتحدة، كندا، أستراليا.
- وذلك ضمن خدمات مصرفية مقدمة عبر ترخيص من Evolve Bank & Trust، مما يجعل البيئة المكانية للدراسة متعددة الجنسيات وتعكس واقع مالي دولي متنوع.

6.1 أهمية الدراسة:

تتبع أهمية هذه الدراسة من الحاجة المتزايدة إلى حلول ذكية وفعالة لمواجهة التهديدات المالية المتزايدة، وعلى رأسها الاحتيال في بطاقات الائتمان، الذي يمثل خطراً كبيراً على استقرار المؤسسات المالية والاقتصاد الوطني. تبرز الدراسة من خلال توظيف خوارزمية الغابة العشوائية إحدى أكثر خوارزميات تعلم الآلة دقة ومرونة لتصميم نموذج ذكي قادر على كشف المعاملات الاحتيالية بدقة عالية. كما تساهم في تعزيز أمان الأنظمة المصرفية عبر تقديم منصة تفاعلية قابلة للتطبيق الفعلي في البيئات المالية، مما يسهل اتخاذ قرارات فورية ودقيقة. وتكمن أهمية إضافية للدراسة في إمكانية توظيف النموذج من قبل الأجهزة الأمنية لتحليل الأنماط الاحتيالية والتنبؤ بها قبل وقوعها، مما يعزز من فعالية الرقابة المالية والوقاية من الجرائم الاقتصادية.

7.1 الدراسات السابقة:

- يتوسع محتوى هذا البحث ويكمل أهم بعض الدراسات السابقة ذات الصلة ومنها:
- دراسة بعنوان استطلاع واسع النطاق حول تقنيات التعلم الآلي وطرق إثبات صحة بيانات المستخدم للكشف عن الاحتيال ببطاقات الائتمان. لكل من: ² Karthick & Berkman سنة 2023، تناولت الدراسة في القسم الأول نماذج تقليدية، تركز على (استراتيجية المعرفة). أما القسم الثاني، ركزت بشكل أكبر على عملية التحقق من العميل، وقياسات سلوكه الحيوية لتحديد السلوكيات غير العادية للشخص أثناء استخدام أجهزته الإلكترونية.
- دراسة بعنوان مجموعة شبكات عصبية مع هندسة الميزات لتحسين اكتشاف الاحتيال ببطاقات الائتمان. لكل من المؤلفين: ³ Esenogho, E., والآخرين سنة 2022، اقترحت نهجاً فعالاً للكشف عن احتيال بطاقات الائتمان باستخدام مصنف مجموعة الشبكة العصبية طويلة وقصيرة المدى مع طريقة إعادة أخذ العينات الهجينة للبيانات. ثم تقارن أداء النهج المقترح

¹ <https://www.nerdwallet.com/article/credit-cards/credit-card-data>

بالخوارزميات التالية: آلة الدعم المتجه، الشبكة العصبية متعددة الطبقات، و شجرة القرار، وخوارزمية تعزيز التدرج التقليدية، أظهرت النتائج أن أداء المصنّفات كان أفضل عند تدريبها باستخدام البيانات المعاد أخذ عينات منها، وأن مجموعة الشبكة العصبية طويلة وقصيرة المدى المقترحة تفوقت على الخوارزميات الأخرى.

- دراسة بعنوان الكشف عن الاحتيال في بطاقات الائتمان باستخدام الانحدار اللوجستي. لكل من: M. Devika; S. Ravi Kishan والآخرين، حيث قاموا باستخدام أنواعاً مختلفة من نماذج وخوارزميات التعلم الآلي لتحديد الاحتيال في المعاملات المالية. في النتائج تم اختيار خوارزمية الانحدار اللوجستي التي قدمت دقة عالية للكشف عن الاحتيال في البطاقات المالية.

2. الإطار النظري: مفاهيم حول جوانب الدراسة

يشهد العالم تطوراً متسارعاً في تقنيات الذكاء الاصطناعي وتعلم الآلة، الأمر الذي مكن القطاعات المالية من تحسين إجراءاتها الأمنية، خاصة في كشف الاحتيال في بطاقات الائتمان. يواجه القطاع المالي تحديات متزايدة مع تزايد العمليات الرقمية، مما يفرض الحاجة إلى حلول مبتكرة وقوية تستند إلى التحليل الذكي للبيانات الضخمة، واستخدام أدوات وتقنيات حديثة وسريعة.

1.2 الاحتيال في بطاقة الائتمان:

الاحتيال في بطاقة الائتمان هو أي استخدام غير مصرح به أو غير قانوني لمعلومات البطاقة المالية من أجل إجراء معاملات مالية دون علم أو موافقة صاحبها. يمكن أن يشمل ذلك سرقة البطاقة الفعلية، أو اختراق بياناتها عبر الإنترنت، أو استخدام بيانات مسروقة لإنشاء بطاقات وهمية. يشكل هذا النوع من الجرائم تهديداً كبيراً للمؤسسات المالية والأفراد، مما يستلزم تطوير أنظمة متقدمة للكشف عن الاحتيال ومنعه قبل وقوعه.⁵

2.2 أنواع الاحتيال في بطاقات الائتمان وأهمية خوارزميات التعلم الآلي:

يأخذ الاحتيال في بطاقات الائتمان عدة أشكال، منها الاحتيال المباشر من خلال سرقة البطاقة، أو سرقة البيانات عبر الاختراقات الرقمية، أو الاحتيال غير المباشر مثل تزوير الهويات لإنشاء حسابات وهمية. لكل نوع من هذه الأنواع تحدياته الخاصة التي تتطلب تقنيات مختلفة لكشفها، تعد خوارزميات تعلم الآلة وهي أحد أهم فروع وتقنيات الذكاء الاصطناعي، ومن الأساليب الفعالة في تحليل أنماط الاحتيال في وقت قصير. حيث يمكن لهذه الخوارزميات استيعاب كميات هائلة من البيانات، والتعرف على السلوكيات المشبوهة التي قد يصعب كشفها عبر الطرق التقليدية. ومن خلال التعلم والتدريب المستمر، تتطور النماذج الذكية لتصبح أكثر دقة في التعرف على الاحتمالات المستقبلية.⁶

3.2 الخوارزميات المستخدمة في كشف الاحتيال وتحديات المواجهة في تطبيقنا

تتنوع خوارزميات تعلم الآلة المستخدمة في كشف الاحتيال في بطاقات الائتمان، حيث تشمل خوارزميات الانحدار اللوجستي التي تستخدم خاصية من أجل تصنيف المعاملات المشبوهة بناءً على مجموعة من المتغيرات المستقلة والمتغير المستهدف. والشبكات العصبية العميقة، التي تستخدم في تحليل البيانات المعقدة والتعلم العميق لاكتشاف أنماط غير معروفة سابقاً. وبعض من تقنيات التعلم غير الخاضع للإشراف، الذي يبرز استخداماته لاكتشاف الحالات غير الاعتيادية دون الحاجة إلى بيانات مصنفة من قبل. حالياً تتوجه معظم الأدبيات إلى استخدام خوارزمية شجرة القرار والغابة العشوائية حيث تعتمد على بناء عدة

نماذج تنبؤية تعمل في آن واحد من أجل زيادة دقة كشف الاحتيال. على الرغم من الفوائد الكبيرة لهذه الاستخدامات، إلا أن هناك عدة تحديات تواجه تطبيقها، من أهمها طرح اشكالات:

- التوازن بين الدقة و سهولة الاستخدام: حيث تحتاج الأنظمة إلى تحقيق توازن بين تقليل الأخطاء الإيجابية الكاذبة وتقليل عمليات الاحتيال الفعلية.
- مشكلات البيانات: مثل نقص البيانات أو تحيزها، أو انعدام التوازن في البيانات ذاتها مما قد يؤثر على دقة النماذج.
- الهجمات المضادة: التي يلجأ إليها المحتال، حيث يسعى إلى التكيف مع الأنظمة الذكية وإيجاد طرق جديدة لتجاوزها باستمرار.⁷

3.2 العلاقة بين كشف الاحتيال والاستقرار الاقتصادي

تسبب عمليات الاحتيال في الأموال حالة من عدم الاستقرار الاقتصادي؛ حيث غالباً ما ترتبط الأنشطة الاحتيالية بالمصارف والبنوك والمؤسسات المالية؛ إذ يحاول المتورط الإضرار بالمصرف أو بالعملاء، مما يؤثر سلباً على النظام بأسره، فقد يُسبب في فقدان عملاء ذوي قيمة معتبرة، إضافة إلى ما ينطوي عليه الوضع من تأثير المصرف مالياً خاصة إذا كانت العمليات الاحتيالية متكررة وبمجم ضخم، أو تضمن الاحتيال تحويل مبالغ إلى خارج الوطن، فتتأثر بذلك قيمة العملة الوطنية، وتبرز مشاكل فقدان السيولة، وعرقلة مشاريع الاستثمارات، حيث تستدعي الحاجة إلى وضع المزيد من القيود التنظيمية، مثل الحد من الاحتيال بجميع أنواعه، والتي يتم بموجبها يحصل تعزيز الثقة في الأنظمة المالية، وتقليل الخسائر الناجمة عن التهديدات المحتملة ومختلف جرائم الاحتيال، مما يساهم في كفاءة الأنظمة المالية خاصة باستخدام تقنيات الذكاء الاصطناعي، وبالتالي تؤثر إيجاباً على استقرار الاقتصادات الوطنية.⁸

4.2 التعلم الموجه

هو نوع من التعلم الآلي يعتمد على تدريب النموذج باستخدام بيانات مصنفة (مدخلات مع مخرجات معروفة). يتم تعليمه التنبؤ بالمخرجات الجديدة من خلال التعميم من بيانات التدريب.

5.2 شجرة القرار

نموذج يُستخدم للتصنيف أو التنبؤ، يعتمد على اتخاذ قرارات وفقاً لمجموعة شروط منطقية تتفرع بشكل هرمي. كل عقدة تمثل خاصية، وكل فرع يمثل قيمة أو شرط، وكل ورقة مدرجة تمثل نتيجة. ومجموعة من أشجار القرار تشكل الغابة العشوائية.

6.2 ماهية الغابة العشوائية

الغابة العشوائية⁹ هي خوارزمية من خوارزميات تعلم الآلة، وتعلم الآلة بدوره فرع من فروع الذكاء الاصطناعي، تعتبر الغابة في هذا السياق مجموعة من أشجار القرار العشوائية كطريقة شائعة تعتمد على هيكل شجري لاتخاذ القرارات بخصوص الافراد الداخلة في عملية التعلم. شجرة القرار سهلة الفهم والإدراك لا يحتاج المستخدم إلى الكثير من المعرفة الخلفية في الرياضيات والاحصاء لكي ينفذ عملية التعلم من البيانات باستخدام الغابة العشوائية، وغالباً ما يكون إعداد البيانات بسيطاً أو غير ضروري في الكثير من الأحيان، كما يمكن التعامل مع عدد ضخم من البيانات وتحليل السمات المعقدة في نفس الوقت، سواء كانت العلاقة التي تحكمها خطية أو غير خطية حيث لا تحتاج إلى إجراء افتراضات خاصة ثم يتم استخراج نتائج قابلة للتطبيق وفعالة في وقت قصير نسبياً. تنوزع البيانات بواسطة كتل العوامل فتجمع الميزات المتشابهة في العينة تحت العقدة الواحدة ثم تتفرع تدريجياً على شكل عقد تحت شروط معينة. تتشكل خاصية التجميع عندما تكون العقدة الفرعية لشجرة القرار تحتوي على

جميع العينات. كلما كان نقاء تجميع العينات تحت العقدة أعلى، كلما كان ذلك أفضل. عند إجراء التنبؤات في مجموعة العينة ذات الخصائص المتشابهة، يتم استخدام المتغير التابع كنتيجة متوقعة (أي تصنيف المتغير المستهدف بالنسبة للمتغيرات المستقلة)، يتم استخدام طريقة التصويت بالأغلبية للحصول على النتائج المتوقعة. عندما يكون المتغير التابع كميًا، يتم استخدام متوسط المتغير التابع كنتيجة للتنبؤ. كما يمكن التعامل بشكل جيد أيضا مع المتغيرات المستقلة النوعية بعد تحويلها وهذا ما يميز خوارزمية شجرة القرار عن غيرها من الخوارزميات.

1.6.2 لفرق بين الغابة العشوائية وبقية خوارزميات التعلم الموجه التصنيفي

تتكون الغابة العشوائية كما تقدم من مجموعة من أشجار القرار، تستخدم أسلوب "تصويت الأغلبية" في التصنيف، مما يحسن من الدقة ويقلل من فرط التكيف، تتميز بالقدرة على التعامل مع البيانات غير المتوازنة والمتغيرات التوضيحية المتنوعة. أما الخوارزميات الأخرى مثل:

- الانحدار اللوجستي: فهي مناسبة للبيانات الخطية، بسيطة الاستعمال وغالبا ما تعطي نتائج أقل دقة في حالات التعقيد.

- الشبكات العصبية: هي خوارزميات قوية، لكنها تتطلب موارد كبيرة أي بيانات ضخمة ووقت تدريب طويل.

- شجرة القرار المنفردة: هي سريعة ولكن عرضة لفرط التكيف، مما يجعل الغابة العشوائية أكثر استقراراً منها.

3. الإطار التطبيقي: استخدام خوارزمية الغابة العشوائية لكشف الاحتيال في بطاقة الائتمان

تعتبر خوارزمية الغابة العشوائية من أكثر أساليب تعلم الآلة فعالية في كشف الاحتيال المالي، نظراً لقدرتها على تحليل بيانات ضخمة والتعامل مع المتغيرات المتعددة. تعتمد مجموعة من أشجار القرار التي تعمل معا لاكتشاف الأنماط المشبوهة¹⁰ في البيانات المالية، مما يزيد من دقة التنبؤ مقارنة بالنماذج الأخرى. بفضل قدرتها على التكيف مع التغيرات في بيانات الاحتيال، توفر الغابة العشوائية نموذجاً ديناميكياً يمكنه رصد السلوكيات غير الاعتيادية والتعرف على العمليات الاحتمالية بسهولة.

1.3 أداة الدراسة:

تم استخدام برمجية بايثون وتوظيف مكتبة **Scikit-learn** وهي ال شاملة للتعلم الآلي، والأكثر شيوعاً لا استخدام الغابة العشوائية، توفر خوارزميات الغابة العشوائية لكل من التصنيف والانحدار مع واجهة متسقة توفر أدوات لتحليل أداء النموذج كدقة التنبؤ، مصفوفة الارتباك، وأهمية الميزات.¹¹

2.3 إجراءات الدراسة:

- بعد معالجة البيانات، بدءاً من تنظيفها وموازنتها أي تصحيح عدم توازن الفئات **Class Imbalance Correction** أثناء التدريب. ثم تعيين القيم الرقمية المناسبة لكل الفئات وتحديد المتغيرات المستقلة والمتمثلة في:

Time: وهو الزمن محدد بالساعات، محدد على الجدول رقم (2)، في المجال: 0-167.

Amount: المبلغ بالدولار الأمريكي. وأعلى مبلغ في العمليات الاحتمالية محدد 5000 دولار.

Transaction_Type: نوع المعاملة، وهي على ثلاث مستويات، أعلى مستوى يعني تسجيل معاملة غير مألوفة أو مشتبهاً بها.

Location: الموقع وهي مجموعة المواقع المختلفة التي تستقبل منها العمليات، موقع 4 تعني منطقة غير متوافقة مع سجل العميل.

Device_Type: نوع الجهاز ، هو على ثلاث أنواع كمبيوتر، مصرف آلي، أو هاتف (جهاز جديد أو غير مألوف).
Num_Transactions: عدد المعاملات في آخر 24 ساعة، حيث يحدد الاحتيال في الغالب بعدد مرتفع جداً من العمليات في وقت قصير.

Transaction_Time: وقت المعاملة على أربع مستويات كما يظهر في الجدول رقم (2)، حيث اعلى مستوى يعني تنفيذ العملية في ساعة متأخرة من الليل، وهي في الغالب دالة على عملية احتيال.

Distance_From_Home: المسافة بين المصرف المالي وعملية تنفيذ التعامل. كلما كانت المسافة متباعدة مع خلو سجل المستعمل من أي سفر أو تنقل، تزيد احتمالية الاحتيال.

Is_Fraud: وهو المتغير المستهدف الذي يحدد إذا كان هناك احتيال من عدمه. يأخذ مستويان إما 0= ليس هناك احتيال، أم 1= هناك احتيال.

- يتم تقسيم البيانات إلى بيانات تدريب وبيانات اختبار بحدود 80-20% لكل قسم على الترتيب. يوضح الجدولين رقم (1) و رقم (2) الملاحظات والمواصفات الإحصائية لجميع المتغيرات قيد الدراسة.

الجدول رقم (1): عرض بيانات الدراسة

	Time	Amount	Transaction_Type	Location	Device_Type	Num_Transactions_Last_24h	Transaction_Time_Of_Day	Distance_From_Home	Is_Fraud
0	102	3383.750000	2	1	1	10	0	12.240000	0
1	92	652.810000	0	2	0	19	1	18.820000	1
2	14	1174.750000	1	0	1	14	1	15.050000	0
3	106	3934.820000	0	3	0	13	2	5.480000	1
4	71	4853.980000	0	0	2	15	1	24.490000	0
...	...	...	...	...	...	...	...	...	...
3466	100	3697.981821	1	3	0	11	1	32.331364	1
3467	33	4874.350223	0	1	1	5	1	14.566323	1
3468	37	1510.573327	0	0	0	10	0	6.385811	1
3469	156	4960.193523	0	1	2	17	2	33.166719	1
3470	17	2168.624679	0	0	1	9	1	18.865651	1

3471 rows × 9 columns

المصدر: من إعداد الباحثة باستخدام برمجية بايثون وبيانات الدراسة.

يعرض الجدول رقم (1) بيانات تضم معلومات عن المعاملات المالية، وتتكون من 3471 صف و 9 أعمدة و هي المتغيرات الدالة على: وقت المعاملة، وقيمتها، والموقع الذي تمت فيه، والجهاز المستخدم، وعدد المعاملات المحررة خلال 24 ساعة، وتوقيت المعاملة خلال اليوم (مقسم لفترات، مثل: صباح، مساء... إلخ)، ثم المسافة من المنزل (قد تكون بالكيلومترات، ليتم تحديد فيما إذا كانت المعاملة احتيالية أم لا. مثال من البيانات:

الصف الأول يمثل معاملة بقيمة 3383.75، تم إجراؤها من نوع 2، من موقع 1، باستخدام جهاز من نوع 0، وكانت المعاملة العاشرة في آخر 24 ساعة، وقتها في الفترة الزمنية 1، على بعد 12.24 كم من المنزل، ولم تكن

احتيالية أي: **Is_Fraud = 0**

الجدول رقم (2): الموصفات الإحصائية لبيانات الدراسة

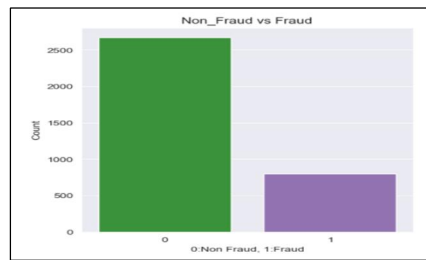
	Time	Amount	Transaction_Type	Location	Device_Type	Num_Transactions_Last_24h	Transaction_Time_Of_Day	Distance_From_Home	Is_Fraud
count	3471.0	3471.0	3471.0	3471.0	3471.0	3471.0	3471.0	3471.0	3471.0
mean	85.0	2514.0	1.0	2.0	1.0	10.0	1.0	25.0	0.0
std	49.0	1430.0	1.0	1.0	1.0	5.0	1.0	14.0	0.0
min	0.0	6.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
25%	41.0	1294.0	0.0	1.0	0.0	5.0	1.0	14.0	0.0
50%	86.0	2520.0	1.0	2.0	1.0	10.0	1.0	25.0	0.0
75%	128.0	3759.0	2.0	3.0	2.0	15.0	2.0	37.0	0.0
max	167.0	4999.0	2.0	4.0	2.0	19.0	3.0	50.0	1.0

المصدر: من إعداد الباحثة باستخدام برمجية بايثون وبيانات الدراسة.

الجدول رقم (2) يُظهر أن المتوسط العام لقيمة المعاملة هو 2514، بينما الحد الأقصى هو 4990، ما يشير إلى تفاوت واضح في القيم. معظم المعاملات تحدث خلال 10 ساعات من اليوم وباستخدام جهاز من نوع 2، وفي موقع 1. نسبة الاحتيال منخفضة جداً (الوسيط والربع الثالث لقيمة $Is_Fraud = 0$ ، مما يشير إلى أن عدد المعاملات الاحتيالية قليل مقارنة بالإجمالي، كما يمكن تحليل الخصائص الإحصائية كما يلي:

- **الانحراف المعياري:** نلاحظ أن قيمة المعاملة **Amount** لها انحراف معياري مرتفع (1430)، وهذا يعني أن هناك تشتت عالي في القيم حول المتوسط. نفس الأمر ينطبق على المتغير: **Distance_From_Home** مما يشير إلى تباين كبير في المسافات بين المعاملات.
- **التوقع:** التوقع يعطي فكرة عن القيمة المتوقعة في العموم. مثلاً، المعاملة المتوسطة قيمتها حوالي 2514، والمسافة المتوسطة من المنزل 25 كم، وهذا يساعد في بناء خوارزميات تصنيف طبيعي مقابل غير طبيعي.
- **القيم الشاذة:** بالنظر إلى الفرق الكبير بين القيم القصوى والربع الثالث مثلاً في المتغيرة: **Amount** حيث الربع الثالث يبلغ 3750 بينما الحد الأقصى هو 4990، بالإضافة إلى الانحراف المعياري المرتفع، يمكن القول بوجود قيم شاذة قد تكون معاملات غير طبيعية أو دالة على حالات احتيالية، تؤكد لها قيم المتغيرات الأخرى في نفس السطر.

الشكل رقم (1): توزيع المعاملات الاحتيالية في بيانات الدراسة



المصدر: من إعداد الباحثة باستخدام برمجية بايثون وبيانات الدراسة.

يظهر الشكل رقم (1) أن من حوالي إجمالي 5000 معاملة، تم تصنيف 2670 معاملة منها على أنها عادية (77%)، وتم تصنيف 801 معاملة فقط على أنها احتيال (23%). ورغم أن هذا قد يبدو صغيراً، إلا أن كل معاملة احتيال يمكن أن تمثل نفقات كبيرة للغاية، والتي قد تصل مجتمعة إلى مليارات الدولارات من الإيرادات المفقودة كل عام.

- يتم تطبيق خوارزمية الغابة العشوائية الخاضعة للإشراف التصنيفي، (الملحق رقم 1) مع تدريب وتحسين النموذج، بالإضافة إلى تحديد المعايير المناسبة وتفعيل الوظيفة التي تقوم بتعيين الميزات من خلال تنبؤ كل شجرة قرار ثم جمع التنبؤات النهائية

على شكل غابة القرار (أنظر الملحق رقم 2، 3، 4)، كما هو موضح أدناه في المعادلة رقم (1)، إلى المتغيرات المتوقعة. الميزة الرئيسية في هذا النهج يتم بناءً على الملاحظات المستهدفة ثم يدرّب البيانات تحت الإشراف حتى يحقق قيم الدقة الجيدة.

$$h_k(X) : X \rightarrow Y \equiv H(X) = (h_1(X), h_2(X), \dots, h_K(X)) \quad (\text{المعادلة رقم 1})$$

حيث الوظيفة FI_i تحدد الميزات الأكثر أهمية في النموذج كما في المعادلة رقم (2):

$$FI_i = \frac{1}{K} \sum_{k=1}^K \Delta y_i^k \quad (\text{المعادلة رقم 2})$$

وظيفة أهمية الميزات² في الغابة العشوائية تُستخدم لتحديد مدى تأثير كل ميزة على أداء النموذج. فعند بناء الأشجار يتم تقسيم البيانات عند كل عقدة انطلاقاً من ميزة معينة. هذه التقسيمات تهدف إلى تحسين نقاء العقدة (أي تقليل الأخطاء أو تحسين تصنيف البيانات). (أنظر الملحق رقم 3) Δ فارق تمثل مقدار التحسين في نقاء العقدة مثل مؤشر جيني أو مقياس آخر الناتج عن استخدام ميزة معينة. كلما كانت قيمته أكبر، زاد تأثير تلك الميزة على التنبؤات النهائية.

4. نتائج الدراسة وتحليلها:

- عرض نتائج الدراسة

انطلاقاً من النتائج الأولية المتحصل عليها، تبين وجود فارق كبير بين قيم توزيع المتغير المستهدف الأصلي والفعلي، ولغرض تبسيط أكثر للنموذج وتسهيل التحليل وكذا زيادة دقة النموذج يتم إعادة تشكيل الفئات المصنفة بقصد إعادة التوازن للتصنيف باستعمال وظيفة تصحيح التوازن، حيث النتائج النهائية على الجدول رقم (2).

- شرح مبسط لكيفية عمل وظيفة تصحيح التوازن:

لمعالجة عدم توازن الفئات، تم استخدام خاصية `class_weight='balanced'` في مكتبة Scikit-learn، والتي تضبط وزن كل فئة عكسياً لحجمها في العينة. كما تمت تجربة إعادة أخذ العينات باستخدام طريقة Undersampling للفئة الكبرى وذلك من أجل تحقيق توازن أفضل دون فقد البيانات. تستخدم تقنية "سموت" لإنشاء عينات صناعية للفئة شبه النادرة.

الجدول رقم (2): مخرجات نتائج التنبؤ التصنيفي لخوارزمية الغابة العشوائية

	precision	recall	f1-score	support
0	0.81	0.96	0.88	795
1	0.70	0.27	0.39	247
accuracy			0.80	1042
macro avg	0.76	0.62	0.63	1042
weighted avg	0.78	0.80	0.76	1042

المصدر: من إعداد الباحثة باستخدام برمجية بايثون ومعايير الملحق رقم (1).

² ضبط معلمات نموذج الغابة العشوائية:

- يمكن تحسين النموذج عبر ضبط المعلمات (Hyperparameter Tuning) مثل:

- عدد الأشجار (n_estimators)

- أقصى عمق للشجرة (max_depth)

- الحد الأدنى من العينات لكل تقسيم (min_samples_split).

يظهر من الجدول أن 81 % من الحالات التي توقعها النموذج على أنها صحيحة، وكما تعرف بالاسترجاع على 96% من الحالات الفعلية التي تنتمي للفئة الصحيحة. وحسب المقياس f_1 -Score وهو المقياس الذي يجمع بين الدقة والاسترجاع للحصول على متوسط توافقي بينهما. بالنسبة للفئة صحيحة هي: $f_1 = 0.88$ ، وغير صحيحة هي:

$f_1 = 0.39$ مما يلخص أن الدقة الاحتمالية للنموذج تصل إلى حدود: $f_1 = 0.8$ ما يعني أن النموذج عموماً يعمل بشكل جيد لكنه يعاني من انخفاض في قيم الاستدعاء و f_1 -Score، يتم تحسين معلمات النموذج، حيث إحدى التقنيات التي يمكن استخدامها للعثور على أفضل قيم لهذه المعلمات هي **Grid Search CV** تقدم قائمة بالقيم المحتملة والنتيجة المستخدمة لقياس كفاءة النموذج (أنظر الملحق رقم 2)، وسيقوم بتشغيل التحقق المتبادل مع جميع التركيبات الممكنة وفي النهاية نتوصل إلى التركيبة التي تلخص أفضل نتيجة ممكنة، وبذلك ترتفع دقة النموذج إلى حوالي 91%.

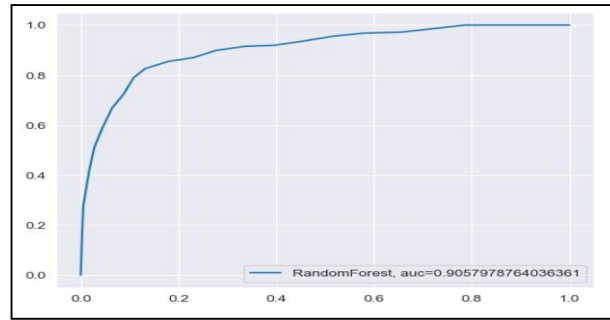
ملاحظة: Grid Search CV تعني البحث الشبكي مع التحقق المتقاطع. وهي تقنية تُستخدم في تعلم الآلة لاختيار أفضل مجموعة من المعلمات (**Parameters**) لنموذج معين. يتم ذلك عبر:

- البحث الشبكي (**Grid Search**): تجربة كل التركيبات الممكنة من المعلمات المحددة (مثل عدد الأشجار أو عمقها).

- التحقق المتقاطع (**Cross-Validation**): تقسيم البيانات إلى أجزاء، وتقييم النموذج على أكثر من جزء لضمان دقة التقييم وعدم الاعتماد على تقسيم واحد فقط.

يمكن التأكد من دقة القياس باستعمال منحنى **AUC** الظاهر على الشكل رقم (3).

الشكل رقم (2): توزيع المعاملات الاحتمالية في بيانات الدراسة



المصدر: من إعداد الباحثة باستخدام برمجية بايثون.

حسب المساحة تحت المنحنى. وكلما زادت قيمة **AUC**، كان النموذج أفضل في التنبؤ بالصفر كصفر والواحد كواحد. في هذه الحالة، كلما زادت **AUC**، كان النموذج أفضل في التمييز بين المعاملات الاحتمالية والعادية. تتراوح قيمة **AUC** من 0.0 إلى 1.0، تصل قيمتها في النموذج المحسن إلى 0.905، مما يؤكد دقة القياس، وقبول نتائج الدراسة للتنبؤ بالحالات الاحتمالية في بطاقات الائتمان قيد بيانات الدراسة.

- تحليل وعرض الميزات الأكثر أهمية في تحديد التعاملات الاحتمالية.

الشكل رقم (3): تحديد الميزات الأكثر أهمية في نموذج الدراسة



المصدر: من إعداد الباحثة باستخدام برمجية بايثون وتفعيل وظيفة تحديد الميزات.

يعرض الرسم البياني في الشكل رقم (3) أهمية المتغيرات في نموذج الغابة العشوائية لكشف الاحتيال في بطاقات الائتمان، حيث:

- متغير المبلغ المسحوب: هو الأكثر نسبة تأثيراً في المتغير المستهدف حيث يفوق 25 بالمائة، يشير إلى أن قيمة المعاملة تلعب دوراً كبيراً في تحديد ما إذا كانت المعاملة احتيالية أم لا. فمن العادة تكون عمليات الاحتيال بمبالغ غير اعتيادية أو كبيرة مقارنةً بسلوك المستخدم العادي.
- متغير المسافة من المنزل: يأتي بنسبة أقل قليلاً من المبلغ المسحوب، يوضح مدى بعد موقع المعاملة عن الموقع المعتاد للعميل، كإشارة إلى عمليات الاحتيال غالباً تتم في أماكن غير مألوفة للمستخدم.
- متغير الزمن أو الوقت: على شكل سلسلة زمنية، محدد بالساعات يأتي في المرتبة الثالثة بنسبة أقل من 20 بالمائة، توقيت ضبط المعاملة قد يكشف عن سلوك غير اعتيادي، مثل تلك العمليات المشبوهة التي تتم تكرارها في أزمدة متفاوتة.
- متغير عدد المعاملات في آخر 24 ساعة: ونسبته أقل من 15 بالمائة، إذا قام المستخدم بعدد كبير من المعاملات خلال فترة قصيرة، فقد يشير ذلك إلى سلوك احتيالي.
- متغير الموقع الجغرافي: بنسبة تأثير أقل من 10 بالمائة له تأثير متوسط، يمكن أن يكون دليلاً على عمليات غير اعتيادية، خاصة إذا كانت المعاملة تتم في بلد آخر.
- متغير فترة إجراء المعاملة في اليوم: بنسبة ضعيفة لكنها تعتبر مؤثرة في النموذج حيث بعض فترات اليوم قد يكون فيها الاحتيال أكثر شيوعاً، مثل المعاملات الليلية المشبوهة، أو فترات الغذاء وخارج المداومة من العمل.
- متغير نوع المعاملة: له تأثير أقل نسبياً مثل عمليات السحب النقدي أو التحويلات، التي قد تكون أقل أهمية مقارنة بالعوامل الأخرى.
- متغير نوع الجهاز: هو الأقل أهمية بنسبة أقل من 5 بالمائة، على الرغم من أنه قد يكون مفيداً في بعض الحالات، إلا أنه لا يعتبر كعامل أساسي في كشف الاحتيال.

- تحليل تأثير كل متغير على النتيجة النهائية

الجدول رقم (3): تحليل ومستويات التأثير للمتغيرات الدراسة على المتغير المستهدف

المتغير	مستوى التأثير	التحليل
قيمة المعاملة	مرتفع جداً	المعاملات ذات المبالغ المرتفعة أو غير المعتادة مقارنة بسلوك العميل المعتاد تعتبر مؤشراً قوياً على احتمال الاحتيال.
المسافة من المنزل	مرتفع	كلما زادت المسافة بين مكان تنفيذ المعاملة والموقع الجغرافي المعروف للعميل، زادت احتمالية أن تكون المعاملة غير شرعية.
وقت المعاملة	مرتفع	الوقت كمتغير زمني يساعد في كشف الأنماط المشبوهة، خاصة إذا كانت المعاملة في ساعات غير معتادة منتصف الليل.
عدد المعاملات في آخر 24 ساعة	متوسط إلى مرتفع	وجود عدد كبير من المعاملات خلال فترة قصيرة قد يدل على هجوم أو استغلال مكثف للبطاقة.
الموقع	متوسط	المعاملات التي تتم من أماكن غير مألوفة أو بلدان غير معتادة للعميل قد تكون مؤشراً على الاحتيال.
توقيت المعاملة في اليوم	متوسط	الفترات الليلية أو غير المعتادة خلال اليوم غالباً ما ترتبط بسلوك احتيالي، لكونها أقل مراقبة.
نوع المعاملة	منخفض إلى متوسط	أنواع معينة مثل التحويلات الفورية أو السحب النقدي يمكن أن تكون أكثر عرضة لسوء الاستخدام مقارنة بعمليات الدفع العادية.
نوع الجهاز	منخفض	استخدام جهاز غير مألوف مثلاً لأول مرة أو من جهاز جديد قد يرفع من احتمال الشك في المعاملة.
المتغير المستهدف	هو الناتج المتوقع	يتم التنبؤ به بناءً على العوامل أعلاه. لا يؤثر بل يتأثر بها.

المصدر: من إعداد الباحثة باستخدام نتائج وظيفة تحديد الميزات.

- تحليل نتائج Grid Search CV: كما هو مبين على الجدول التالي

- الجدول رقم (4): عرض نتائج التحليل الشبكي مع تحقق المقاطع

max_depth	n_estimators	Accuracy	F1-Score	AUC
5	100	0.87	0.78	0.89
7	150	0.89	0.80	0.90
10	200	0.91	0.82	0.905
12	250	0.91	0.82	0.907

المصدر: من إعداد الباحثة باستخدام النتائج المخرجة.

أظهر تحليل Grid Search CV أن أفضل أداء للنموذج تحقق عند القيمة $\text{max_depth} = 10$ و $\text{n_estimators} = 200$ ، حيث بلغت الدقة 91% و F1-Score وصل إلى 0.82، مما يعكس توازناً جيداً بين الدقة والاسترجاع. القيم الأصغر مثل (5, 100) كانت أقل دقة، بسبب ضعف عمق الأشجار وعدم قدرتها على تمثيل التعقيد. في المقابل، القيم الأعلى مثل (12, 250) لم تقدم تحسناً كبيراً، مما يشير إلى نقطة تشبع في الأداء. بناءً عليه، تم اعتماد القيم المتوسطة كخيار أمثل يجمع بين الكفاءة والدقة.

- عرض نتائج مصفوفة الالتباس Confusion Matrix

تستخدم مصفوفة الالتباس لتقييم أداء نماذج التصنيف في تعلم الآلة، وتظهر كيف صنف النموذج البيانات بشكل صحيح أو خاطئ. حيث من خلال النتائج المذكورة حول عدد المعاملات الذي يقدر بـ 3471، وكانت نسبة الاحتيال بـ 23%، وما تم استخراجه من نتائج مصفوفة الالتباس:

- **True Positives (TP) = 710** تعبر عن حالة احتيال تم اكتشافها بنجاح.
 - **False Positives (FP) = 90** تعبر عن 90 معاملة عادية صنفها النموذج كخطأ احتيالي.
 - **False Negatives (FN) = 91** تعبر عن معاملة احتيالية لم يتم اكتشافها.
 - **True Negatives (TN) = 2580** تعبر عن معاملة عادية صنفها بشكل صحيح.
- يتم تأكيد الدقة النهائية للنموذج التي وصلت إلى حوالي: 91%.

- تخريج النموذج على منصة تفاعلية على واجهة محلية ستريمليت Strimlit، حيث:

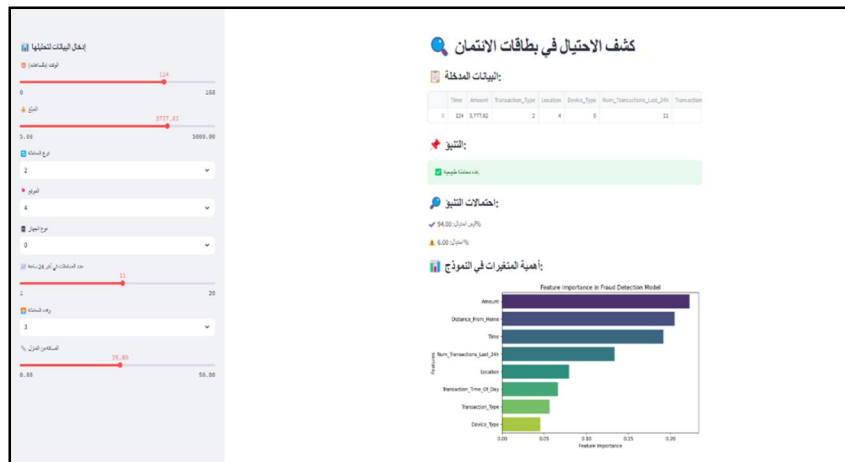
تُستخدم هذه المنصة داخلياً من طرف موظفي المؤسسات المالية أو الجهات الأمنية للتحقق من المعاملات المشبوهة في الوقت الحقيقي، وليست موجهة للاستخدام الفردي من قبل العميل. يتم قراءة البيانات المدخلة، ثم يحول القيم النصية إلى أرقام باستخدام التشفير ثم يتم توظيف النموذج المحسن ويحفظ باستخدام وظائف خاصة، ثم تصميم الواجهة، بحيث يتيح إدخال بيانات جديدة وتظهر النتيجة فيما إذا كانت المعاملة احتيالية أم لا. كما يتم عرض مخطط أهمية المتغيرات المساهمة في كل تغيير للقيمات المدخلة.

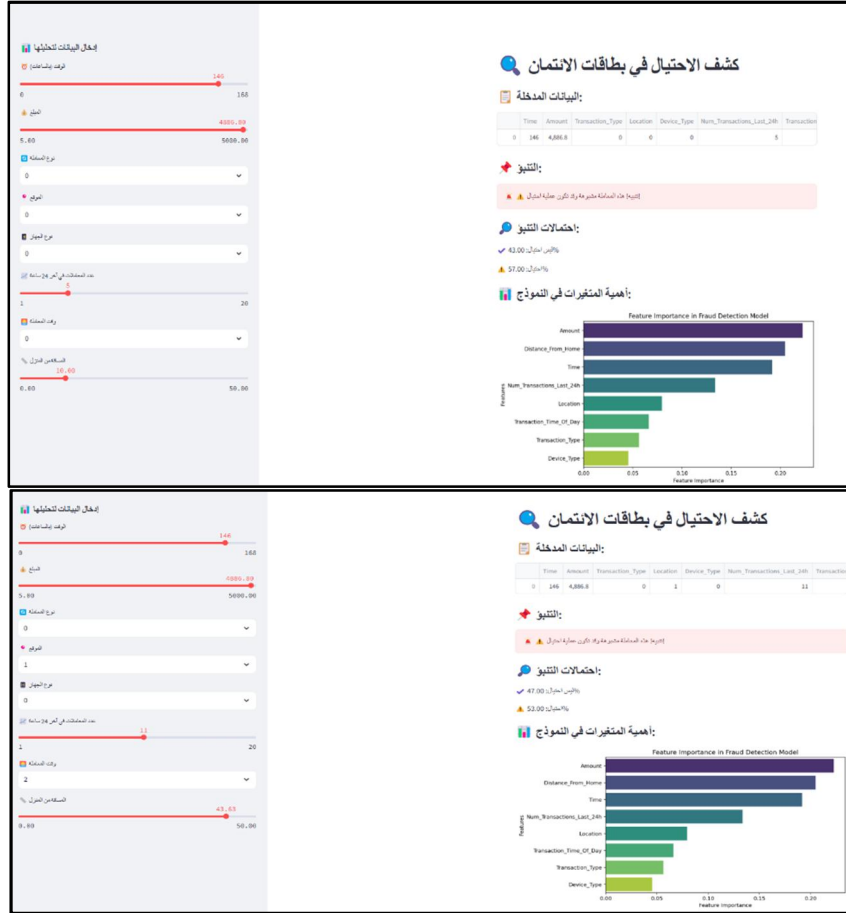
- كيف يتم تطبيق الأداة عملياً؟

يجب توفر مكتبات بايثون: **Scikit-learn** و **Streamlit**، تستخدم عملياً من قبل جهة مكلفة بالمراقبة الامنية (مثلاً موظف مراقب في بنك من البنوك)، وفق ما يلي:

أولاً يتم فتح الواجهة التفاعلية. ثم إدخال بيانات المعاملة. يتم الضغط على التشغيل ثم يستعرض التطبيق النتيجة (احتمالية أو لا)، مع توضيحات أخرى مثل نسبة الاحتمال.

الشكل رقم (4): صور ملتقطة لواجهة المنصة التفاعلية عند قيم مختلفة للمدخلات





المصدر: من إعداد الباحثة باستخدام برمجية بايثون ومكتبة سترميكت.

يظهر من اختلاف صور الشكل رقم (3) حسب مواضيع المتغيرات مدى تأثير المتغير المستهدف الذي يكشف العمليات الاحتمالية:

- بالمبلغ المسحوب، حيث كلما زاد مبلغ المعاملة، زادت احتمالية كونها احتيالية. ويمكن الحماية بتحديد سقف يومي للمعاملات الكبيرة، وطلب تحقق ثنائي للمعاملات المشبوهة.
- المسافة بين موقع المعاملة والموقع المعتاد للعميل: تعد من المتغيرات المهمة في النماذج الذكية لكشف الاحتيال. عند تنفيذ معاملة في موقع بعيد عن الموقع المعتاد للعميل، فإن ذلك قد يشير إلى احتمال وقوع عملية احتيال، خاصة إذا لم يكن هناك سجل سابق لسفر العميل إلى ذلك الموقع.
- الموقع، حيث المعاملات التي تتم من مواقع غير مألوفة (دولة أخرى) تكون أكثر عرضة للاشتباه. ويمكن الحماية باستخدام تقنيات تتبع الموقع الجغرافي وإشعار العميل عند تسجيل الدخول من مكان جديد.
- وقت المعاملة، غالباً ما تتم عمليات الاحتيال في ساعات متأخرة لتجنب اكتشافها بسرعة. وقد تكون الحماية من خلال تفعيل التنبيهات الفورية عند تنفيذ عمليات خلال ساعات غير معتادة.
- اختبار صحة الفرضيات:

الفرضية الأولى وتنص: يمكن لنموذج الغابة العشوائية تحسين دقة كشف الاحتيال، والتقليل من نسبة المعاملات الاحتمالية قبل تنفيذها. من خلال ما تقدم يمكن تأكيد صحة الفرضية نظرياً وتطبيقياً، فنموذج الغابة العشوائية هو أحد أكثر خوارزميات تعلم

الآلة كفاءة في كشف الاحتيال، حيث يعتمد على تجميع نتائج عدة أشجار قرار لزيادة دقة التنبؤ. كما أنه قادر على تصنيف المعاملات المشبوهة بناءً على أنماط سابقة، مما يساعد في الحد من الاحتيال قبل تنفيذه.

الفرضية الثانية وتنص: يمكن تحديد المتغيرات الأكثر أهمية التي يعتمد عليها نموذج الغابة العشوائية من أجل كشف معاملات الاحتيال المالي. فإن تحليل البيانات وتحسين جودتها، واستخدام وظيفة أهمية الميزات في الغابة العشوائية، يمكن تحديد المتغيرات الأكثر تأثيراً، مثل قيمة وتوقيت المعاملة، الموقع ونوع الجهاز، وفترة التنفيذ، عدد العمليات في فترة قصيرة، المسافة. كل هذه المتغيرات تساعد في تصنيف العمليات كاحتمالية أو غير احتمالية، وهو ما يدعم صحة هذه الفرضية.

- تحليل مقارنة للدراسات السابقة:

رغم تشابه الهدف العام للدراسات السابقة والدراسة الحالية في كشف الاحتيال المالي باستخدام الذكاء الاصطناعي، إلا أن هذه الدراسة تتميز بدمج الجانب التطبيقي العملي من خلال منصة تفاعلية، مما يعزز من قابلية استخدامها في الواقع، ويجعلها أكثر فائدة لصناع القرار في المؤسسات المالية، مقارنة بالدراسات السابقة التي ركزت غالباً على التحليل النظري أو التجريبي فقط داخل بيئة البرمجة.

تتوافق النتائج المتحصل عليها في هذه الدراسة مع ما ذهبت إليه الأدبيات السابقة، خاصة ما ورد في دراسة **Esenogho** وآخرين (2022) حول أهمية معالجة توازن البيانات وفعالية النماذج التجميعية في تحسين دقة الكشف عن الاحتيال. كما تعزز دقة النموذج الحالي (91%) ما توصلت إليه دراسة **Devika** وآخرين بشأن فعالية خوارزميات التصنيف عند اختيار المتغيرات المؤثرة بعناية. غير أن الدراسة الحالية انفردت بتقديم نموذج عملي جاهز للاستخدام في بيئة تفاعلية، وهو ما لم تتطرق إليه الدراسات السابقة، مما يعزز من قيمة نتائجها التطبيقية ويوسع آفاق استخدامها في المؤسسات المالية والأمنية.

5. الخلاصة

تناولت هذه الدراسة دور الذكاء الاصطناعي وخوارزميات تعلم الآلة في كشف الاحتيال في بطاقات الائتمان، حيث تم الاعتماد على نموذج الغابة العشوائية كأحد أكثر الخوارزميات كفاءة في تحليل الأنماط الاحتمالية والتنبؤ بها. أظهرت نتائج النموذج، بعد تحسينه عبر اختيار الميزات الأكثر تأثيراً، دقة تصل إلى 90% في تصنيف العوامل التي يشتبه بها غالباً في عمليات الاحتيال المالية. فمثلاً من بين أهم المتغيرات التي أثرت على كشف الاحتيال في بطاقة الائتمان هي: قيمة المعاملة، وقت تنفيذها، المسافة، نوع الجهاز المستخدم، عدد المعاملات في فترة قصيرة، والموقع الجغرافي. ثم تم نشر هذا النموذج على منصة تفاعلية تتيح للمؤسسات المالية والإدارات الأمنية إدخال البيانات وتحليلها في الوقت الفعلي، مما يعزز سرعة وكفاءة اتخاذ القرار في منع عمليات الاحتيال وتقليل الخسائر المالية التي قد تهدد الاستقرار الاقتصادي للدول.

6. التوصيات

على ضوء ما تقدم يمكن من خلال نتائج الدراسة اقتراح بعض التوصيات كالتالي:

- توسيع استخدام خوارزميات الغابة العشوائية في المؤسسات المالية من أجل الكشف المبكر عن الاحتيال في مجالات أخرى مثل التأمين والقروض.
- تحديث النماذج بشكل دوري لمواكبة تطور أساليب الاحتيال.
- تعزيز الوعي بين المستخدمين حول كيفية حماية بياناتهم المالية.

- تعزيز التعاون بين المؤسسات المالية والأجهزة الأمنية للاستفادة من النماذج الذكية في كشف الجرائم المتصلة بعمليات الاحتيال المالية.

7. الهوامش والإحالات:

¹ Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. **The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature.** *Decision support systems*, 50(3), (2011). 559-569.

² Berkman, T. J., & Karthick, S. **A widespread survey on machine learning techniques and user substantiation methods for credit card fraud detection.** *International Journal of Business Intelligence and Data Mining*, 22(1-2), (2023). 223-247.

³ Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. **A neural network ensemble with feature engineering for improved credit card fraud detection.** *IEEE access*, 10, (2022). 16400-16407.

⁴ Devika, M., Kishan, S. R., Manohar, L. S., & Vijaya, N. **Credit card fraud detection using logistic regression.** In *2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, (2022, December). (pp. 1-6). IEEE.

⁵ د. علي بن الضب. دور الذكاء الاصطناعي وتعلم الآلة في تعزيز اكتشاف الاحتيال على البطاقات الائتمانية، العدد 22، صندوق النقد العربي، 2023، صفحة 8.

⁶ بن خضرة حمزة، و بضياف صالح. دور الذكاء الاصطناعي في الكشف والحد من الاحتيال على البطاقات الائتمانية البنكية على المستوى الدولي مجلة الإبداع، 14(2)، 274-292. (2024). <https://asjp.cerist.dz/en/article/249048>

⁷ موقع الكتروني، بعنوان يخلق الذكاء الاصطناعي فرصاً جديدة لمكافحة الاحتيال، الجريدة الدولية U.S. GAO، الفصل الرابع. 2020. تاريخ التصفح 20 فبراير 2025، https://intosaijournal.org/ar/journal-entry/artificial-intelligence-combat-fraud/?utm_source=chatgpt.com

⁸ عبد الرحمن قابيل، سامي، طارق أحمد حافظ، سماح، أحمد محمد إبراهيم سعده، و إكرام. نموذج مقترح لمراجعة الأداء للتنبؤ بالفساد المالي في شركات قطاع الأعمال العام المقيدة في سوق الأوراق المالية المصرية باستخدام تقنية التنقيب في البيانات. المجلة المصرية للدراسات التجارية، 46(4)، 467-530. (2022).

⁹ Louppe, G. **Understanding Random Forests: From Theory to Practice.** (2015). Retrieved 15 March 2025, from arXiv: <https://arxiv.org/abs/1407.7502>

¹⁰ Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. **Random forest for credit card fraud detection.** In *2018 IEEE 15th international conference on networking, sensing and control (ICNSC)*, IEEE. (pp. 1-6). (2018, March).

¹¹ Testas, A. **Random Forest Classification with Scikit-Learn and PySpark.** *Distributed Machine Learning with PySpark*, 243-258. (2023).

8. الملاحق

الملحق رقم (1): تحديد معايير تطبيق خوارزمية الغابة العشوائية

```
(bootstrap=True, ccp_alpha=0.0, class_weight=None,
 criterion='gini', max_depth=None, max_features='auto',
 max_leaf_nodes=None, max_samples=None,
 min_impurity_decrease=0.0, min_impurity_split=None,
 min_samples_leaf=1, min_samples_split=2,
 min_weight_fraction_leaf=0.0, n_estimators=100,
 n_jobs=None, oob_score=False, random_state=None,
 verbose=0, warm_start=False)
```

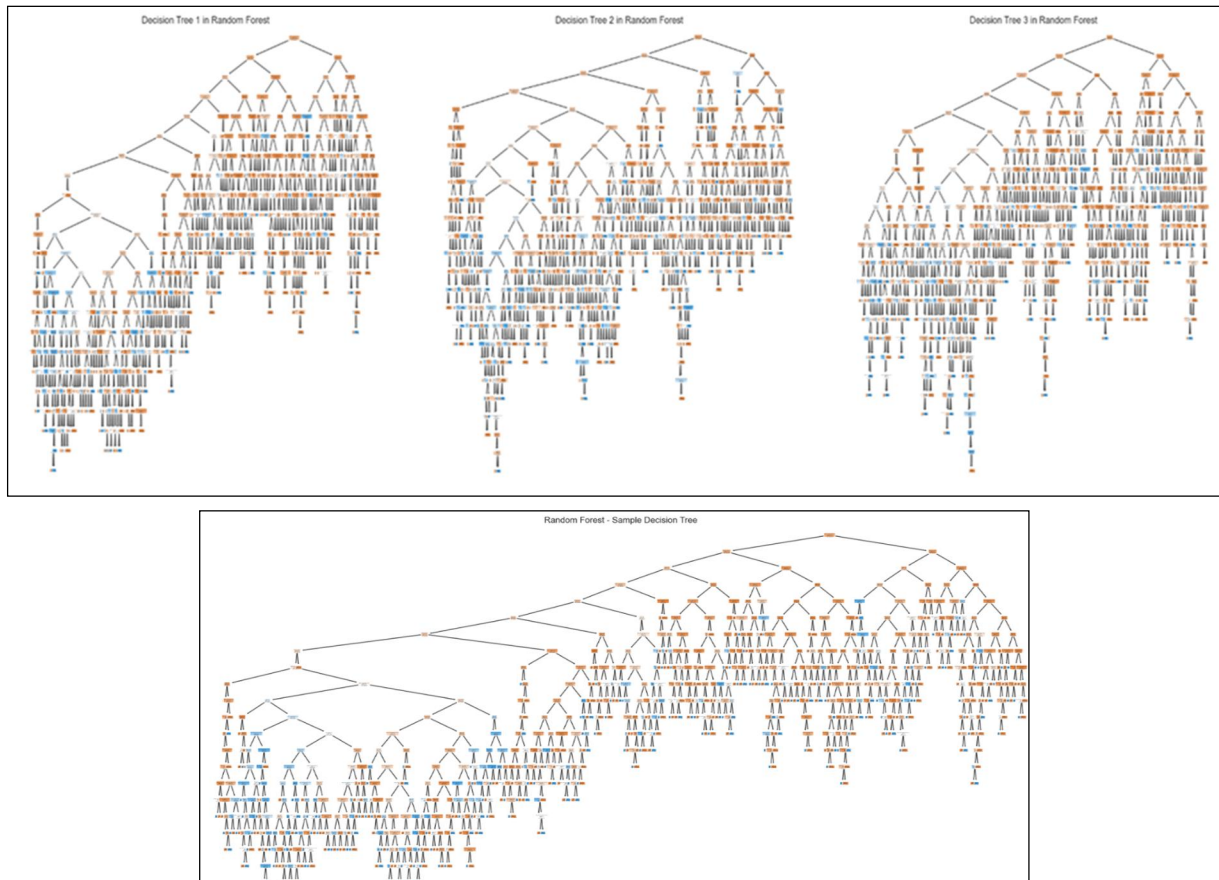
المصدر: من إعداد الباحثة باستخدام برمجية بايثون والحزم المناسبة.

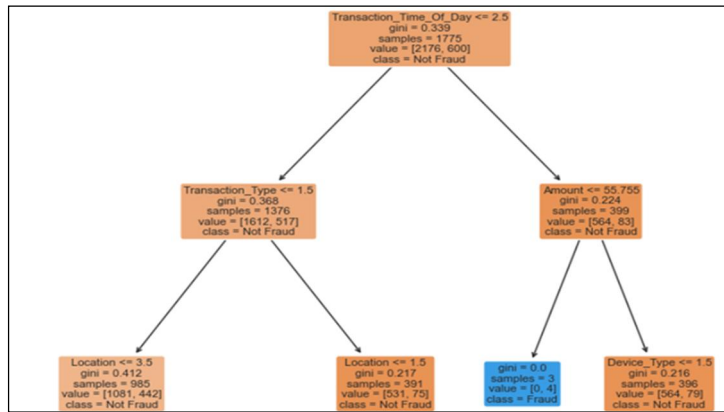
الملحق رقم (2): تحديد معايير تحسين نموذج الغابة العشوائية

```
({'criterion': 'entropy',
 'max_features': 'auto',
 'n_estimators': 25,
 'n_jobs': 1},
 0.9109896480331263)
```

المصدر: من إعداد الباحثة باستخدام برمجية بايثون والحزم المناسبة.

الملحق رقم (3،4): الرسم التوضيحي لفروع وأشجار القرار في الغابة العشوائية





المصدر: من إعداد الباحثة باستخدام برمجية بايثون ونتائج النموذج المحسن.