Comprehensive Journal of Science
Volume (11), Issue (42), (Sapt, 2026)
ISSN: 3104-6266

مجلة العلوم الشاملة
المجلد (11) العدد 42 سبتمبر 2026
ردمد: 3104-6266

إطار هجين قابل للتفسير قائم على تعلم الآلة لكشف مواقع التصيد الاحتيالي باللغة العربية: دمج خصائص الروابط وخصائص المحتوى

وائل عثمان محمود النعاس [1]، ايمان سالم علي مذكور[2]، فؤاد فرج علي الطرشاني [3]، فرج احمد محمود الفارسي [4]
[1.2.3] المعهد العالي للتقنيات الهندسية الماجوري – ليبيا
[4] كلية العلوم والتقنية قمينس

An Explainable Hybrid Machine Learning Framework for Arabic Phishing Website Detection: Integrating URL and Content-Based Features

Wael Othman Mahmoud Al-Naas[1], Iman Salem Ali Al-Madkour[2], Fouad Faraj Ali Al-Tarshani [3] , Faraj Ahmed Mahmoud Al-Farsi [4]
1.2.3 Higher Institute of Engineering Technologies, Al-Majouri – Benghazi, Libya
College of Science and Technology – Qaminis, Libya
1 https://orcid.org/0009-0009-2253-2810 , 2 https://orcid.org/0009-0001-8427-6455
Corresponding Author's Email:F.altrshani@gmail.com

تاريخ الاستلام: 2026/09/02 –تاريخ المراجعة:2026/09/08– تاريخ القبول: 2026/10/06– تاريخ للنشر: 2026/11/17

Abstract

Phishing attacks targeting Arabic-speaking users have increased significantly, exploiting linguistic and cultural characteristics that are often overlooked by detection systems designed primarily for English content. Existing models struggle to generalize to Arabic phishing websites due to morphological complexity, dialectal variation, and deceptive textual patterns. This paper presents an explainable hybrid machine learning framework for phishing website detection in Arabic content, focusing on three dimensions: performance, efficiency, and interpretability. Four models were evaluated: Logistic Regression, Random Forest, XGBoost, and AraBERT. A balanced Arabic dataset comprising legitimate and phishing websites was constructed, incorporating both URL-based features and TF-IDF content features. Experimental results demonstrate that AraBERT achieved the highest accuracy (98.4%), while XGBoost offered the best trade-off between performance and efficiency (97.8% accuracy). Logistic Regression was the fastest and lightest model, but with lower accuracy. Interpretability analysis using SHAP revealed that sensitive Arabic keywords such as "حدث بياناتك" and "تم إيقاف حسابك" were the most influential features in phishing classification. The findings highlight that traditional machine learning models remain highly competitive for Arabic phishing detection when combined with language-aware features, while transformer-based models provide marginal gains at a significantly higher computational cost. This work provides practical guidance for deploying lightweight, explainable, and effective phishing detection systems for Arabic web content.

Keywords: Phishing Detection, Arabic Content, Machine Learning, XGBoost, AraBERT, Interpretability, SHAP.

1. المقدمة

يُعد التصيد الاحتيالي (Phishing) من أكثر أشكال الجرائم الإلكترونية انتشاراً وضرراً، حيث ينتحل المهاجمون صفة جهات موثوقة لخداع المستخدمين والإيقاع بهم للكشف عن معلومات حساسة مثل كلمات المرور والبيانات البنكية والبيانات الشخصية. وعلى الرغم من التقدم الكبير في أنظمة الأمن السيبراني، لا يزال التصيد الاحتيالي فعالاً بشكل كبير لأنه يستغل الثقة البشرية بدلاً من الثغرات التقنية. وتشير التقارير الحديثة إلى أن هجمات التصيد مستمرة في الازدياد من حيث الحجم والتعقيد، مما يسبب خسائر مالية وخسائر في السمعة على مستوى العالم.

وعلى الرغم من أن أغلب الأبحاث ركزت على كشف التصيد في المحتوى الإنجليزي، إلا أن المحتوى العربي لم يحظَ بالاهتمام الكافي. يمثل المستخدمون الناطقون بالعربية شريحة كبيرة ومتنامية من مستخدمي الإنترنت، ومع ذلك يتم استهدافهم بحملات تصيد تستغل الخصائص اللغوية العربية وأنماط الثقة الثقافية. وتعتمد هذه الحملات في الغالب على رسائل ملحة ومقنعة اجتماعياً، مثل الادعاء بالفوز بجوائز، أو إيقاف الحساب، أو الحاجة إلى التحقق الفوري، وذلك للضغط على المستخدم لاتخاذ إجراء سريع.

ويشكل كشف التصيد في المحتوى العربي تحديات فريدة. فاللغة العربية لغة غنية صرفياً، تتميز بتعقيد أشكال الكلمات، وتنوع اللهجات، ومرونة أساليب الكتابة. إذ يمكن أن تظهر الكلمة الواحدة بأشكال متعددة، وغالباً ما تمزج رسائل التصيد بين اللغة العربية الفصحى واللهجات المحلية لتبدو أكثر إقناعاً. بالإضافة إلى ذلك، تعتمد العديد من أنظمة الكشف الحالية على ميزات الروابط أو القوائم السوداء الإنجليزية، وهي غير كافية للتعامل مع مواقع التصيد العربية التي تستخدم نصاً عربياً خادعاً ونطاقات مشابهة ومحتوى مصمماً ثقافياً.

وقد أظهر تعلم الآلة نتائج واعدة في مجال كشف التصيد، لما يوفره من قدرة على تعلم الأنماط من بنية الرابط ومحتوى الصفحة معاً. ومع ذلك، فإن معظم الدراسات السابقة قيمت النماذج بناءً على الدقة فقط، دون الأخذ في الاعتبار عاملين حاسمين للنشر في العالم الحقيقي، وهما: الكفاءة وقابلية التفسير. فالنموذج الدقيق لكنه بطيء أو صعب التفسير قد لا يكون مناسباً للحماية في الوقت الحقيقي أو لمحللي الأمن الذين يحتاجون إلى فهم سبب تصنيف موقع ما على أنه خبيث.

وتعالج هذه الورقة هذه الفجوة من خلال تقديم إطار هجين قابل للتفسير قائم على تعلم الآلة لكشف مواقع التصيد بالمحتوى العربي، مع التركيز على ثلاثة أبعاد: الأداء، والكفاءة، وقابلية التفسير. وتتم مقارنة أربعة نماذج: الانحدار اللوجستي، والغابة العشوائية، و XGBoost، و AraBERT. وتتمثل مساهمات هذه الورقة فيما يلي:

- بناء مجموعة بيانات متوازنة لمواقع التصيد العربية، تجمع بين ميزات الروابط وميزات المحتوى المستخرجة باستخدام TF-IDF.
- تحليل مقارن بين النماذج التقليدية والنماذج القائمة على المحولات من حيث الدقة والتكلفة الحسابية.
- تحليل قابلية التفسير باستخدام SHAP لتحديد الكلمات العربية والميزات الأكثر تأثيراً في تصنيف التصيد.

2. مشكلة البحث

على الرغم من التطور الكبير في أنظمة كشف التصيد الاحتيالي، لا تزال الأنظمة الحالية تواجه قصوراً واضحاً عند التعامل مع المحتوى العربي. فمعظم النماذج المتوفرة تم تصميمها وتدريبها على بيانات إنجليزية، مما يجعلها غير قادرة على فهم الخصائص اللغوية والثقافية التي يستغلها المهاجمون في استهداف المستخدمين الناطقين بالعربية.

وتتمثل المشكلة الرئيسية في أن مواقع التصيد العربية تستخدم أساليب خداع لغوية معقدة، مثل المزج بين اللغة الفصحى واللهجات المحلية، واستخدام كلمات ملحة ومقنعة مثل "حدث بياناتك" و "تم إيقاف حسابك" و "مبروك ربحت"، وهي أساليب لا تستطيع النماذج الإنجليزية اكتشافها. كما تعتمد العديد من

الأنظمة الحالية على ميزات الرابط فقط، متجاهلة المحتوى النصي العربي الذي يمثل المؤشر الأقوى على الخداع.
بالإضافة إلى ذلك، تركز معظم الدراسات السابقة على الدقة فقط، دون مراعاة الكفاءة من حيث سرعة التنفيذ وحجم النموذج، ودون توفير قابلية التفسير التي تسمح لمحللي الأمن بفهم سبب تصنيف موقع ما على أنه تصيد. وهذا يجعل النماذج الحالية غير عملية للنشر في بيئات حقيقية، مثل متصفحات الويب وأنظمة الحماية في الوقت الحقيقي، حيث تكون السرعة والشفافية بنفس أهمية الدقة.

وبناءً على ذلك، يمكن صياغة مشكلة البحث في السؤال الرئيسي التالي:

كيف يمكن بناء نموذج فعال لكشف مواقع التصيد بالمحتوى العربي يحقق توازناً بين الأداء العالي والكفاءة الحسابية وقابلية التفسير؟

ويتفرع من هذا السؤال الأسئلة الفرعية التالية:

1.ما هي الميزات اللغوية وميزات الروابط الأكثر تأثيراً في كشف مواقع التصيد العربية؟
2. أي نموذج من نماذج تعلم الآلة يحقق أفضل أداء في كشف التصيد بالمحتوى العربي؟
3.ما هي التكلفة الحسابية لكل نموذج، وهل تستحق النماذج المعقدة مثل AraBERT[5] هذه التكلفة؟
4.كيف يمكن تفسير قرارات النموذج لتحديد الكلمات العربية التي تدفعه لتصنيف الموقع كتصيد؟

3. أهداف البحث

يهدف هذا البحث إلى بناء وتقييم نموذج فعال لكشف مواقع التصيد الاحتيالي بالمحتوى العربي، يحقق التوازن بين الأداء العالي والكفاءة الحسابية وقابلية التفسير، بما يسمح بنشره في بيئات حقيقية مثل متصفحات الويب وأنظمة الحماية في الوقت الحقيقي.

1. بناء مجموعة بيانات متوازنة لمواقع التصيد العربية، تجمع بين ميزات الروابط (URL Features) وميزات المحتوى النصي المستخرجة باستخدام TF-IDF.
2. تحديد الميزات اللغوية وميزات الروابط الأكثر تأثيراً في الكشف عن مواقع التصيد بالمحتوى العربي.
3. مقارنة أداء أربعة نماذج من تعلم الآلة: الانحدار اللوجستي، والغابة العشوائية، و XGBoost، و AraBERT ، من حيث الدقة (Accuracy) والدقة (Precision) والاستدعاء (Recall) و F1-Score و AUC .
4. تقييم الكفاءة الحسابية لكل نموذج من حيث زمن التدريب وزمن الاستدلال وحجم النموذج، لتحديد النموذج الأنسب للنشر في الوقت الحقيقي.
5. تفسير قرارات النموذج الأفضل باستخدام تقنية SHAP، لتحديد الكلمات العربية والميزات التي تدفع النموذج لتصنيف الموقع كتصيد، وزيادة شفافية النظام وثقة محللي الأمن به.

أهمية البحث

تنبع أهمية هذا البحث من تزايد استهداف المستخدمين الناطقين بالعربية بهجمات التصيد الاحتيالي، في مقابل قلة الأبحاث والأنظمة المتخصصة في كشف التصيد بالمحتوى العربي. وتتمثل أهمية البحث في جانبين: نظري وعملي .

أولاً: الأهمية النظرية

1. يسد هذا البحث فجوة بحثية واضحة في أدبيات كشف التصيد، التي ركزت بشكل كبير على المحتوى الإنجليزي وأهملت المحتوى العربي، رغم خصوصيته اللغوية والثقافية.
2. يقدم البحث إطاراً هجيناً يقارن بين النماذج التقليدية والنماذج القائمة على المحولات مثل AraBERT، مما يساعد الباحثين على فهم متى تستحق النماذج المعقدة تكلفتها الحسابية، ومتى تكون النماذج البسيطة كافية.
3. يسلط البحث الضوء على أهمية دمج ميزات الروابط مع ميزات المحتوى النصي، بدلاً من الاعتماد على نوع واحد من الميزات، مما يثري الإطار النظري لتصميم أنظمة كشف التصيد.

4. يعزز البحث أهمية قابلية التفسير في نماذج تعلم الآلة، من خلال استخدام SHAP لتفسير قرارات النموذج، وهو جانب غالباً ما يتم تجاهله في الدراسات السابقة.

ثانياً: الأهمية العملية

1. **حماية المستخدمين العرب**: يساعد النموذج المقترح على حماية المستخدمين الناطقين بالعربية من سرقة بياناتهم البنكية وكلمات المرور، من خلال الكشف المبكر عن مواقع التصيد.
2. **دعم المؤسسات والبنوك**: يمكن للمؤسسات المالية والبنوك العربية الاستفادة من النموذج في حماية عملائها، وتقليل الخسائر المالية الناتجة عن هجمات التصيد.
3. **النشر في الوقت الحقيقي**: من خلال تقييم الكفاءة الحسابية، يحدد البحث النموذج الأنسب للنشر في متصفحات الويب وأنظمة الحماية، حيث تكون السرعة وحجم النموذج بنفس أهمية الدقة.
4. **زيادة الثقة في أنظمة الذكاء الاصطناعي**: من خلال تفسير قرارات النموذج باستخدام SHAP، يستطيع محللو الأمن فهم سبب تصنيف موقع ما على أنه تصيد، مما يزيد من ثقتهم في النظام ويسهل عليهم اتخاذ القرار الصحيح.

4. فرضيات البحث

الفرضية الرئيسية:

يوجد نموذج تعلم آلة قادر على كشف مواقع التصيد بالمحتوى العربي بدقة عالية، مع تحقيق توازن بين الأداء والكفاءة وقابلية التفسير.

الفرضيات الفرعية:

H1: يؤدي دمج ميزات الروابط (URL Features) مع ميزات المحتوى النصي (TF-IDF) إلى تحسين أداء الكشف عن مواقع التصيد العربية مقارنة باستخدام ميزات الروابط فقط.

H2: يتفوق نموذج XGBoost على نماذج الانحدار اللوجستي والغابة العشوائية من حيث الدقة و F1-Score في كشف التصيد بالمحتوى العربي.

H3: يحقق نموذج AraBERT أعلى دقة من بين جميع النماذج، لكن بتكلفة حسابية أعلى بكثير من حيث زمن التدريب والاستدلال، مما يجعله أقل ملاءمة للنشر في الوقت الحقيقي مقارنة بـ XGBoost.

H4: يمكن لتقنية SHAP تفسير قرارات النموذج بشكل فعال، وتحديد الكلمات العربية الملحة مثل "حدث بياناتك" و "تم إيقاف حسابك" كأكثر الميزات تأثيراً في تصنيف مواقع التصيد.

5 منهجية البحث

اعتمد هذا البحث المنهج التجريبي، من خلال بناء وتقييم أربعة نماذج من تعلم الآلة لكشف مواقع التصيد بالمحتوى العربي، ومقارنة أدائها وكفاءتها وقابليتها للتفسير. تتكون المنهجية المقترحة من ست مراحل رئيسية :

5.1 نظرة عامة على المنهجية المقترحة

تتكون المنهجية المقترحة من ست مراحل رئيسية:

1. جمع مجموعة البيانات
2. المعالجة المسبقة للنص العربي
3. استخراج الميزات
4. تدريب النماذج
5. تقييم الأداء والكفاءة
6. تفسير قرارات النموذج باستخدام SHAP

5.2 مجموعة البيانات Dataset

تم بناء مجموعة بيانات متوازنة تحتوي على مواقع شرعية ومواقع تصيد بالمحتوى العربي.

- مصادر المواقع الشرعية: تم جمعها من مواقع عربية موثوقة مثل المواقع الحكومية والبنوك والمواقع الإخبارية.

- مصادر مواقع التصيد: تم جمعها من قواعد بيانات عامة مثل PhishTank [10] و OpenPhish [11] مع التركيز على المواقع التي تستهدف المستخدمين العرب.
- حجم مجموعة البيانات: تحتوي مجموعة البيانات على 2000 موقع، مقسمة بالتساوي: 1000 موقع شرعي و 1000 موقع تصيد.
- تقسيم البيانات: تم تقسيم البيانات إلى 80% للتدريب و 20% للاختبار، مع استخدام التحقق المتقاطع Cross-Validation بـ 5 طيات لضمان موثوقية النتائج.

5.3 المعالجة المسبقة للنص العربي Preprocessing

يتميز النص العربي بخصائص خاصة تتطلب معالجة دقيقة، وتم تطبيق الخطوات التالية:

1. إزالة علامات HTML والوسوم والروابط.
2. إزالة التشكيل وعلامات الترقيم والأرقام غير الضرورية من المحتوى النصي فقط، مع الاحتفاظ بها في ميزات الرابط.
3. توحيد أشكال الحروف العربية، مثل توحيد (أ، إ، آ) إلى (ا)، و (ة) إلى (ه).
4. إزالة كلمات التوقف العربية مثل (في، من، على، أن).
5. تطبيق التجذير باستخدام مكتبة CAMeL Tools للوصول إلى الجذر اللغوي للكلمة

5.4 استخراج الميزات FeatureExtraction

تم استخراج نوعين من الميزات:

أولاً: ميزات الرابط URL Features (8 ميزات)

1. طول الرابط
2. عدد النقاط
3. وجود عنوان IP
4. وجود رمز @
5. وجود https
6. عدد الأرقام
7. وجود كلمة login
8. Entropy

Listing 1: مثال على استخراج ميزات الرابط

```
import re

def extract_features(url):
    """Extract URL-based features for phishing detection"""
    return {
        'url_length': len(url),
        'num_dots': url.count('.'),
        'has_ip': 1 if re.search(r'\d+\.\d+', url) else 0,
        'has_at': 1 if '@' in url else 0,
        'has_https': 1 if 'https' in url else 0,
        'num_digits': sum(c.isdigit() for c in url),
        'has_login': 1 if 'login' in url else 0,
        'url_entropy': len(set(url)) / len(url)
    }
```

تم استخراج الميزات من الرابط فقط دون تحميل محتوى الصفحة، مما يجعل النموذج سريعاً ومناسباً للكشف الفوري. على سبيل المثال، الرابط الكامل https://support-charge-backs324354.com/index.php يعطي طول 48 و 6 أرقام و Entropy = 0.54، مما يساعد النموذج على تصنيفه كموقع تصيد.
مثال تطبيقي:
عند إدخال الرابط https://support-charge-backs324354.com/index.php ، كانت النتيجة:

```
{'url_length': 48, 'num_dots': 2, 'has_ip': 0, 'has_at': 0,
 'has_https': 1, 'num_digits': 6, 'has_login': 0, 'url_entropy': 0.54}
```

الرابط كموقع تصيد.
ثانياً: ميزات المحتوى النصي (TF-IDF Features)
تم تحويل المحتوى النصي للمواقع إلى أرقام باستخدام تقنية TF-IDF، لقياس أهمية كل كلمة في الموقع. حيث تعطي هذه التقنية وزناً أعلى للكلمات المميزة التي تتكرر في مواقع التصيد مثل "حدث بياناتك" و "تم إيقاف حسابك"، ووزناً أقل للكلمات الشائعة..

5.5 النماذج المقترحة Models

تم تدريب وتقييم أربعة نماذج:
1. الانحدار اللوجستي Logistic Regression:
نموذج خطي بسيط يستخدم كخط أساس Baseline للمقارنة.
2. الغابة العشوائية Random Forest:
نموذج جماعي يعتمد على 100 شجرة قرار، قادر على التعامل مع الميزات غير الخطية.
3. XGBoost [7]:
نموذج تعزيز تدرجي يحقق توازناً ممتازاً بين الدقة والسرعة، ويعتبر النموذج الرئيسي المقترح في هذا البحث.
4. AraBERT [5]:
يعتمد نموذج AraBERT على معمارية BERT [9]، وهو نموذج محولات Transformer تم تدريبه مسبقاً على مجموعة كبيرة من النصوص العربية، مما يسمح له بفهم سياق الكلمات العربية بشكل أفضل من النماذج التقليدية. وتم ضبطه الدقيق Fine-tuning على مجموعة بيانات التصيد العربية.

5.6 مقاييس التقييم Evaluation Metrics

تم تقييم النماذج باستخدام المقاييس التالية:
مقاييس الأداء:
- الدقة الكلية Accuracy
- الدقة Precision
- الاستدعاء Recall
- F1-Score
- AUC-ROC

مقاييس الكفاءة الحسابية:
- زمن التدريب Training Time
- زمن الاستدلال Inference Time لكل عينة
- حجم النموذج Model Size

تم استخدام اختبار t المزدوج (Paired t-test) على نتائج الطيات الخمس للتحقق من الدلالة الإحصائية للفروق بين النماذج.

5.7 تفسير النموذج باستخدام SHAP [8]

تم استخدام تقنية SHAP لتفسير قرارات النموذج الأفضل، وذلك من خلال:

1. تحديد أهم 20 ميزة مؤثرة في تصنيف مواقع التصيد.
2. تحديد الكلمات العربية الأكثر تأثيراً، مثل "حدث"، "مبروك"، "ربحت"، "إيقاف".
3. تحليل أمثلة فردية لتوضيح لماذا تم تصنيف موقع معين كتصيد، مما يساعد محللي الأمن على فهم القرار.

5.8 البيئة التجريبية

تم تنفيذ جميع التجارب باستخدام لغة Python، ومكتبات Scikit-learn و XGBoost و Transformers، على جهاز بمعالج Intel Core i7 وذاكرة GB16 وكرت شاشة GPU.

6 الدراسات السابقة Related Work

تم تقسيم الدراسات السابقة إلى ثلاثة محاور رئيسية: الدراسات المعتمدة على ميزات الرابط، والدراسات المعتمدة على المحتوى النصي، والدراسات المعتمدة على نماذج اللغة العربية.

6.1 الدراسات المعتمدة على ميزات الرابط URL-Based Studies

ركزت العديد من الدراسات المبكرة على استخراج الميزات من الرابط فقط، دون تحليل محتوى الصفحة. أجرى Sahingoz وآخرون [1] دراسة لكشف التصيد باستخدام ميزات الرابط فقط، وحقق نموذج الغابة العشوائية دقة 97.98% على مجموعة بيانات إنجليزية. ورغم الدقة العالية، إلا أن الدراسة تجاهلت المحتوى النصي تماماً، مما يجعلها غير قادرة على كشف مواقع التصيد التي تستخدم روابط تبدو شرعية.وفي دراسة أخرى، قام Mohammad وآخرون [2] ببناء مجموعة ميزات من الرابط وخصائص الصفحة ، وحققوا دقة 95.8% باستخدام نموذج شجرة القرار. لكن الدراسة اعتمدت على ميزات يدوية تحتاج إلى تحديث مستمر مع تطور أساليب المهاجمين.

نقطة الضعف المشتركة : هذه الدراسات لا تستطيع كشف التصيد العربي، لأنها تعتمد على أنماط الروابط الإنجليزية فقط، وتتجاهل الكلمات العربية الملحة التي يستخدمها المهاجمون.

6.2 الدراسات المعتمدة على المحتوى النصي Content-Based Studies

مع تطور هجمات التصيد، بدأ الباحثون في تحليل المحتوى النصي للصفحة.استخدم Rao و Pais [3] تقنية TF-IDF مع نموذج الغابة العشوائية، وحققوا دقة 98.1% على بيانات إنجليزية. وأثبتت الدراسة أن المحتوى النصي يحتوي على مؤشرات خداع أقوى من الرابط وحده.كما استخدم Jain و Gupta [4] تحليل المحتوى النصي والروابط معاً، وحققوا تحسناً ملحوظاً مقارنة باستخدام نوع واحد من الميزات. وأكدت الدراسة أن دمج الميزات هو الأفضل.نقطة الضعف: كل هذه الدراسات كانت على المحتوى الإنجليزي، ولم يتم اختبارها على اللغة العربية التي تتميز بتعقيد صرفي، والمزج بين الفصحى واللهجات.

6.3 الدراسات المعتمدة على نماذج اللغة العربية Arabic NLP Studies

في السنوات الأخيرة، بدأ الاهتمام بنماذج اللغة العربية مثل AraBERT. قام Antoun وآخرون [5] بتطوير نموذج AraBERT، وهو نموذج يعتمد على معمارية BERT [9] تم تدريبه على اللغة العربية، وحقق نتائج ممتازة في مهام تصنيف النصوص العربية مثل تحليل المشاعر وكشف الأخبار الكاذبة.

وأظهرت دراسات سابقة مثل Mubarak وآخرون [6] أن كشف الرسائل الاحتيالية العربية يمثل تحدياً بسبب طبيعة اللغة العربية ولهجاتها المختلفة، مما يبرر الحاجة إلى استخدام نماذج متخصصة مثل AraBERT في هذه الدراسة.

6.4 الدراسات المعتمدة على قابلية التفسير Explainability Studies

اهتمت دراسات قليلة جداً بتفسير قرارات نماذج كشف التصيد. استخدم Lundberg وآخرون [8] تقنية SHAP لتفسير قرارات نماذج تعلم الآلة، حيث أثبتت هذه التقنية قدرتها على تحديد الميزات الأكثر تأثيراً في قرار النموذج. لكن أغلب هذه الدراسات طبقت على المحتوى الإنجليزي، ولم يتم تطبيقها على المحتوى العربي.

6.5 جدول المقارنة

يوضح الجدول (1) ملخصاً لأهم الدراسات السابقة في مجال كشف التصيد الاحتيالي، من حيث الميزات المستخدمة، والنموذج، واللغة، والدقة المحققة، وأوجه القصور في كل دراسة.

الدراسة	الميزات	النموذج	اللغة	الدقة	القصور
Sahingoz et al. [1]	URL فقط	Random Forest	إنجليزي	%97.98	تجاهل المحتوى النصي
Rao and Pais [3]	TF-IDF	Random Forest	إنجليزي	%98.1	لا يدعم العربية
Jain and Gupta [4]	URL + Content	Logistic	إنجليزي	%96.4	لا يوجد تفسير
Antoun et al. [5]	AraBERT	BERT	عربي	%95.0	تكلفة عالية، ليس للتصيد
Mubarak et al. [6]	Text	Traditional	عربي	-	تويتر فقط، ليس مواقع

تُظهر الدراسات السابقة اتجاهين واضحين. الاتجاه الأول هو الدراسات التي اعتمدت على اللغة الإنجليزية فقط، مثل Sahingoz وآخرون [1] و Rao و Pais [3]، حيث حققت دقة مرتفعة تجاوزت 97%، لكنها لا يمكن تعميمها على المحتوى العربي بسبب اختلاف بنية اللغة وخصائصها الصرفية. أما الاتجاه الثاني، فهو الدراسات العربية التي استخدمت نماذج اللغة العربية، مثل Antoun وآخرون [5] و Mubarak وآخرون [6]، حيث حققت نتائج جيدة، لكنها ركزت إما على التصنيف العام للنصوص أو على رسائل تويتر، ولم تركز على مواقع التصيد بشكل خاص. كما يلاحظ أن أغلب الدراسات اعتمدت على نوع واحد من الميزات، إما ميزات الرابط URL فقط، أو ميزات المحتوى TF-IDF فقط، ولم تدمج بينهما، بالإضافة إلى غياب تقييم الكفاءة الحسابية وقابلية التفسير في معظمها.

6.6 الفجوة البحثية Research Gap

من خلال تحليل الدراسات السابقة الموضحة في الجدول (1)، يتضح ما يلي:

1. لا توجد دراسة تجمع بين ميزات الرابط والمحتوى العربي مع تقييم الكفاءة وقابلية التفسير معاً.
2. معظم الدراسات العربية ركزت على الدقة فقط، وتجاهلت زمن الاستدلال وحجم النموذج.
3. لا توجد دراسة استخدمت SHAP لتفسير الكلمات العربية في مواقع التصيد.

وبناءً على ذلك، يقترح هذا البحث نموذجاً هجيناً يدمج بين ميزات URL وميزات المحتوى العربي TF-IDF، مع تقييم الكفاءة الحسابية، واستخدام SHAP لتفسير القرارات، وذلك لسد هذه الفجوة.

7 النتائج والمناقشة Results and Discussion

7.1 نتائج مقارنة النماذج

يوضح الجدول (2) نتائج تقييم النماذج الأربعة على مجموعة الاختبار.

جدول (2): مقارنة أداء النماذج

النموذج	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regressior	%92.4	%91.8	%92.1	%91.9	0.952
Random Forest	%95.6	%95.2	%95.8	%95.5	0.978
XGBoost	%97.8	%97.5	%98.1	%97.8	0.992
AraBERT	%98.4	%98.2	%98.5	%98.3	0.996

المناقشة: تفوق AraBERT على جميع النماذج من حيث الدقة، حيث حقق 98.4%، بفارق 0.6% فقط عن XGBoost. يعود هذا التفوق إلى قدرة AraBERT على فهم سياق الكلمات العربية وخصائصها الصرفية، بينما تفوق XGBoost بشكل واضح على النموذجين التقليديين، مما يؤكد صحة الفرضية H2. كما يلاحظ أن الفارق الصغير بين AraBERT و XGBoost يفتح المجال مستقبلاً لدمج دقة AraBERT مع كفاءة XGBoost.

7.2 نتائج اختبار الفرضية H1: تأثير دمج الميزات

يوضح الجدول (3) تأثير دمج ميزات الرابط مع ميزات المحتوى النصي.

جدول (3): تأثير دمج الميزات على نموذج XGBoost

الميزات المستخدمة	Accuracy	F1-Score	AUC
ميزات الرابط فقط	%93.2	%92.8	0.961
ميزات المحتوى فقط TF-IDF	%95.1	%94.9	0.975
الرابط + المحتوى	%97.8	%97.8	0.992

المناقشة: أدى دمج الميزات إلى تحسن بنسبة 4.6% في الدقة مقارنة باستخدام ميزات الرابط فقط، وبنسبة 2.7% مقارنة باستخدام ميزات المحتوى فقط، وهو تحسن ذو دلالة إحصائية ($p < 0.05$). هذا يثبت صحة الفرضية H1، ويؤكد أن المحتوى العربي يحتوي على مؤشرات خداع لا تظهر في الرابط، وأن الجمع بين النوعين يعطي صورة أكمل للموقع.

7.3 نتائج الكفاءة الحسابية ـ اختبار H3

جدول (4): مقارنة الكفاءة الحسابية

النموذج	زمن التدريب	زمن الاستدلال / عينة	حجم النموذج
Logistic Regression	4 ثواني	ms 0.8	MB 2.1
Random Forest	28 ثواني	ms 12.4	MB 48
XGBoost	35 ثواني	ms 3.2	MB 15
AraBERT	1840 ثواني (30 دقيقة)	ms 85.6	MB 543

المناقشة: رغم أن AraBERT هو الأعلى دقة، إلا أنه أبطأ بـ 26 مرة في الاستدلال،

وأكبر بـ 36 مرة في الحجم من XGBoost. هذا يجعله غير عملي للنشر في الوقت الحقيقي، مثل إضافات المتصفح. بينما يحقق XGBoost توازناً مثالياً بين الدقة والسرعة، مما يثبت صحة الفرضية H3.

7.4نتائج التفسير باستخدام SHAP - اختبار H4

يوضح الجدول (5) أهم الميزات المؤثرة في قرار النموذج.

جدول (5): أهم الميزات المؤثرة في تصنيف التصيد

النوع	الميزة المؤثرة
محتوى عربي	حدث بياناتك
محتوى عربي	تم إيقاف حسابك
محتوى عربي	مبروك ربحت
محتوى عربي	اضغط هنا للتأكيد
محتوى عربي	سيتم حذف حسابك
محتوى عربي	أدخل كلمة المرور
محتوى عربي	تحقق من هويتك
محتوى عربي	جائزة مالية
ميزة رابط	وجود عنوان IP
ميزة رابط	طول الرابط > 75 حرف
ميزة رابط	عدد النقاط > 4
ميزة رابط	Entropy عالي

المناقشة: أثبتت SHAP أن النموذج لا يعتمد على قرارات عشوائية، بل يركز على نفس الكلمات التي يستخدمها المهاجمون لاستعجال الضحية وخلق شعور بالخوف والطمع. وهذا يثبت صحة الفرضية H4، ويجعل النموذج قابلاً للثقة من قبل محللي الأمن .

7.5 تحليل الأخطاء Error Analysis

تم تحليل 44 عينة تم تصنيفها بشكل خاطئ من قبل نموذج XGBoost **على كامل مجموعة البيانات خلال التحقق المتقاطع**، وكانت موزعة على النحو التالي:

جدول (6): توزيع أخطاء التصنيف لنموذج XGBoost

نوع الخطأ	العدد	النسبة
مواقع تصيد تحاكي المواقع الحكومية بشكل احترافي جداً، بدون كلمات ملحة	22	%50.0
مواقع شرعية تحتوي على كلمات تسويقية مثل "مبروك" و "عرض خاص"، مما سبب إنذاراً كاذباً	15	%34.1
صفحات فارغة أو تحتوي على صورة فقط بدون نص	7	%15.9
الإجمالي	44	%100

الحل المقترح: إضافة ميزات بصرية (لقطة شاشة) في الأعمال المستقبلية لحل مشكلة الصفحات التي تعتمد على الصور.

7.6 مناقشة عامة

أثبتت النتائج أن النموذج المقترح يحقق 3 أهداف معاً: دقة عالية، كفاءة في الوقت الحقيقي، وقابلية للتفسير. ويعتبر XGBoost الخيار الأنسب للتطبيق العملي، بينما يبقى AraBERT الخيار الأنسب عندما تكون الدقة هي الأولوية القصوى ولا يوجد قيد على الموارد. وتؤكد النتائج مجتمعة صحة الفرضيات الأربع: دمج الميزات يحسن الدقة (H1)، وتفوق نماذج اللغة العربية على النماذج التقليدية (H2)، ويحقق كفاءة مناسبة للوقت الحقيقي (H3)، ويعتمد على ميزات مفسرة وقابلة للثقة (H4).

8 الخاتمة والتوصيات Conclusion and Recommendations

8.1 الخاتمة

هدفت هذه الدراسة إلى بناء نموذج لكشف مواقع التصيد بالمحتوى العربي، يجمع بين الأداء العالي والكفاءة الحسابية وقابلية التفسير. وأظهرت النتائج أن نموذج AraBERT حقق أعلى دقة بلغت 98.4%، بينما حقق نموذج XGBoost دقة قريبة جداً بلغت 97.8%، مع تفوق واضح في الكفاءة، حيث كان أسرع بـ 26 مرة وأصغر بـ 36 مرة في الحجم. كما أثبتت الدراسة أن دمج ميزات الرابط مع ميزات المحتوى النصي العربي أدى إلى تحسن بنسبة 4.6% في الدقة مقارنة باستخدام ميزات الرابط فقط، وأن استخدام تقنية SHAP ساعد في تفسير قرارات النموذج وتحديد الكلمات العربية الأكثر استخداماً في هجمات التصيد، مثل "حدث بياناتك" و "تم إيقاف حسابك". وبناءً على ذلك، يُوصى باستخدام نموذج XGBoost كحل عملي للكشف عن التصيد في الوقت الحقيقي، بينما يمكن استخدام AraBERT في الحالات التي تكون فيها الدقة هي الأولوية القصوى.

8.2 التوصيات

1. تطوير إضافة للمتصفح تعتمد على نموذج XGBoost، لتنبيه المستخدم العربي قبل الدخول إلى أي موقع مشبوه.
2. نشر الوعي لدى المستخدمين العرب حول الكلمات والعبارات الملحة التي يستخدمها المهاجمون، مثل "حدث بياناتك" و "تم إيقاف حسابك".
3. على البنوك والمؤسسات الحكومية العربية اعتماد أنظمة كشف تعتمد على المحتوى العربي، وليس على الأنظمة الإنجليزية فقط.

8.3 الأعمال المستقبلية Future Work.

1. توسيع مجموعة البيانات لتشمل أكثر من 10,000 موقع، وتغطي اللهجات العربية المختلفة مثل اللهجة الليبية والمصرية والخليجية.
2. إضافة الميزات البصرية، مثل تحليل لقطة شاشة الصفحة، للكشف عن مواقع التصيد التي تعتمد على الصور فقط بدون نص.
3. تطوير نموذج هجين يجمع بين سرعة XGBoost ودقة AraBERT.اختبار النموذج على هجمات التصيد في الوقت الحقيقي، وقياس أدائه ضد هجمات جديدة غير موجودة في بيانات التدريب.

References

[1] O. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345-357, 2019.

[2] R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing websites using an automated technique," in Proc. Int. Conf. Internet Technology and Secured Transactions, 2014, pp. 492-497.

[3] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," Neural Computing and Applications, vol. 31, pp. 3851-3872, 2019.

[4] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, pp. 687-700, 2018.
[5] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in Proc. 4th Workshop on Open-Source Arabic Corpora, 2020, pp. 9-15.
[6] H. Mubarak, A. Abdelali, S. Hassan, and K. Darwish, "Spam detection on Arabic Twitter," in Proc. Int. Conf. on Social Informatics, 2020, pp. 237-251.
[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD, 2016, pp. 785-794.
[8] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems, 2017, pp. 4765-4774.
[9] J. Devlin, et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171-4186.
[10] PhishTank, "PhishTank: Join the fight against phishing," [Online]. Available: https://www.phishtank.org .
[11] OpenPhish, "OpenPhish: Phishing Intelligence," [Online]. Available: https://www.openphish.com .