الشريحة
01

كشف ملفات PDF الخبيثة باستخدام Machine Learning

المدة المستهدفة: 90 ثانية

هدف الشرح

تقديم فكرة المشروع، هدفه الأمني، وما الذي ينتجه النظام.

0:00 - 0:30

هذا المشروع يطبق منهج **Data Science Methods for Cybersecurity** لبناء نظام يساعد على اكتشاف ملفات **PDF** الخبيثة. هذه الملفات قد تبدو طبيعية للمستخدم، لكنها قد تحتوي **JavaScript** أو **actions** أو **embedded objects** أو بنية غير سليمة تستغل قارئ الملفات. لذلك نتعامل مع الملف بوصفه مصدر أدلة أمنية، وليس مجرد مستند.

0:30 - 1:00

يستخرج النظام خصائص ثابتة من بنية الملف، ثم يحولها **Feature Engineering** إلى تمثيل مناسب للنماذج. بعد ذلك ينتج **Classifier** احتمالًا لكون الملف **malicious.** القرار لا يعتمد على **Accuracy** وحدها، بل على **Threshold** مضبوط وفق تكلفة **False Positive** و**False Negative**، مع إمكانية **Abstention** عندما تكون الثقة غير كافية.

1:00 - 1:30

القيمة الأساسية هي ربط **Machine Learning** بعملية أمنية كاملة. النتيجة تعرض الأدلة المرصودة، مساهمة الخصائص، ودرجة الخطر، ثم تقترح إجراءً مثل **release** أو **manual review** أو **sandbox** أو **containment.** بهذا يصبح النموذج أداة لدعم قرار المحلل، وليس حكمًا مستقلًا عن السياق الأمني.

الشريحة
02

يبدأ الهدف الأمني من Threat Model واضح

هدف الشرح

شرح الأصل المحمي، التهديد، تأثير **CIA**، ودور الدفاعات المتعددة.

المدة المستهدفة: 90 ثانية

0:00 - 0:30

الأصل الذي نحميه هو مسار تداول المستندات الموثوقة داخل المؤسسة. المستخدمون والأنظمة الطرفية يحتاجون فتح ملفات **PDF**، لكن هذا السلوك الطبيعي قد يتحول إلى نقطة دخول. التهديد هو ملف خبيث صُمم لإخفاء سلوك ضار داخل بنية تبدو مقبولة أو مربكة لأدوات التحليل.

0:30 - 1:00

يرتبط ذلك بمثلث **CIA.** بالنسبة إلى **Confidentiality** قد يؤدي الملف إلى كشف بيانات أو **credentials.** وبالنسبة إلى **Integrity** قد يغيّر إجراءات أو ينفذ تعليمات غير موثوقة. أما **Availability** فقد تتأثر إذا سببت البنية المعقدة استهلاكًا عاليًا للذاكرة أو وقت المعالجة، لذلك نحمي الموارد أيضًا.

1:00 - 1:30

يفترض **Threat Model** أن المهاجم يستطيع تغيير أسماء العناصر، ترتيبها، **filters**، **metadata**، **nesting**، أو تمثيل **cross-reference** مع الحفاظ على السلوك. لذلك لا يكفي **Model** واحد. نستخدم **Parsing** مضبوطًا، خصائص متعددة، **Calibration**، **Abstention**، و**Sandbox** كطبقات متكاملة تقلل أثر أي فشل منفرد.

الشريحة
03

كل ملف PDF يتحول إلى Supervised Learning Observation

هدف الشرح

توضيح وحدة الملاحظة، الفئة الإيجابية، وعقد **Train-Serve.**

المدة المستهدفة: 90 ثانية

0:00 - 0:30

وحدة الملاحظة هي ملف **PDF** واحد. يحصل كل ملف على **Label** يحدد **benign** أو **malicious**، والفئة الإيجابية هي **malicious** لأنها تمثل التهديد الذي نبحث عنه. من المهم أن يكون مصدر **Label** معروفًا وأن تتم إزالة القيم الغامضة، لأن خطأ **Label** ينتقل مباشرة إلى التدريب والتقييم.

0:30 - 1:00

يمر الملف عبر **Bounded Parser** يستخرج **facts** فقط، مثل عدد **objects** و**streams** و**actions**، وجود **JavaScript** أو **embedded files**، مؤشرات **encryption**، اختلافات **xref**، وتحذيرات **parser health.** بعد ذلك يحول **Feature Vectorizer** هذه الحقائق إلى نفس الأعمدة والترتيب والتحويلات المستخدمة في التدريب.

1:00 - 1:30

هذا هو **Train-Serve Contract**: أي معنى أو ترتيب أو **preprocessing** استُخدم أثناء التدريب يجب أن يتكرر نفسه أثناء **Inference.** يخرج **Classifier probability** ثم تطبق **Calibration** و**Threshold policy.** النتيجة النهائية ليست مجرد **Label**؛ بل قرار مع **Probability** و**Explanation** وإمكانية **Abstention** عند عدم صلاحية الإدخال.

الشريحة
04

Pipeline تربط الأدلة بالاستجابة الأمنية

المدة المستهدفة: 90 ثانية

هدف الشرح

شرح **Workflow** الكامل بالترتيب من **Context** حتى **Monitoring.**

0:00 - 0:30

تبدأ **Pipeline** بتحديد **Context**: ما الأصل المحمي، ما الفئة الإيجابية، وما **Threat Model.** ثم تأتي **Data Validation** لفحص **provenance** و**schema** و**missing values** و**duplicates** وسلامة **Labels.** بعد ذلك تقوم **EDA** بدراسة التوزيعات والاختلال والعلاقات وأي علامة يمكن أن تسبب **Leakage.**

0:30 - 1:00

بعد التحقق، تحول **Feature Engineering** بنية **PDF** إلى **counts** و**densities** و**consistency gaps** و**interactions** و**complexity** و**parser-health signals.** يتم تنفيذ **Split** قبل تعلم أي تحويلات. ثم نقارن **baselines** و**tree-based models** و**neural controls** باستخدام نفس **partitions** ونفس **preprocessing** حتى تكون المنافسة عادلة.

1:00 - 1:30

تأتي **Calibration** لتحويل **score** إلى **Probability** أكثر موثوقية، ثم يختار **Threshold** وفق تكلفة الأخطاء وقدرة المحللين. في **Inference** نحفظ الأدلة ونطبق **OOD checks** و**Abstention.** أخيرًا ترتبط النتيجة بعمليات **triage** و**review** و**sandbox** و**containment** و**monitoring**، وتستخدم الحالات المؤكدة فقط في **Retraining** مضبوط.

الشريحة
05

EDA تمنع Leakage قبل بدء Modeling

المدة المستهدفة: 90 ثانية

هدف الشرح

شرح جودة البيانات، **Split-first preprocessing**، والاختلال الطبيعي.

0:00 - 0:30

الهدف من **EDA** ليس رسم **charts** فقط، بل تحديد ما إذا كان التقييم اللاحق جديرًا بالثقة. نفحص **data types** و**ranges** و**sentinel values** و**missingness** ونسب الفئات. كما نراجع **duplicates** والملفات المترابطة وحقول **provenance**، لأن وجود نسخ من الملف نفسه في **Train** و**Test** يعطي نتيجة مرتفعة لكنها مضللة.

0:30 - 1:00

ننفذ **Group-aware split** قبل أي **Imputation** أو **scaling** أو **feature selection.** كل **Transformer** يتعلم من **Train** فقط، ثم يطبق دون إعادة تعلم على **Validation** و**Test.** هذا يمنع تسرب معلومات من المستقبل. كما نحافظ على **Prevalence** الطبيعي بدل تغيير توزيع الفئات بواسطة **SMOTE** أو توليد سجلات صناعية.

1:00 - 1:30

نستخدم **EDA** أيضًا للبحث عن **Shortcut Features** مثل اسم مصدر أو قيمة ثابتة مرتبطة بالـ **Label** دون معنى أمني. نقارن التوزيعات بين **splits** ونراجع **drift** و**missing patterns.** بهذه الخطوات يصبح أداء النموذج انعكاسًا للقدرة على التعميم، لا نتيجة لتسرب أو فصل بيانات غير سليم.

الشريحة
06

Feature Engineering تمثل السلوك البنيوي لا الكلمات فقط

هدف الشرح

ربط عائلات الخصائص بمعنى أمني واضح وقابل للفحص.

المدة المستهدفة: 90 ثانية

0:00 - 0:30

تبدأ الخصائص من **Raw PDF Structure**: **objects** و**streams** و**actions** و**filters** و**xref** وحجم الملف وحالة **Parser.** **Counts** مفيدة، لكنها تتأثر بالحجم، لذلك نضيف **Scale-stable features** مثل **ratios** و**densities.** مثال ذلك نسبة **actions** إلى **objects** أو **JavaScript markers** إلى **streams** بدل الاعتماد على العدد المطلق فقط.

0:30 - 1:00

نضيف **Consistency Gaps** لقياس التعارض بين ما تعلنه البنية وما يستطيع **Parser** استخراجه، و**Interaction Features** لالتقاط اجتماع مؤشرات ضعيفة تصبح خطرة معًا. كما تمثل **Complexity** عدد مستويات **nesting** وتنوع **filters** وعدم انتظام البنية، بينما توضح **Parser-health features** مقدار **uncertainty** أثناء الاستخراج.

1:00 - 1:30

كل **Feature** موثقة في **versioned schema** وتستخدم في **Train** و**Inference** بالطريقة نفسها. نختبر **Abbreviation** و**Ablation** و**Shortcut Audit** و**Train-Inference Parity.** الهدف ليس زيادة عدد الأعمدة، بل إنشاء **Representation** ذات معنى أمني، مستقرة تحت اختلاف حجم الملف، ويمكن تفسير أثرها واختبار فائدتها.

الشريحة
07

Model Selection تتبع طبيعة الأدلة الجدولية

المدة المستهدفة: 90 ثانية

هدف الشرح

مقارنة Baselines وTrees وNeural Models ضمن شروط متساوية.

0:00 - 0:30

نبدأ بـ Logistic Regression كـ transparent baseline. فائدته أنه يختبر إن كانت الإشارة شبه خطية ويعطي مرجعًا بسيطًا. بعد ذلك نختبر Extra Trees وBoosting لأنها مناسبة للبيانات الجدولية وتلتقط thresholds وinteractions وnonlinear relationships دون افتراض شكل خطي.

0:30 - 1:00

نضيف MLP كـ control ولا نفترض أنه الأفضل لمجرد أنه Neural Network. كما نستخدم FT-Transformer كمرشح Neural أكثر ملاءمة للـ tabular data. كل نموذج يتلقى splits متطابقة وpreprocessing متطابق، ثم تخضع Scores إلى Calibration مستقلة كي نستطيع مقارنة جودة Probability وليس الترتيب فقط.

1:00 - 1:30

يتم الاختيار باستخدام Validation data وOperating Objective محدد، مثل تكلفة False Negative مقابل False Positive وقدرة فريق التحليل. نستخدم Confidence Intervals وstability عبر seeds، ونفضّل أبسط نموذج يحقق فائدة موثوقة. التعقيد يجب أن يثبت قيمته بدل أن يكون هدفًا بحد ذاته.

الشريحة
08

النتائج تُفهم من خلال Operational Metrics

هدف الشرح

شرح نتائج 90%+ من خلال **Precision** و**Recall** و**F-scores** و**ROC.**

المدة المستهدفة: 90 ثانية

0:00 - 0:30

أظهرت المقاييس الأساسية في الاختبارات المحفوظة أداءً ضمن نطاق 90%+. لكن الرقم المجمع لا يكفي. **Precision** تجيب: من كل الملفات التي صنفناها **malicious**، كم ملفًا كان خبيثًا فعلًا؟ ارتفاعها يقلل **False Positives** ويحمي وقت المحللين من التنبيهات غير الضرورية.

0:30 - 1:00

Recall تجيب: من كل الملفات الخبيثة الموجودة، كم ملفًا اكتشفنا؟ ارتفاعها يقلل **False Negatives. F1** توازن **Precision** و**Recall** بالتساوي، بينما **F-beta** يسمح بإعطاء وزن أكبر لـ **Recall** عندما يكون فقدان تهديد أخطر من مراجعة إنذار إضافي، أو العكس وفق سياسة التشغيل.

1:00 - 1:30

ROC-AUC يقيس قدرة النموذج على ترتيب **malicious** أعلى من **benign** عبر **Thresholds** متعددة، لكنه لا يحدد **Threshold** التشغيلي. لذلك نضيف **PR-AUC** و**partial ROC** عند **Low FPR** و**Calibration** و**Confusion Matrix** و**Confidence Intervals.** القرار النهائي يربط **Threshold** بالـ **Prevalence** الحقيقي وتكلفة الأخطاء وقدرة **SOC.**

الشريحة
09

Explainability تحول Score إلى تحقيق قابل للتنفيذ

المدة المستهدفة: 90 ثانية

هدف الشرح

الفصل بين Observed Facts وModel Attribution وAnalyst Action.

0:00 - 0:30

الطبقة الأولى هي Observed Facts التي ينتجها Parser: actions وJavaScript markers وembedded content وencryption وstructural gaps وتحذيرات parser health. هذه حقائق عن الملف، لكنها لا تثبت النية وحدها. لذلك تظهر منفصلة عن تفسير النموذج حتى لا نخلط بين الملاحظة والاستنتاج.

0:30 - 1:00

الطبقة الثانية هي Model Attribution. نستخدم Local explanations لبيان الخصائص التي رفعت أو خفضت Probability مقارنة بخلفية Training حقيقية. وعلى المستوى العام نستخدم Permutation Importance وnative importance وALE وinteractions وfeature-family ablation وstability عبر bootstrap وseeds.

1:00 - 1:30

لا نقبل Explanation قبل Sanity Checks مثل label randomization وnoise controls وfaithfulness tests وsubgroup analysis. ثم ترتبط النتيجة بـ Analyst Action: release للحالة الموثوقة، review أو sandbox للحالة المشبوهة، وAbstention عند parser failure أو OOD. Attribution ليست Causality ولا دليل إدانة؛ إنها أداة لترتيب التحقيق.

الشريحة
10

الدفاعات المتعددة ضد Adversarial Attacks

المدة المستهدفة: 90 ثانية

هدف الشرح

شرح أساليب **Evasion** والدفاعات المطبقة والمقترحة.

0:00 - 0:30

في **Adversarial Attack** قد يغير المهاجم أسماء **objects** أو **whitespace** أو **metadata** أو **filters** أو **nesting** أو **xref structure** مع الحفاظ على السلوك الخبيث. نختبر هذه الفكرة باستخدام **Inert Mutations** آمنة لا تنفذ **Payloads** حقيقية، ثم نقارن **Prediction** للملف الأصلي والنسخة المعدلة.

0:30 - 1:00

الدفاع الأول هو **Canonicalization** لتقليل الاختلافات الشكلية. ويأتي **Multi-view Parsing** لاستخراج الأدلة بأكثر من تفسير مستقل. تحمي **Resource Bounds** مثل حدود **size** و**time** و**nesting** من استنزاف **Availability.** كما تكشف **Parser-health checks** و**OOD Detection** الملفات التي لا تشبه ظروف **Training.**

1:00 - 1:30

عندما تقع **Probability** قرب **Threshold** أو ترتفع **Uncertainty**، يستخدم النظام **Abstention** بدل قرار قسري. تشمل الطبقات الإضافية **Monitoring** لـ **drift** وأنماط **evasion**، **Adversarial Training** على حالات آمنة ومتحقق منها، **Sandbox** خارجي، **Retraining** مضبوط، و**Rollback** لإصدار سابق. القوة هنا ناتجة عن اجتماع الدفاعات لا عن تقنية واحدة.

الشريحة
11

ضوابط البيانات والنمذجة تحمي Validity

المدة المستهدفة: 90 ثانية

هدف الشرح

شرح Scale وData Quality وFeature Schema والمقارنة العادلة.

0:00 - 0:30

تحتوي Data Pipeline على ضوابط تجعل العمل قابلًا لإعادة الإنتاج على حجم كبير. يتم التحقق من provenance وchecksum وschema وrow quality وsplit integrity باستخدام Streaming Processing دون تحميل الجدول كاملًا في الذاكرة. كما تمنع Duplicate Control وGroup-aware Split وصول عينات مترابطة إلى أكثر من Partition.

0:30 - 1:00

يعتمد Feature Engineering على Versioned Schema يجمع scale-stable transformations وdensities وconsistency gaps وrisk interactions وcomplexity وparser-health indicators. يختبر Shortcut Audit ما إذا كان النموذج يستغل مؤشرًا غير أمني، بينما يختبر Ablation أهمية كل Feature Family، وتضمن Parity Tests تطابق Training وInference.

1:00 - 1:30

تستخدم جميع النماذج partitions وpreprocessing وCalibration وOperating Objective نفسها. تشمل المقارنة Logistic Regression وExtra Trees وBoosting وMLP control وFT-Transformer. بهذه الشروط تكون المقارنة عادلة، ويمكن تفضيل نموذج أبسط عندما يقدم أداءً واستقرارًا وقيمة تشغيلية مماثلة لنموذج أكثر تعقيدًا.

الشريحة
12

التقييم والتفسير والدفاع تشكل Evidence Chain واحدة

المدة المستهدفة: 90 ثانية

هدف الشرح

تلخيص المشروع من **Metrics** حتى القرار والمراقبة المستمرة.

0:00 - 0:30

يجمع التقييم **Precision** و**Recall** و**F1** و**F-beta** و**ROC-AUC** و**PR-AUC** و**partial ROC** و**Calibration** و**Confidence Intervals** وتحليل الأخطاء. هذه المجموعة تمنع الاعتماد على **Metric** واحدة، وتوضح جودة الفصل وجودة **Probability** والتكلفة المتوقعة عند **Threshold** التشغيلي.

0:30 - 1:00

تربط **Explainability** بين **Feature Evidence** وسلوك **Model** باستخدام **Local Attribution** و**Permutation** و**ALE** و**Interactions** و**Ablation** و**Stability** و**Sanity Checks.** لا يتحول أي نمط إلى **Conclusion** قبل اتفاق أكثر من طريقة. بعدها يمكن اتخاذ قرار مثل **keep** أو **remove** أو **re-engineer** أو **monitor** أو **Abstain.**

1:00 - 1:30

تختبر **Robustness** سلوك النموذج قبل وبعد **Inert Mutations** و**Canonicalization** و**Multi-view Parsing** و**Resource Bounds** و**OOD Abstention.** وفي التشغيل تستمر **Monitoring** ومراجعة المحللين و**Sandbox** و**Controlled Retraining.** الخلاصة أن **Machine Learning** يسرع **Triage**، بينما الثقة تأتي من **Data Quality** و**Evaluation** و**Explainability** و**Layered Defense** معًا.