

Title Slide

Intrusion Detection System Using Machine Learning

This presentation focuses on the development of an intrusion detection system (IDS) utilizing machine learning techniques to enhance anomaly detection in network traffic. The study compares two machine learning models, Isolation Forest and One-Class Support Vector Machine (SVM), and explores the implementation of a user-friendly graphical user interface (GUI) for practical deployment.

Slide 1: Introduction

Intrusion Detection Systems (IDS) play a crucial role in monitoring network traffic to identify and respond to suspicious activities. Traditional IDS relies heavily on predefined rules and signature-based detection, which limits their ability to address advanced and evolving threats such as zero-day attacks. Machine learning (ML) provides a more adaptive and scalable solution for detecting anomalies effectively. By leveraging ML, IDS can continuously learn and adapt to new patterns, enhancing their effectiveness in modern cybersecurity environments.

Slide 2: Objectives

The primary goal of this project is to develop a robust IDS that leverages machine learning for anomaly detection. The study aims to compare the effectiveness of Isolation Forest and One-Class SVM models in detecting anomalies. Additionally, the project focuses on creating a user-friendly GUI to streamline data processing, model training, and visualization, paving the way for scalable and efficient IDS solutions.

Slide 3: Dataset Description

The dataset utilized in this project is NSL-KDD, a cleaned and improved version of the KDD Cup 1999 dataset. This dataset is widely recognized for its balance and reduced redundancy, making it suitable for evaluating IDS systems. It consists of 41 features classified into three categories: Basic, Content, and Traffic features. Basic features describe general network attributes, Content features focus on session-specific activities, and Traffic features capture aggregated behavior over time. The dataset categorizes traffic into normal and malicious classes, with malicious traffic further divided into DoS, Probe, U2R, and R2L attacks.

Slide 4: Methodology

The methodology involves a series of steps to ensure the effective implementation of the IDS. First, the dataset is loaded and preprocessed, which includes cleaning, normalization, and encoding of categorical features. Subsequently, the user selects a model, either Isolation Forest or One-Class SVM, and trains it using the preprocessed data. The trained model is then tested to evaluate its performance, followed by visualizing results such as anomaly detection plots. The final step involves exporting the results for further analysis or reporting.

Slide 5: Tools and Environment

The implementation was performed using Python, supported by essential libraries such as Scikit-learn for model implementation, NumPy for numerical computations, Matplotlib for visualizations, and Tkinter for GUI development. The system was developed on a hardware setup consisting of an Intel Core i7 processor, 16GB RAM, and 512GB SSD, ensuring optimal performance.

Slide 6: Machine Learning Models

The study explores two machine learning models for anomaly detection:

- **Isolation Forest:** Detects anomalies by isolating them in the dataset, making it suitable for high-dimensional data and large-scale environments.
 - **One-Class SVM:** Defines a boundary around normal data points and flags anything outside this boundary as an anomaly. This model is particularly effective in scenarios where normal behavior is well-defined.
-

Slide 7: Graphical User Interface (GUI)

The GUI developed for this project simplifies the use of the IDS by providing a seamless interface for loading datasets, selecting models, and visualizing results. Key features include dataset validation, parameter configuration, and the ability to export results. The GUI ensures accessibility for users with minimal technical expertise, enhancing the practicality of the system.

Slide 8: Implementation Workflow

The implementation follows a structured workflow beginning with data loading and preprocessing. The user then selects a machine learning model, trains it on the dataset, and evaluates its performance using testing data. The results are visualized to provide insights into anomaly detection, and the outputs are exported for further use. This workflow ensures a logical and efficient approach to developing the IDS.

Slide 9: Challenges

The project faced several challenges, including handling imbalanced datasets, optimizing model hyperparameters, and ensuring scalability for real-time applications. Addressing these challenges involved implementing robust preprocessing techniques and thorough parameter tuning to achieve optimal performance.

Slide 10: Future Work

Future work aims to expand the system's capabilities by integrating real-time deployment for dynamic network environments. Additional validation using advanced datasets like CIC-IDS2017 will be conducted to evaluate the system's adaptability. Furthermore, exploring advanced techniques such as deep learning and improving the GUI for enhanced interactivity will be prioritized.

Slide 11: Conclusion

The project demonstrates the potential of machine learning in enhancing IDS effectiveness. Isolation Forest and One-Class SVM provide complementary strengths, offering scalability and precision in anomaly detection. By combining these models with a user-friendly GUI, the system sets a solid foundation for developing scalable and robust IDS solutions.

Slide 12: Questions

Thank you for your attention!

Are there any questions?

الشريحة 1: المقدمة

دورًا حيويًا في مراقبة حركة المرور الشبكية لتحديد الأنشطة المشبوهة والتعامل معها. تعتمد أنظمة (IDS) تلعب أنظمة كشف التسلل كشف التسلل التقليدية بشكل كبير على القواعد المحددة مسبقًا والكشف المعتمد على التوقعات، مما يحد من قدرتها على معالجة التهديدات حلاً أكثر تكيفًا وقابلية للتوسع للكشف عن الشذوذ (ML) يقدم التعلم الآلي. "Zero-Day" المتقدمة والمتطورة مثل الهجمات من نوع بفعالية. من خلال الاستفادة من التعلم الآلي، يمكن لأنظمة كشف التسلل التعلم باستمرار والتكيف مع الأنماط الجديدة، مما يعزز فعاليتها في بيانات الأمن السيبراني الحديثة.

الشريحة 2: الأهداف

الهدف الأساسي من هذا المشروع هو تطوير نظام كشف تسلل قوي يستفيد من التعلم الآلي للكشف عن الشذوذ. تهدف الدراسة إلى مقارنة في الكشف عن الشذوذ. بالإضافة إلى ذلك، يركز المشروع على إنشاء واجهة Isolation Forest وOne-Class SVM فعالية نماذج لتبسيط معالجة البيانات وتدريب النماذج وعرض النتائج، مما يمهّد الطريق لأنظمة كشف تسلل قابلة (GUI) رسومية سهلة الاستخدام للتوسع وفعالة.

الشريحة 3: وصف مجموعة البيانات

KDD Cup 1999 ، وهي نسخة مُحسّنة ومُنقّحة من مجموعة بيانات NSL-KDD مجموعة البيانات المستخدمة في هذا المشروع هي تُعرف هذه المجموعة بتوازنها وتقليل التكرار فيها، مما يجعلها مناسبة لتقييم أنظمة كشف التسلل. تتكون المجموعة من 41 ميزة مصنفة تصف الميزات (Traffic) ، وميزات حركة المرور (Content) ، وميزات المحتوى (Basic) إلى ثلاث فئات: الميزات الأساسية الأساسية سمات الشبكة العامة، وتركز ميزات المحتوى على الأنشطة الخاصة بالجلسات، بينما تلتقط ميزات حركة المرور السلوك الإجمالي عبر الزمن. تُصنف حركة المرور في المجموعة إلى فئات طبيعية وخبيثة، مع تقسيم الحركة الخبيثة إلى هجمات من نوع DoS ، Probe ، U2R ، و R2L.

الشريحة 4: المنهجية

تتضمن المنهجية سلسلة من الخطوات لضمان التنفيذ الفعال لنظام كشف التسلل. أولاً، يتم تحميل مجموعة البيانات ومعالجتها مسبقاً، بما في ذلك One-Isolation Forest أو Isolation Forest في ذلك تنظيف البيانات، وتطبيعها، وترميز الميزات القوية. بعد ذلك، يختار المستخدم نموذجاً، سواء كان Class SVM ، ويتم تدريبه باستخدام البيانات المعالجة مسبقاً. ثم يتم اختبار النموذج المدرب لتقييم أدائه، يتبع ذلك عرض النتائج مثل Class SVM . الرسوم البيانية للكشف عن الشذوذ. تتضمن الخطوة النهائية تصدير النتائج لتحليلها أو الإبلاغ عنها.

الشريحة 5: الأدوات والبيئة

للمعاملات NumPy لتنفيذ النماذج، و Scikit-learn ، بدعم من مكتبات أساسية مثل Python تم تنفيذ المشروع باستخدام لغة البرمجة Intel Core لتطوير الواجهة الرسومية. تم تطوير النظام على جهاز مزود بمعالج Tkinter للرسوم البيانية، و Matplotlib للحسابية، و i716 ، مما يضمن الأداء الأمثل 512GB بسعة SSD ، ووحدة تخزين GB ، وذاكرة وصول عشوائي i716 .

الشريحة 6: نماذج التعلم الآلي

تستكشف الدراسة نموذجين للتعلم الآلي للكشف عن الشذوذ:

- **Isolation Forest:** يقوم بالكشف عن الشذوذ من خلال عزلها في مجموعة البيانات، مما يجعله مناسباً للبيانات ذات الأبعاد العالية والبيئات واسعة النطاق.
- **One-Class SVM:** يقوم بتحديد حدود حول النقاط الطبيعية ويصنف أي شيء خارج هذه الحدود كعنصر شاذ. يُعد هذا النموذج فعالاً بشكل خاص في السيناريوهات التي يكون فيها السلوك الطبيعي محدداً بوضوح.

(GUI) الشريحة 7: الواجهة الرسومية

لهذا المشروع لتبسيط استخدام نظام كشف التسلل من خلال توفير واجهة سلسلة لتحميل مجموعات (GUI) تم تطوير واجهة رسومية للبيانات، واختيار النماذج، وعرض النتائج. تشمل الميزات الرئيسية التحقق من صحة البيانات، تكوين المعلمات، والقدرة على تصدير النتائج. تضمن الواجهة سهولة الوصول للمستخدمين ذوي الخبرة التقنية المحدودة، مما يعزز من عملية النظام.

الشريحة 8: سير عمل التنفيذ

يتبع التنفيذ سير عمل منظم يبدأ بتحميل البيانات ومعالجتها مسبقاً. ثم يختار المستخدم نموذج التعلم الآلي، ويدربه على مجموعة البيانات، ويقيم أدائه باستخدام بيانات الاختبار. يتم عرض النتائج لتوفير رؤية حول الكشف عن الشذوذ، ويتم تصدير المخرجات للاستخدام الإضافي. يتضمن هذا النهج سيراً منطقياً وفعالاً لتطوير نظام كشف التسلل.

الشريحة 9: التحديات

واجه المشروع عدة تحديات، بما في ذلك التعامل مع مجموعات البيانات غير المتوازنة، وتحسين معلمات النماذج، وضمان القابلية للتوسع في التطبيقات الزمنية الحقيقية. تضمنت معالجة هذه التحديات تطبيق تقنيات معالجة مسبقة قوية وضبطاً دقيقاً للمعلمات لتحقيق الأداء الأمثل.

الشريحة 10: الأعمال المستقبلية

تهدف الأعمال المستقبلية إلى توسيع قدرات النظام من خلال دمجها في تطبيقات زمنية حقيقية للبيانات الشبكية الديناميكية. سيتم إجراء تقييم قابلية النظام للتكيف. علاوة على ذلك، سيتم التركيز على CIC-IDS2017 المزيد من التحقق باستخدام مجموعات بيانات متقدمة مثل على استكشاف تقنيات متقدمة مثل التعلم العميق وتحسين واجهة المستخدم الرسومية لتعزيز التفاعلية.

الشريحة 11: الخاتمة

One-Class SVM و Isolation Forest يوضح المشروع إمكانيات التعلم الآلي في تعزيز فعالية أنظمة كشف التسلل. يوفر كل من نقاط قوة مكملة، مما يتيح قابلية التوسع والدقة في الكشف عن الشذوذ. من خلال دمج هذه النماذج مع واجهة رسومية سهلة الاستخدام، يوفر النظام أساسًا قويًا لتطوير حلول أنظمة كشف تسلل قابلة للتوسع والموثوقة.

الشريحة 12: الأسئلة

شكرًا لاهتمامكم

هل هناك أي أسئلة؟

Q:

1. لماذا اخترت موضوع نظام كشف التسلل (IDS)؟

الإجابة:

اخترت هذا الموضوع نظرًا لأهميته في مجال الأمن السيبراني. مع تزايد الهجمات السيبرانية، أصبحت الحاجة إلى أنظمة كشف تسلل فعالة أمرًا ضروريًا لحماية الشبكات. أردت استكشاف كيف يمكن للتعلم الآلي أن يعزز من قدرات هذه الأنظمة ويجعلها أكثر تكيفًا مع التهديدات المتطورة.

2. ولماذا اخترتهما؟ One-Class SVM و Isolation Forest ما الفرق بين

الإجابة:

- **Isolation Forest:** يعمل على عزل النقاط الشاذة بطريقة تعتمد على تقسيم البيانات بشكل عشوائي. مناسب للبيانات ذات الأبعاد العالية والبيئات الكبيرة.
- **One-Class SVM:** يقوم بإنشاء حدود حول السلوك الطبيعي ويصنف أي نقطة خارج هذه الحدود كعنصر شاذ. فعال مع البيانات التي يكون فيها السلوك الطبيعي واضحًا. اخترتهما لأنهما نموذجين غير خاضعين للإشراف ويمكنهما التعامل مع السيناريوهات المختلفة للشبكات.

3. كمجموعة بيانات؟ NSL-KDD لماذا اخترت

الإجابة:

الأصلية. تقلل من التكرار KDD Cup 1999 لأنها مجموعة بيانات متوازنة ومُحسّنة مقارنةً بمجموعة بيانات NSL-KDD اخترت وتحسن التقييم، مما يجعلها مناسبة لاختبار فعالية نماذج التعلم الآلي.

ما هي أبرز التحديات التي واجهتها أثناء 4.*

You said:

اعد الاسئلة مع إجابات أكاديمية وعامية

ChatGPT said:

ChatGPT

أسئلة مقترحة مع إجابات أكاديمية وعامية

1. لماذا اخترت العمل على نظام كشف التسلل (IDS)؟

- **الإجابة الأكاديمية:**
اخترت هذا الموضوع نظراً لأهميته المتزايدة في مجال الأمن السيبراني. مع تزايد التهديدات والهجمات الإلكترونية، أصبحت الحاجة إلى أنظمة كشف تسلل قادرة على التعامل مع التهديدات المتطورة أمراً ضرورياً. أردت استكشاف دور تقنيات التعلم الآلي في تحسين قدرات هذه الأنظمة من خلال الكشف عن الشذوذ بفعالية أكبر.
- **الإجابة العامة:**
لأن الأمن السيبراني شيء مهم جداً في حياتنا اليومية، والهجمات الإلكترونية في تزايد مستمر. أنظمة كشف التسلل هي خط الدفاع الأول، واستخدام التعلم الآلي يمكن أن يجعلها أذكى وأكثر كفاءة.

2. لماذا اخترت خوارزميات Isolation Forest وOne-Class SVM؟

- **الإجابة الأكاديمية:**
لأنه يعمل بكفاءة على البيانات ذات الأبعاد العالية ويتعامل مع الكميات الكبيرة من البيانات Isolation Forest اخترت بقدرته العالية على الكشف عن الشذوذ في البيانات التي يكون فيها السلوك الطبيعي One-Class SVM بسهولة، بينما يتميز محدداً بوضوح. هذان النموذجان يكملان بعضهما البعض مما يوفر تغطية شاملة لأنماط البيانات المختلفة.
- **الإجابة العامة:**
اخترت هذين النموذجين لأنهما فعالان في التعامل مع البيانات بطرق مختلفة. الأول ممتاز في التعامل مع كميات كبيرة، والثاني دقيق في الحالات التي نعرف فيها السلوك الطبيعي بشكل واضح.

3. كمجموعة بيانات؟ NSL-KDD لماذا استخدمت

- **الإجابة الأكاديمية:**
الأصلية، حيث تقلل من التكرار وتوازن البيانات. هذا KDD Cup 1999 هي نسخة محسنة من مجموعة NSL-KDD مجموعة يجعلها أكثر ملاءمة لاختبار فعالية النماذج دون التأثير على التقييم بسبب البيانات المتكررة أو غير المتوازنة.
- **الإجابة العامة:**
اخترتها لأنها مجموعة بيانات نظيفة ومتوازنة، ما يساعدني على تدريب النماذج بشكل أفضل ويعطيني نتائج دقيقة.

4. Isolation Forest كيف تعمل خوارزمية

- **الإجابة الأكاديمية:**

تعتمد على تقسيم البيانات بشكل عشوائي. النقاط التي يتم عزلها بسرعة (بعد عدد قليل من التقسيمات) يتم Isolation Forest اعتبارها شذوذاً. تعمل هذه الطريقة بكفاءة مع البيانات الكبيرة والعالية الأبعاد.

- **الإجابة العامة:**

الخوارزمية تقسم البيانات بطريقة عشوائية. النقاط الشاذة تكون سهلة العزل، لذلك يمكن التعرف عليها بسرعة.

5. ما هي نقاط القوة والضعف في One-Class SVM؟

- **الإجابة الأكاديمية:**

نقاط القوة تشمل دقته العالية في تحديد الشذوذ في الحالات التي يكون فيها السلوك الطبيعي واضحاً. ومع ذلك، فهو حساس جداً لوجود بيانات غير متوازنة ويتطلب وقتاً أطول لمعالجة البيانات الكبيرة.

- **الإجابة العامة:**

قوته أنه دقيق جداً لو البيانات طبيعية ومفهومة. لكن مشكلته أنه بطيء مع البيانات الكبيرة ويتأثر لو البيانات مش متوازنة.

6. ما الذي يجعل التعلم الآلي أفضل من الطرق التقليدية في IDS؟

- **الإجابة الأكاديمية:**

تقنيات التعلم الآلي لا تعتمد على التوقعات أو القواعد الثابتة، مما يسمح لها بالتكيف مع التهديدات الجديدة مثل الهجمات غير بالإضافة إلى ذلك، يمكنها تحليل البيانات الكبيرة والمعقدة بكفاءة. (Zero-Day) المعروفة.

- **الإجابة العامة:**

التعلم الآلي أذكى لأنه يتعلم من البيانات بدل ما يعتمد على القواعد الثابتة. هذا يجعله قادراً على كشف هجمات جديدة ما كنا نعرفها من قبل.

7. ما هي التحديات التي واجهتها في المشروع؟

- **الإجابة الأكاديمية:**

تضمنت التحديات التعامل مع البيانات غير المتوازنة، تحسين معلمات النماذج لتحقيق الأداء الأمثل، وضمان كفاءة الأداء في البيانات الزمنية الحقيقية. تمت معالجة هذه التحديات من خلال تقنيات معالجة البيانات المسبقة وضبط المعلمات بدقة.

- **الإجابة العامة:**

كان عندي تحديات مثل البيانات اللي مش متوازنة وضبط إعدادات النماذج. لكن باستخدام تقنيات تنظيف البيانات وضبط الإعدادات، قدرت أحسن الأداء.

8. كيف يمكن تحسين المشروع مستقبلاً؟

- **الإجابة الأكاديمية:**

يمكن تحسين المشروع من خلال دمج مع أنظمة كشف التسلل في الوقت الحقيقي، واختباره على مجموعات بيانات أكثر تقدماً، واستكشاف تقنيات التعلم العميق لتحسين الأداء والدقة مثل CIC-IDS2017.

- **الإجابة العامة:**
يمكن أطور المشروع عشان يشتغل في الوقت الحقيقي وأستخدم مجموعات بيانات أحدث وأكبر. كمان ممكن أجرب تقنيات أذكى زي التعلم العميق.
-

9. ما هي الفائدة العملية من هذا المشروع؟

- **الإجابة الأكاديمية:**
يوفر المشروع أساساً قوياً لتطوير أنظمة كشف التسلل القابلة للتوسع والفعالة. يمكن دمجها في بيئات حقيقية لحماية الشبكات من الهجمات المتطورة.
 - **الإجابة العامة:**
الفائدة إن المشروع ممكن يساعد الشركات والمؤسسات في حماية شبكاتهم من الهجمات بسهولة وكفاءة.
-

10. في مشروعك؟ (GUI) ما هي ميزة واجهة المستخدم الرسومية.

- **الإجابة الأكاديمية:**
الواجهة تتيح للمستخدمين تحميل البيانات، اختيار النماذج، وضبط المعلمات بسهولة، مما يجعل النظام مناسباً حتى لغير المتخصصين في البرمجة.
- **الإجابة العامة:**
الواجهة تسهل التعامل مع النظام. أي شخص يقدر يستخدمها بدون ما يحتاج خبرة تقنية كبيرة.

1. كيف تضمن أن النظام يعمل بكفاءة على بيانات لم يرها من قبل؟

- **الإجابة الأكاديمية:**
لضمان كفاءة النظام على البيانات غير المرئية، قمنا بتقسيم البيانات إلى مجموعات تدريب واختبار. مجموعة التدريب تُستخدم لتعليم النموذج، بينما مجموعة الاختبار تُستخدم لتقييم الأداء على بيانات جديدة. كما أننا استخدمنا تقنيات التحقق المتقاطع (Cross-Validation) لتعزيز موثوقية النتائج.
 - **الإجابة العامة:**
قسّمنا البيانات لمجموعة تدريب ومجموعة اختبار. التدريب يعلم النموذج، والاختبار يثبت إنه يشتغل صح على بيانات جديدة.
-

2. كيف يمكن للنظام التعامل مع الهجمات في الزمن الحقيقي؟

- **الإجابة الأكاديمية:**
SIEM (Security Information and Event Management) يمكن تعديل النظام ليعمل في الزمن الحقيقي من خلال دمجها مع بيئات شبكية مثل أدوات يتم جمع البيانات بشكل مباشر من الشبكة ومعالجتها عبر النظام للكشف عن أي شذوذ في الوقت الفعلي.
 - **الإجابة العامة:**
يمكن نوصّل النظام بالشبكة مباشرة عشان يراقب البيانات أول بأول ويكشف الهجمات في نفس اللحظة.
-

3. Isolation Forest في (contamination) كيف تختار نسبة الشذوذ.

- **الإجابة الأكاديمية:**
auto-tune.نسبة الشذوذ يمكن تحديدها بناءً على تحليل مسبق للبيانات، أو ترك النموذج يحددها تلقائيًا باستخدام أدوات مثل
في حالة عدم المعرفة المسبقة، تُستخدم القيم الافتراضية مثل 0.1 كنسبة تقريبية.
 - **الإجابة العامة:**
إما نحلل البيانات ونحدد النسبة أو نستخدم النسبة الافتراضية التي تعطي نتائج جيدة في معظم الحالات.
-

4. هل يمكن استخدام نموذج واحد فقط بدلاً من الاثنين؟

- **الإجابة الأكاديمية:**
يناسب البيانات الكبيرة والمعقدة، بينما Isolation Forest. نعم، لكن استخدام نموذج واحد قد يكون محدودًا في بعض الحالات
يحقق دقة أعلى في البيانات ذات السلوك الطبيعي الواضح. لذا، الجمع بين النموذجين يعزز القدرة على One-Class SVM
التعامل مع سيناريوهات متنوعة.
 - **الإجابة العامة:**
ممكن، لكن استخدام الاثنين معًا يعطينا نتائج أفضل لأن كل واحد عنده نقاط قوة مختلفة.
-

5. كيف يمكن تحسين النظام أكثر؟

- **الإجابة الأكاديمية:**
 - (Deep Learning) إضافة نماذج أكثر تعقيدًا مثل التعلم العميق.
 - تحسين واجهة المستخدم لتشمل خيارات أكثر.
 - لمعرفة مدى تكيفه CIC-IDS2017 اختبار النظام على مجموعات بيانات جديدة مثل.
 - **الإجابة العامة:**
ممكن نضيف نماذج أذكى، نحسن الواجهة لتكون أسهل، ونختبره على بيانات أكبر وأحدث.
-

6. ما هو أكبر تحدٍ واجهته أثناء المشروع؟

- **الإجابة الأكاديمية:**
كان التحدي الأكبر التعامل مع البيانات غير المتوازنة وتحسين معلمات النماذج للحصول على أفضل أداء. كما أن دمج واجهة
المستخدم الرسومية مع النماذج كان يتطلب الكثير من الاختبارات لضمان السلاسة.
 - **الإجابة العامة:**
أكبر تحدي كان البيانات غير المتوازنة وضبط الإعدادات. كمان الواجهة الرسومية أخذت وقت عشان تشتغل بشكل سلس.
-

7. ما هي أهم الأخطاء التي يمكن أن تواجه النظام؟

- **الإجابة الأكاديمية:**
 - يتم تصنيف حركة طبيعية على أنها شذوذ: (False Positives) أخطاء إيجابية كاذبة.

- يتم تصنيف الشذوذ على أنه طبيعي (False Negatives) أخطاء سلبية كاذبة وتحديد نسبة الشذوذ بدقة في One-Class SVM و γ و ν لتحسين ذلك، نقوم بتعديل المعلمات مثل Isolation Forest.

- **الإجابة العامة:**
النظام ممكن يغلط ويعتبر شيء طبيعي كأنه شذوذ أو العكس. ضبط الإعدادات يقلل الأخطاء.

8. هل يمكن استخدام النظام في مجالات أخرى غير الأمن السيبراني؟

- **الإجابة الأكاديمية:**
نعم، يمكن استخدام النظام في مجالات أخرى مثل اكتشاف الاحتيال في البنوك، مراقبة الأنظمة الصناعية، أو حتى في تحليل البيانات الطبية للكشف عن الحالات غير الطبيعية.
- **الإجابة العامة:**
أكيد! ممكن نستخدمه في اكتشاف الاحتيال أو في أنظمة المصانع وحتى في تحليل البيانات الطبية.

9. ما هي ميزات واجهة المستخدم مقارنة بالأنظمة الأخرى؟

- **الإجابة الأكاديمية:**
واجهة المستخدم تم تصميمها لتكون سهلة الاستخدام وتتيح للمستخدم تحميل البيانات، اختيار النماذج، وضبط المعلمات دون الحاجة إلى خبرة برمجية. كما أنها تعرض النتائج بشكل رسومي لفهم أفضل.
- **الإجابة العامة:**
الواجهة سهلة جدًا، أي شخص يقدر يستخدمها بدون ما يكون خبير. فيها تحميل بيانات، اختيار النماذج، وعرض النتائج بشكل بسيط.

10. كيف يمكن دمج النظام في بيئات عمل حقيقية؟

- **الإجابة الأكاديمية:**
لتوفير كشف في الوقت الحقيقي. كما يمكن تعديله ليعمل كخدمة (SIEM) يمكن دمج النظام مع أنظمة إدارة الأحداث الأمنية. سحابية بحيث يتمكن المستخدمون من تشغيله عن بُعد.
- **الإجابة العامة:**
ممكن نوصل النظام بالشبكة أو نخليه خدمة على السحابة عشان الناس تستخدمه بسهولة.

1. C++ أو Java لماذا لم تستخدموا تقنيات أخرى مثل 1.

- **الإجابة الأكاديمية:**
Python، Matplotlib، و NumPy و Scikit-learn لأنها توفر مكتبات متقدمة وسهلة الاستخدام مثل Python اخترنا ، فهي قوية، لكنها C++ أو Java تُستخدم على نطاق واسع في مجالات التعلم الآلي ولها مجتمع كبير يقدم الدعم والموارد. أما ، تتطلب وقتًا أطول لتطوير النماذج وتنفيذ الخوارزميات.
- **الإجابة العامة:**
تحتاج وقت أطول وبرمجة معقدة، وما كان عندنا وقت كافي C++ و Java. لأنها أسهل وأسرع للتعلم الآلي Python اخترنا.

2. لماذا لم تطوروا المشروع كتطبيق هاتف محمول؟

- **الإجابة الأكاديمية:**
استخدام التعلم الآلي في كشف التسلل. تطوير (Proof of Concept) كان الهدف الرئيسي من المشروع هو إثبات مفهوم تطبيق هاتف يتطلب تركيزًا إضافيًا على التصميم والتفاعل، وهو خارج نطاق المشروع الحالي. يمكننا في المستقبل تحويل Flutter أو Kivy النظام إلى تطبيق هاتف باستخدام أدوات مثل.
 - **الإجابة العامة:**
الهدف كان نركز على الخوارزميات نفسها، وتطوير تطبيق موبايل يحتاج وقت وخبرة زيادة. ممكن نعمله بعدين لو في وقت.
-

3. بدلاً من التعلم الآلي التقليدي؟ (Deep Learning) لماذا لم تستخدموا تعلم عميق

- **الإجابة الأكاديمية:**
في هذا المشروع، اخترنا (GPU) التعلم العميق يتطلب بيانات ضخمة وموارد حوسبة عالية مثل وحدات معالجة الرسومات صغيرة نسبياً، والنماذج التقليدية أثبتت فعاليتها على هذه البيانات NSL-KDD التعلم الآلي التقليدي لأن مجموعة البيانات.
 - **الإجابة العامة:**
التعلم العميق يحتاج بيانات أكبر وأجهزة أقوى مثل كروت شاشة. مشروعنا كان بسيط وبياناته مش كبيرة، والنماذج اللي اخترناها كانت كافية.
-

4. (SIEM مثل أدوات) لماذا لم ندمجوا النماذج في نظام موجود؟

- **الإجابة الأكاديمية:**
يتطلب وقتاً وموارد SIEM دمج النماذج في أنظمة مثل (Prototype) الهدف من المشروع هو بناء نظام مبدئي وتجريبي إضافية، وهو يمكن تنفيذه كجزء من التطوير المستقبلي للمشروع.
 - **الإجابة العامة:**
عملنا النظام كنموذج بسيط عشان نثبت فكرته. لو في وقت أكثر ممكن نطوره ونربطه بأنظمة متقدمة.
-

5. (Hybrid IDS) لماذا اخترتم التركيز على الكشف عن الشذوذ فقط ولم تستخدموا نظاماً هجيناً؟

- **الإجابة الأكاديمية:**
(Unsupervised Models) التركيز على الكشف عن الشذوذ فقط يسمح لنا بدراسة أداء النماذج غير الخاضعة للإشراف وتحليل فعاليتها. الأنظمة الهجينة تتطلب دمج بين الكشف المعتمد على التوقعات والكشف عن الشذوذ، مما يزيد التعقيد والجهد المطلوب في هذه المرحلة.
 - **الإجابة العامة:**
اخترنا الشذوذ لأنه أسهل نبدأ فيه ونفهم كيف الخوارزميات تشتغل. النظام الهجين يحتاج شغل أكبر ووقت أطول.
-

6. Zero-Day؟ هل يمكن لنظامكم التعامل مع هجمات متقدمة مثل

- **الإجابة الأكاديمية:**
لأنه يعتمد على الكشف عن الشذوذ بدلاً من التوقعات المحددة مسبقاً. Zero-Day نعم، نظامنا قادر على التعامل مع هجمات. عندما يظهر نشاط غير مألوف، يقوم النظام بتصنيفه كعنصر شاذ.

- **الإجابة العامة:**
نعم، لأنه النظام يركز على أي شيء غريب أو جديد، مو بس على الأنماط المعروفة.
-

7. لماذا لم تستخدموا بيانات أحدث مثل CIC-IDS2017؟

- **الإجابة الأكاديمية:**
CIC-IDS2017 لأنها متوازنة وسهلة التحليل، مما يسمح لنا بتطوير نماذج مبدئية بسرعة. استخدام بيانات حديثة مثل KDD-NSL اخترنا يتطلب المزيد من الموارد ووقت إضافي، ويمكن أن يكون جزءًا من العمل المستقبلي IDS2017.
 - **الإجابة العامة:**
يمكن استخدامها لو في وقت أطول CIC-IDS2017. كانت أبسط وتناسب المشروع KDD-NSL بيانات.
-

8. كيف تضمنون أن النظام قابل للتوسع؟

- **الإجابة الأكاديمية:**
Django أو Flask النظام مصمم ليكون مرناً وقابلًا للتوسع. يمكن دمجه بسهولة مع منصات أخرى باستخدام تقنيات مثل GPU) لتوفير خدمات في الزمن الحقيقي. كما يمكن تحسين الأداء باستخدام وحدات معالجة متقدمة.
 - **الإجابة العامة:**
النظام بسيط ومرن. يمكن نضيف له ميزات أو نخليه أسرع لو احتجنا.
-

9. هل قمتم بتقييم مدى قابلية النظام للاستخدام في سيناريوهات حقيقية؟

- **الإجابة الأكاديمية:**
في هذه المرحلة، تم اختبار النظام على مجموعة بيانات محددة كدليل إثبات. في المستقبل، يمكن توسيع الاختبارات لتشمل بيانات واقعية باستخدام بيانات حية.
 - **الإجابة العامة:**
لنا اختبارناه على بيانات محددة، بس ممكن نظوره عشان يشتغل على بيانات حقيقية في المستقبل.
-

10. ما هو الوقت المتوقع لتطوير المشروع ليصبح نظامًا كاملاً؟

- **الإجابة الأكاديمية:**
يعتمد ذلك على الموارد المتاحة. مع فريق عمل وبيئة تطوير مناسبة، يمكن تحويل النظام إلى منتج كامل خلال 6-12 شهرًا، يشمل الميزات الإضافية مثل التكامل مع الأنظمة الأخرى.
- **الإجابة العامة:**
لو عندنا فريق ووقت كافي، ممكن نظوره بالكامل خلال حوالي سنة.