CIS 350: AI for Cybersecurity

Week 5 Activity Worksheet: Anomaly Detector False Alarms

CISA-Style Incident Tabletop Exercise

Team members (names): _____

Your assigned role: _____

Activity Overview

In this activity your small group (3-5 people) runs a CISA-style incident tabletop. There are no live systems - you talk through a crisis and record your decisions, just like real security teams do when they rehearse. You will move through three timed rounds of about five minutes each. In every round you must record the team decision AND the reasoning behind it, using risk, cost, and evidence. Model your discussion on the CISA Tabletop Exercise Packages (cisa.gov/resources-tools/services/cisa-tabletop-exercise-packages). Note: Exam 1 (Weeks 1-4) is online this week, so this is a focused in-class group exercise, not heavy homework.

Key background. An unsupervised anomaly detector flags activity that looks UNUSUAL without any labeled examples of attacks (think Isolation Forest or clustering outliers). It says 'this is weird,' not 'this is an attack.'

Scenario Briefing (read aloud before Round 1)

Your company, Harborline Payments, processes online card payments for businesses. Security deployed an AI anomaly detector on login and payment sessions. It has no list of known attacks - it just flags sessions that look unusual. Every day it flags about 100 sessions. Analysis shows roughly 98 percent are legitimate (a customer traveling, a big end-of-month payment run, a new device) and only about 2 percent - around two sessions a day - are real attacks. Your overnight analyst team is three people and cannot hand-review 100 alerts. The CFO just asked: 'Should we just auto-block them all?'

Task 1: Assign Roles and Read the Resource

Assign one role to each person, then write who has each role. If your group has 3-4 people, the Scribe may also take a second role. Before you start, skim the CISA Tabletop Exercise Packages resource so your discussion mirrors how professionals run a tabletop.

Role	Responsibility	Team member
Incident Commander	Runs the clock, calls the final decision	
Technical Lead	Interprets the detector numbers and tuning	
Communications Lead	Drafts messages to customers and the CFO	
Legal / Compliance	Flags liability from blocks and from misses	

Scribe	Records every decision and its reasoning	

Decision framework - use this in every round: RISK (who gets hurt if we are wrong - blocked customers, or a missed breach), COST (analyst hours, lost payments, refunds, churned clients, reputation), EVIDENCE (what the numbers prove: 100 alerts/day, about 98 percent false, about 2 real attacks).

Round 1: Triage Today's 100 Alerts (~5 min)

One hundred sessions are flagged right now and your team is three analysts. You cannot hand-review all 100 tonight. How do you handle today's flood?

Decision space (weigh these):
- Auto-block every flagged session immediately
- Ignore the queue tonight and review in the morning
- Triage: rank by account value and dollar amount, review the top slice by hand, and step-up-verify (extra login check) the rest

Before you decide, settle this as a team: **Is a flagged session the same as a confirmed attack?** Note what the 98 percent / 2 percent split means for treating all 100 alerts equally.

Team decision (Round 1):

Justification (risk / cost / evidence):

Round 2: A Complication Is Injected (~5 min)

Two things happen at once. First, one of last night's un-reviewed alerts WAS a real account takeover - a client just reported $40,000 stolen. Second, the CFO now orders 'auto-block everything,' but a hospital client was auto-blocked during a test, could not run payroll, and is threatening to cancel its contract. Does this change your Round 1 plan?

Consider:
- Do you turn on auto-block now, given a real attack slipped through?
- What is the cost of blocking 98 legitimate customers to stop 2 attackers?
- Would you rather tighten or loosen the detector, add a human reviewer, or step up verification?

What changed from Round 1, and why:

Your message to the CFO (2-3 sentences - is auto-block the right call?):

Round 3: Final Operating Decision and Documentation (~5 min)

Time to commit. Decide how Harborline will OPERATE this detector going forward. Set the tuning dial (the contamination rate: too strict misses real attacks, too loose floods analysts), decide what is automated versus human-reviewed, and how you will staff it. Justify it one more time with risk, cost, and evidence. Then write the operating / communication artifact an auditor or your VP would read later.

Final operating decision and justification (include where you set the tuning dial and why):

Operating / communication artifact. Write 4-6 sentences that state: (1) how you will handle the 100 daily alerts and why, (2) how you will MEASURE whether the detector actually prevents attacks (not just generates alerts), (3) what you tell the CFO and the hospital client, and (4) what governance change prevents this next time.

Task 2: Tie It to Lab 5 (Clustering and Anomaly Detection)

Lab 5, released next week, builds this hands-on. You will run an Isolation-Forest-style anomaly detector and tune its CONTAMINATION RATE - the same dial you set in Round 3 - against a business cost rather than a plain accuracy number.

Q1: Explain the contamination-rate trade-off in your own words. What happens to missed attacks and to false alarms when you set the dial too STRICT? Too LOOSE?

Q2: 'Alerts generated' is not the same as 'attacks prevented.' Name two measurements you would track over several weeks to prove the detector is actually protecting Harborline.

Q3: The detector has no labeled examples of attacks. Explain, in plain language, how an unsupervised method can flag a suspicious session anyway - and why 'unusual' is not the same as 'malicious.'

Task 3: Debrief and Governance Gap

Q1: What changed between Round 1 and Round 3, and what forced the change?

Q2: What would you document about tonight so your operating decision survives an external review months from now?

Q3: What governance gap did this reveal? Name a policy, staffing plan, or review step that SHOULD have existed before the detector went live - and say who decides the acceptable balance of missed attacks versus blocked customers.

Grading Rubric (20 points)

Criterion	Description	Points
Completion	Participates in all 3 rounds; every decision recorded	8
Reasoning	Decisions justified with risk, cost, and evidence	7
Communication	Clear stakeholder message and audit documentation	5

Total: 20 points
Scoring: Completion (8) + Reasoning (7) + Communication (5) = 20 points

How to Submit

1. Complete all three rounds with your group, recording each decision and its justification
2. Write the Round 3 operating / communication artifact
3. Answer the Lab 5 tie-in questions (Task 2) and the debrief questions (Task 3)
4. Each student submits their own completed PDF worksheet on Canvas by [DUE DATE]

Questions? Post in the course forum or attend office hours.