CIS 350: AI for Cybersecurity

Week 2 Activity Worksheet: Data Exposure and Privacy Violations

Activity Overview

This week's lecture is about datasets, features, and labels. In this activity you will explore the security side of that topic: how sensitive data gets exposed, why we should collect as little of it as possible, and how poor-quality or mislabeled data can be turned into a weapon against a machine-learning model. You will watch/read a short guided lesson, answer questions, analyze the data quality of the Lab 2 email dataset, and design a data-minimization and labeling checklist you could hand to a real team.

Warm-Up (5 min) - Think, Pair, Share

Before the lesson, take 60 seconds to write a quick prediction, then compare with a partner. This is ungraded - it just primes your thinking.

Predict: A company suffers a data breach. Which do you think exposes people to more harm - that the company was hacked by a skilled attacker, or that the company simply collected and stored more data than it needed? Jot your guess and one reason.

Pair: Share with a partner. Did you pick the same cause? Hold onto your prediction - you will revisit it after the lesson.

Task 1: Watch/Read the Lesson

Resource: Hacker101 Lessons (hacker101.com/lessons) or OWASP Sensitive Data Exposure / OWASP Top Ten (owasp.org/www-project-top-ten)
Estimated time: ~15 minutes

As you watch/read, take notes on:
- What kinds of sensitive data appear (PII, credentials, training data)
- How the data was exposed (collected too much, stored unprotected, left public)
- What data minimization would have changed
- How access control and encryption reduce the damage of a breach

Task 2: Guided Questions

Q1: In your own words, what is 'sensitive data'? Give one example each of PII, a credential, and training data.

Q2: Explain the principle of data minimization. Give one example of data a company might collect but does not actually need.

Q3: What does it mean for a training label to be 'wrong', and why does a mislabeled dataset create a security risk?

Q4: Describe, step by step, how an attacker could poison a spam classifier by tampering with its training labels.

Q5: Give one feature that is useful but privacy-leaking, and explain the tradeoff between keeping it and dropping it.

Task 3: Analyze the Lab 2 Email Dataset

In Lab 2 ('Exploring Security Data') you worked with a small email dataset. Recall the numbers:
- 24 emails total: 8 spam and 16 ham (33.3% of the emails are spam)
- Spam emails average about 6.25 links each; ham emails average about 0.94 links each
- The lab adds data-quality checks (missing values, duplicates) and feature-importance / risk-scoring

Questions:
1. The dataset is imbalanced (only 33.3% spam). If a model simply guessed 'ham' for every email, what accuracy would it get, and why would that number be misleading? Show your reasoning.

2. Suppose 3 of the 8 spam emails were accidentally labeled as 'ham' (a labeling error). Explain how this would change what the model learns about the link-count feature (remember: spam ~6.25 links, ham ~0.94 links). Would the two averages move closer together or farther apart?

3. With only 24 rows, why do a few duplicate or missing-value rows matter more than they would in a dataset of 24,000 rows? Connect your answer to why Lab 2's data-quality checks run BEFORE training.

Checkpoint - discuss with a partner (2 min): In Part 2 you decided whether the two link-count averages move closer together or farther apart after the labeling error. Compare your answer and your reasoning with a partner. If you disagree, re-check which group the high-link emails landed in. Agree on a one-sentence explanation before moving on.

Task 4: Design a Data-Minimization and Labeling Checklist

You are the data lead on a team building an email security filter. Before your team collects and labels any data, you must give them a checklist that protects privacy AND protects the model. Write a checklist of 6-8 items. Cover BOTH sides: (a) minimizing and protecting sensitive data, and (b) catching data-quality and labeling problems before training.

Your Checklist (use a numbered list):

Explain two of your items: Pick two checklist items and explain what specific risk each one prevents (name the attack or failure it defends against).

Task 5: Reflection

Q1: Data minimization means sometimes NOT collecting data you could easily grab. Why is that hard for companies to do in practice? What pressures push them to over-collect?

Q2: This week's lecture treated labels as just 'the answer key.' After this lesson, how has your view of label quality as a SECURITY issue (not just an accuracy issue) changed?

Q3: Name one piece of data an app on your own phone collects that it probably does not need. What is the worst thing that could happen if that data were exposed?

Grading Rubric (20 points)

Criterion	Description	Points
Completion	Lesson watched/read; all 5 tasks and questions attempted	8
Understanding	Q&A answers demonstrate learning about exposure, minimization, and poisoning	7
Application	Lab 2 analysis and checklist are sound, specific, and well-explained	5

Total: 20 points

Scoring: Completion (8) + Understanding (7) + Application (5) = 20 points

How to Submit

1. Complete all 5 tasks in this worksheet
2. Ensure the Lab 2 analysis uses the given numbers and shows your reasoning
3. Make sure your checklist has 6-8 items and explains two of them
4. Submit this PDF on Canvas by [DUE DATE]

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