Sample Quiz: The EAT Button

Pass 1 of 2

ECE 4424 / CS 4824: Machine Learning: Learn by Building — Fall 2026

Instructor: Ming Jin — Virginia Tech

Sample only - not for credit. This shows the format that future in-class quizzes may use. For open-ended judgments, we grade your reasoning and use of evidence, not whether you match one preferred answer. When there is a clear right or wrong answer, such as a calculation or definition, correctness matters.

Name: _____ **Course:** ☐ ECE 4424 ☐ CS 4824 **Section:** _____

TA name: _____

Situation

ForestFork is fictional. Users enter mushroom traits, no human reviews the result, and the app displays **EAT** whenever the estimated $P(\text{poisonous})$ is below its threshold. A false negative means a poisonous mushroom is shown as EAT.

Prompt sent to the AI

"Choose one model today. Give us a clear go/no-go decision. Use normal defaults unless the evidence strongly requires otherwise."

AI response (abridged into five claims)

---

1 Choose logistic regression. At threshold 0.50, it has 42 false negatives, compared with 81 for categorical Naive Bayes.

2 Accuracy is not the main issue: in this app, a false negative means a poisonous mushroom is shown as EAT.

3 Naive Bayes has slightly better log loss, but scores around 0.16 to 0.25 are badly miscalibrated; these apparently low-risk cases are usually poisonous.

4 The supplied priors imply $P(\text{poisonous} \mid \text{foul odor}) \approx 88\%$, so a foul odor should strongly favor DO NOT EAT.

5 Use logistic regression with a 0.20 threshold and launch. This lower threshold is conservative enough to make the public EAT action safe.

Pass 1: Cold audit (3 minutes)

On the response, **underline** the strongest claim, add **?** to one claim you would verify, and add **!** to the most dangerous leap.

Initial disposition: ☐ SHIP ☐ HOLD ☐ CANNOT YET TELL

One-line reason: _____

Sample Quiz: The EAT Button

Pass 2 of 2

ECE 4424 / CS 4824: Machine Learning: Learn by Building — Fall 2026

Instructor: Ming Jin — Virginia Tech

Use a different ink if available; otherwise box every new mark. The AI saw only the 0.50 results before answering. The 0.20 results and data-scope note are revealed now.

Evidence packet

Model / threshold	TN	FP	FN	TP	Accuracy
Logistic / 0.50	811	31	42	741	95.5%
Logistic / 0.20	764	78	11	772	94.5%
Naive Bayes / 0.50	835	7	81	702	94.6%
Naive Bayes / 0.20	830	12	52	731	96.1%

Denominators. The test set contains 842 edible and 783 poisonous examples. $\text{FNR} = \text{FN}/(\text{FN} + \text{TP})$; $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$.

Calibration. For Naive Bayes, score 0.156 corresponds to 87.0% actually poisonous; score 0.253 corresponds to 91.7%. Some bins are small.

Data and pipeline. There are 8,124 hypothetical descriptions but only 23 species. Nominal features were integer-coded, and the split was random rather than species-held-out.

Bayes inputs. $P(\text{poisonous}) = 0.482$, $P(\text{foul} \mid \text{poisonous}) = 0.80$, and $P(\text{foul} \mid \text{edible}) = 0.10$.

1. Claim audit

Choose the claim most needing correction and connect your verdict to evidence.

Claim number: _____ **Verdict:** ☐ supported ☐ unsupported ☐ contradicted

2. Make the risk legible

A. Logistic threshold 0.20: $\text{FNR} = 11/\_\_\_\_\_ = \_\_\_\_\_\%$

B. Reverse the conditional: $P(\text{poisonous} \mid \text{foul}) = \dfrac{0.80(0.482)}{0.80(0.482) + 0.10(0.518)} = \_\_\_\_\_\%$

C. One sentence: Why does a model score below 0.20 not guarantee field safety above 80%?

3. Make the AI-fluent next move

Write one follow-up request that names **a test or artifact** and **a stop condition**.

Final disposition

☐ LAUNCH EAT ☐ PILOT WITHOUT EAT ☐ STOP / REDESIGN

Revision trace: ☐ I changed my mind ☐ Same verdict, stronger reason

Decisive reason: _____