FDA Drug Recall Classifier

Predicting Recall Severity from Reason Text

Jannet Castaneda Sanchez
ID: 105863950
STATS 418-Spring 2025

BACKGROUND

Problem: FDA recall notices are manually classified into Class I (urgent) to Class III (least risk), often with a delay.

Goal: Predict recall severity from the "reason for recall" text using machine learning to support faster regulatory and clinical action.

Motivation:

- Class I recalls involve the risk of serious injury/death.
- Timely classification helps protect public health.

DATASET & ACQUISITION METHOD

Source: FDA Enforcement Report API: https://api.fda.gov/drug/enforcement.json

- Sample size: 16,908 drug recalls (06/2012-05/2025)
- Row: Single drug recall event
- Variables: Recall Number, Classification, Reason, Product Description, Report Date, Recalling Firm, State, & Country

Exploration Highlights:

- Class II dominates (60%), followed by Class III (30%) and Class I (10%)
- Term Frequency used to identify key terms by class
- Word clouds revealed distinct language patterns

FRAMEWORK & MODEL

Problem: FDA recall notices are manually classified into Class I (urgent) to Class III (least risk), often with a delay.

Goal: Predict recall severity from the "reason for recall" text using machine learning to support faster regulatory and clinical action.

Motivation:

- Class I recalls involve the risk of serious injury/death.
- Timely classification helps protect public health.

TAKEAWAY & NEXT STEPS

Takeaways:

- Natural Language Processing-based classifier can assist in the early detection of urgent recalls
- Deployed end-to-end using API + Shiny App

Next Steps:

- Enable batch predictions
- Add model explainability

Repo: https://github.com/jannet1313/fda-drug-recall-classifier

Thank you!