**Initial framing and deviation**

From the initial framing of the problem to report out a food complaint, the analysis is modified to determine if the complaint received qualifies for a recall and what category of recall does the reason for re-call fall into.

**Why the question is important** :  Primarily the Food Manufacturing industry takes significant time and resources to determine whether the complaint classifies as a recall and what kind of class it belongs to for the urgency of the recall process. One of the graphs from determines the lag in number of days since a complaint was filed and a classification was determined because of the heavily human interfered process. The ML initiative is to create a model to identify if the event/complaint qualifies for a recall and which class of recall it belongs to.

 Recalls are actions taken by a firm to remove a product from the market. Recalls may be conducted on a firm's own initiative, by FDA request, or by FDA order under statutory authority.

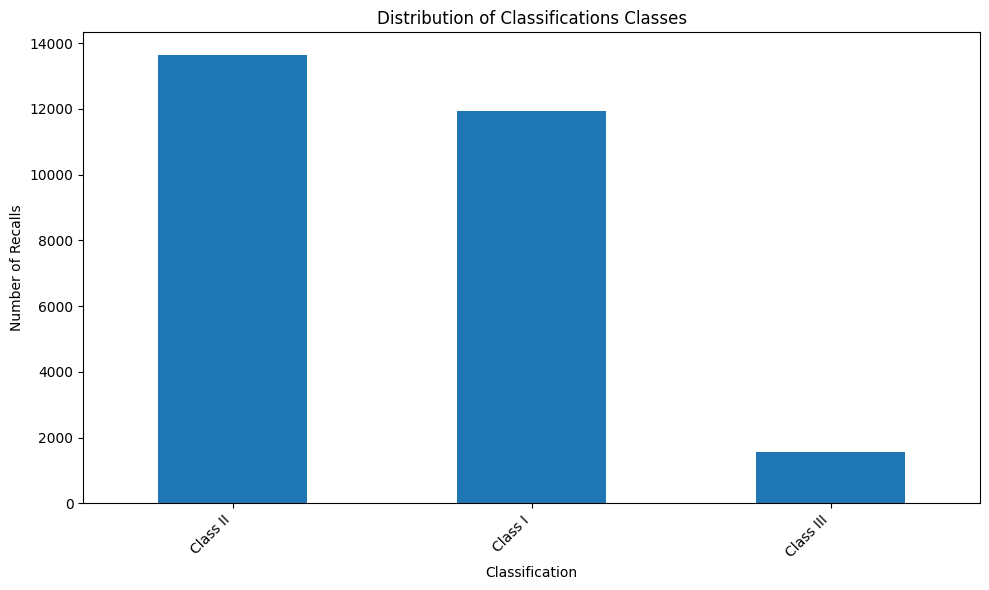
Class I recall: a situation in which there is a reasonable probability that the use of or exposure to a violative product will cause serious adverse health consequences or death.

Class II recall: a situation in which use of or exposure to a violative product may cause temporary or medically reversible adverse health consequences or where the probability of serious adverse health consequences is remote.

Class III recall: a situation in which use of or exposure to a violative product is not likely to cause adverse health consequences.

**Dataset** **and Initial analysis**

The dataset used for generating the model is directly downloaded files in JSON format from the FDA complaints and recalls database. The focus was primarily on the reason for recall field, where some of the key words for recall for grouped to evaluate the type of recalls at large. But mostly this would be a classical NLP based multi-classification model. And for that reason, Naïve Bayes model was chosen as the base line model.



A graph with blue rectangular bars

AI-generated content may be incorrect.

A screenshot of a chart

AI-generated content may be incorrect.

A graph of a number of columns

AI-generated content may be incorrect.

A graph with different colored squares

AI-generated content may be incorrect.

A graph with different colored bars

AI-generated content may be incorrect.

**Baseline Model output**

Accuracy: 0.815033161385409

precision recall f1-score support

Class I 0.86 0.82 0.83 2401

Class II 0.78 0.88 0.83 2718

Class III 0.86 0.23 0.36 309

accuracy 0.82 5428

macro avg 0.83 0.64 0.67 5428

weighted avg 0.82 0.82 0.81 5428

A blue squares with numbers

AI-generated content may be incorrect.

**Evaluation Metrics Interpretation**

The two evaluation metrices: accuracy and the classification report (which includes precision, recall, F1-score, and support).

An accuracy score of 81.5% for a baseline model is pretty descent considering the vast amount of data evaluated.

For each class, precision is the proportion of correctly predicted positive observations to the total predicted positive observations.  A high precision means that when the model predicts a class, it is likely to be correct.

For each class, recall is the proportion of correctly predicted positive observations to the all observations in actual class. A high recall means that the model is good at identifying all positive instances of a given class. The F1-score is the harmonic mean of precision and recall.  It provides a balanced measure that considers both false positives and false negatives.  It's particularly useful when the classes are imbalanced.

While accuracy gives an overall picture, the classification report provides a more granular analysis of the model's performance for each class. This is crucial because the classes might be imbalanced (different numbers of Class I, II, and III recalls).  Accuracy alone could be misleading if one class significantly outnumbers others.  The classification report helps identify if the model is performing well for all classes or if it's biased towards certain classes, particularly important for recall which is how well it finds all of a specific class of recalls.  In this food recall context, correctly identifying all Class I recalls (most serious) is critical even if the model makes a few more errors on less severe recalls. A low recall and F1 Score for Class III due to lesser data.

The confusion matrix depicts the model’s accuracy comparing the actual and predicted classifications. The model is decent, but since the business case is specifically for health, we need to explore more accurate models.