

<https://chicago.github.io/food-inspections-evaluation/>

The Problem

There are over 15,000 establishments serving food in Chicago that means that each inspector is responsible for 470 inspections. The result is that food establishments with critical violations may not be discovered until many patrons have become sick.

The Intuition

If data could be used to predict which establishments would likely score the critical violations, those could be inspected earlier so that the violations may be discovered, improving the overall safety of the community.

The Approach

The team determined that the following predictors would be included in the model:

- Prior critical or serious violations
- 3 day average temperature
- Nearby garbage and sanitation complaints
- Types of facility inspected
- Nearby burglaries
- Whether the establishment has a tobacco license or incidental alcohol consumption
- Length of time since last inspection
- Length of time establishment has been operating
- Inspector assigned

They used a random forest model using the city's open data initiative (complaints, weather, business licenses, crime, inspections), and R to create it, then piloted it in a 2 month initiative. They assigned a probability for

<http://datasmart.ash.harvard.edu/news/article/delivering-faster-results-with-food-inspection-forecasting-631>

The Problem

15% of establishments, on average earn critical violations. Most violations are related to food temperature control, which means they are very likely to spread food-borne illness.

The important thing to note is that the process needs to be designed in a way that helps inspectors focus their efforts in the most efficient manner, not to replace them.

The Approach

The team interviewed inspectors to understand the domain. They partnered city team members with commercial companies (like Allstate) to create an open-source project that other cities could use.

The establishment was assigned a probability based on the number of predicting factors met. They then performed a simulation and compared the results to the actual inspection outcomes to determine accuracy. The simulation showed that they were able to predict 69% of the critical violations (much higher than the manual process yielded at 55%).

How Will Our Project Be Similar/Different?

Our project will be similar in that we'll consider the data points they used (and intuition gained from the interviews), but we'll be using a different model. We'll predict a probability of critical violation using Logistic Regression.

Additionally, we may include some of these data sources found at the open data portal:

- Building permits issued
- Graffiti
- Affordable rent locations

- Vacant building locations
- Water temperature
- West nile virus test results
- Altering the weather forecast dates to include the traditional 3,7, and 10 day
- Energy consumption
- Street sweeping
- Building code violations