JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

OCEANIC, ATMOSPHERIC, & REMOTE SENSING SCIENCES GROUP

FPS-T-17-0045
2 February 2017
X9REDATH

To: Distribution

From: Christopher R. Ratto

Subject: Analysis of Syracuse Health Code Inspection Data

Abstract:

Many state, county, and municipal governments across the country are adopting an *open data* philosophy, in hopes that it will support innovation in how they serve the public. A recent study commissioned by the city of Chicago developed a data-driven approach to scheduling health code inspections of food service establishments based on risk. JHU/APL has been tasked by the JHU Center for Government Excellence (GovEx) to replicate the Chicago study on data from other cities, one of which is Syracuse, NY. Health inspection data was acquired from the New York State Department of Health and a similar predictive modeling approach to Chicago was followed. This memorandum describes the data, risk models, and experimental results for the Syracuse study.

Distribution List:
M. A. Baker
A. M. Bargar
C. J. Costello
J. P. Howard
A. J. Llorens
B. L. Paulhamus
A. G. Perrone
FPS Archives
KTF Files

1 Introduction

Many state, county, and municipal governments across the country are adopting an *open data* philosophy, in which government information of potential interest to the public is made freely available to promote government transparency and accountability, as well as develop a more locally-engaged citizenry. It is also believed that open data can support new and innovative ways to solve the everyday problems addressed by local government. One of these problems is scheduling health code inspections for food service establishments (FSEs), such as restaurants, catering companies, school cafeterias, food stands, etc. A recent study commissioned by the city of Chicago determined that a data-driven approach to scheduling inspections based on predicted risk of critical health code violations would result in an inspection schedule that found violations 7.44 days earlier, on average [1]. JHU/APL has been tasked by the JHU Center for Government Excellence (GovEx) to replicate the Chicago study on data from other cities, one of which is Syracuse, NY. Health inspection data was acquired from the New York State (NYS) Department of Health with the goal of developing a similar model as proposed in [1]. This memorandum describes the data, modeling approach, and experimental results for the Syracuse study.

2 Code

All code has been committed to `restaurant_inspections/Syracuse/code` in the GitHub repository located at `https://github.com/iscoe/restaurant_inspections/`. All analysis was performed using the open-source R programming language and code will be made available so that stakeholders may reproduce our results. The file `formatDataSet.R` was used to prepare the data for analysis. The file `model_analysis.Rmd` was used to perform the analysis and generate figures that are presented in this memorandum.

3 Data Set

3.1 Source of Data

Enforcement of health codes at FSEs in Syracuse, NY is the responsibility of the Onondaga County Health Department, and all data acquired during health code inspections are catalogued by the NYS Department of Health [2]. The data set was composed of all food service inspections performed in the Syracuse metropolitan area[1] between 13 October 2005 and 10 October 2016.

[1]"Syracuse metropolitan area" is defined as the following ZIP codes: 13066, 13202, 13203, 13204, 13205, 13206, 13207, 13208, 13209, 13210, 13211, 13212, 13214, 13215, 13217, 13219, 13224, 13244, and 13290.

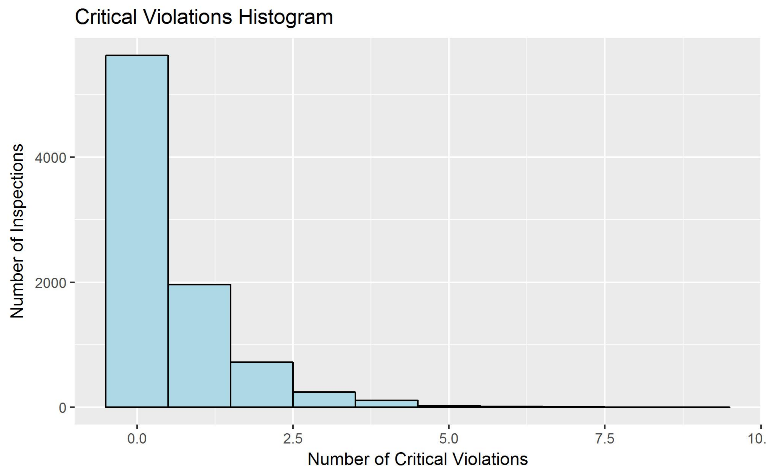

Figure 1: Histogram of the number of critical violations found per inspection.

3.2 Violation Types

Table 1 lists the categories of violations that are considered to be critical and non-critical in New York State [3]. When the county cites an FSE with a critical violation, the FSE must correct it immediately. If the violation is not immediately addressed, the county can force the FSE to close in order to remediate the violation since it is considered to be a public health hazard. When the county cites a FSE with a non-critical violation, it must be remediated within a time frame set by the inspector. The FSE may continue to operate during this time frame, since the violation is not considered to be evidence of an imminent public health hazard. It is surprising to see that several violations considered to be non-critical in NYS might be critical in other jurisdictions, e.g. "inadequate insect/rodent control."

3.3 Food Service Establishments

The data set was originally provided as a table in which each row corresponded to a single violation reported for a single FSE during a single inspection. There were 35,558 rows in this table. To match the Chicago data set, the table was consolidated so that each row summarized a single inspection of a single FSE. The total number of inspections was $N = 8,698$. Histograms of the number of critical and non-critical violations found per inspection are shown in Figures 1 and 2, respectively. More than half of the inspections, about 5,500, yielded no critical violations, and no inspections resulted in more than 9 critical violations. Nearly every inspection resulted in at least one non-critical violation.

Table 1: New York State Health Code Violations

Critical Violations

Category	*Description*	*# Violations*
1	Foods adulterated or received from unapproved sources	8
2	Foods not protected from contamination, temperatures not measured	5
3	Foods not protected from contamination by workers	3
4	Foods not protected from contamination by other sources	3
5	Improper cooling and refrigerated storage of potentially hazardous foods	5
6	Improper hot holding of potentially hazardous foods	2
7	Inadequate cooking and reheating of potentially hazardous foods	8

Non-Critical Violations

Category	*Description*	*# Violations*
8	Food not protected in general	7
9	Poor hygiene and activities of food workers	4
10	Poor sanitary design, construction, installation of equipment and utensils	2
11	Improper cleaning, washing, and sanitizing of equipment and utensils	4
12	Improper sanitary facilities and controls	5
13	Improper garbage and rubbish disposal	2
14	Inadequate insect/rodent control	3
15	Improper construction and maintenance of physical facilities	4
16	Miscellaneous, economic violation, choking poster, training	1

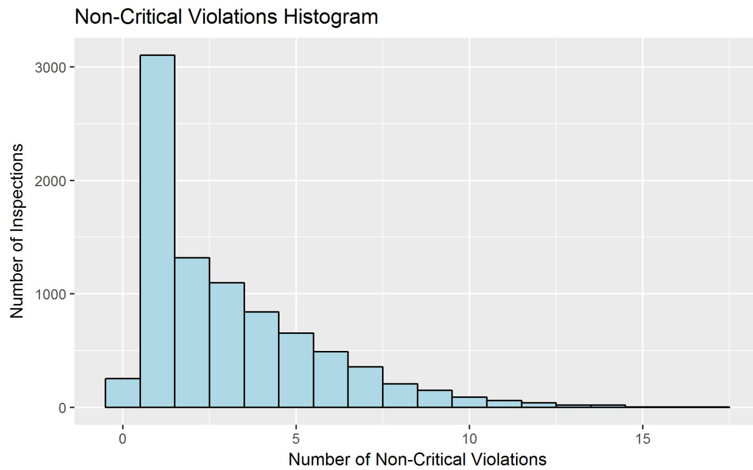

Figure 2: Histogram of the number of non-critical violations found per inspection.

3.4 Violations vs. Inspection Type

Of the 8,698 inspections, 7,557 were regular inspections and 1,141 were re-inspections, presumably in response to a violation that was found by the inspector. Figures 3 and 4 show how the number of critical and non-critical violations are distributed according to inspection type. As one should expect, the number of violations (critical or non-critical) tends to be lower for re-inspections, since the re-inspection is being performed to verify that specific violations have been remediated. The difference was determined to be statistically significant ($p < 0.01$) via a t-test for both critical and non-critical violations.

3.5 Violations vs. FSE Category

Each FSE in the data set (687 FSEs total) was assigned a “Food Service Description” by the health department, which are essentially categories of FSEs. The numbers of inspections (excluding re-inspections) per FSE category are listed in Table 2. The acronym SED stands for “State Education Department” and “SOFA” stands for “State Office For the Aging.” Upon first glance, several of the smaller categories are similar, e.g. “SSED Food Preparation Site”/ “SED Satellite Feeding Site” / “SED Satellite/Preparation Site,” and “SOFA Prep Site” / “SOFA Satellite Site.” According to NYS, a “Satellite” site is one where food is served but not prepared. a “Prep” site is where food is prepared but not served, and a “Sat/Prep Site” is where food is both served and prepared. The category “Food Service Establishment” appears to be a miscellaneous category, which includes private social clubs, concession stands at stadiums, churches, theaters, soup kitchens, apartment complexes, community centers, and what appear to be a few restaurants (including dining facilities on the Syracuse University campus).

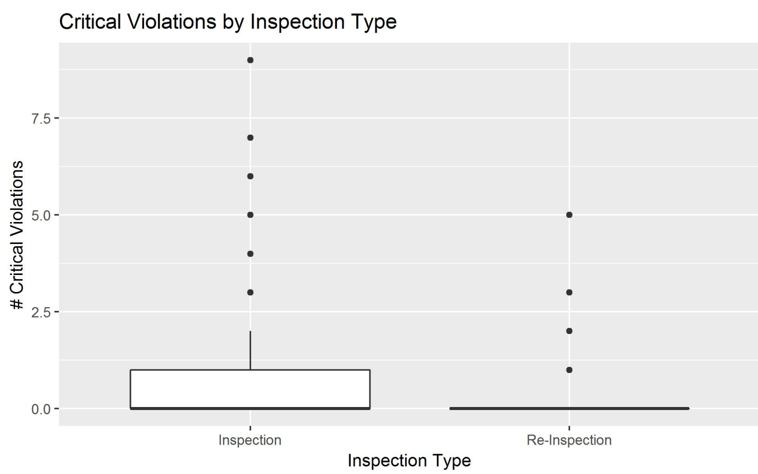

Figure 3: Boxplot showing the distribution of number of critical violations per inspection as a function of inspection type. The upper and lower limits of the box correspond to the first and third quartiles. The whiskers extend from these quartiles to the lowest/highest values within 1.5 times the quartile. The dots correspond to outliers beyond these limits.

Figure 4: Boxplot showing the distribution of number of non-critical violations per inspection as a function of inspection type. The upper and lower limits of the box correspond to the first and third quartiles. The whiskers extend from these quartiles to the lowest/highest values within 1.5 times the quartile. The dots correspond to outliers beyond these limits.

Table 2: Food Service Descriptions

Type	# Establishments
Bakery	9
Catering Operation	5
Commissary	7
Food Service Establishment	221
Ice Cream Store	6
Restaurant	304
School K-12 Food Service	54
SED Food Preparation Site	3
SED Satellite Feeding Site	37
SOFA Prep Site	14
SOFA Satellite Site	5
Tavern	22
Total	687

Figures 5 and 6 show how the distribution of critical and non-critical violations varies with facility type. An interesting result is that the State-managed facilities tend to have few critical and non-critical violations. Restaurants, on the other hand, have the most critical violations on average. Therefore, separate analysis will be performed on restaurants alone as well as restaurants combined with the other facility types.

4 Methodology

4.1 Data Preparation

Following the example set by the Chicago study, we sought to develop a statistical model that would produce a binary prediction of whether an inspection would yield *at least one* critical violation. Since critical violations are considered to pose a public health risk, the ability to predict whether inspecting a particular FSE would yield at least one violation can allow health inspectors to prioritize higher-risk FSEs over lower-risk ones when scheduling inspections. Since the predictor is binary, its performance can be assessed using conventional metrics like the receiver operating characteristic (ROC) curve, and area under the curve (AUC). Furthermore, to compare to the Chicago study, performance was assessed using the number of days saved in finding critical health code violations by scheduling inspections according to the model's predictions.

In order to fairly evaluate the models, the data was divided into training and testing sets. The

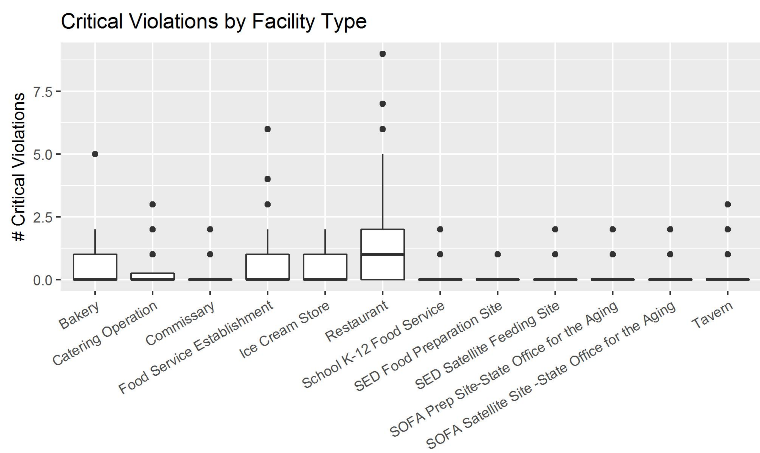

Figure 5: Boxplot showing the distribution of number of critical violations per regular inspection as a function of FES category. The upper and lower limits of the box correspond to the first and third quartiles. The whiskers extend from these quartiles to the lowest/highest values within 1.5 times the quartile. The dots correspond to outliers beyond these limits.

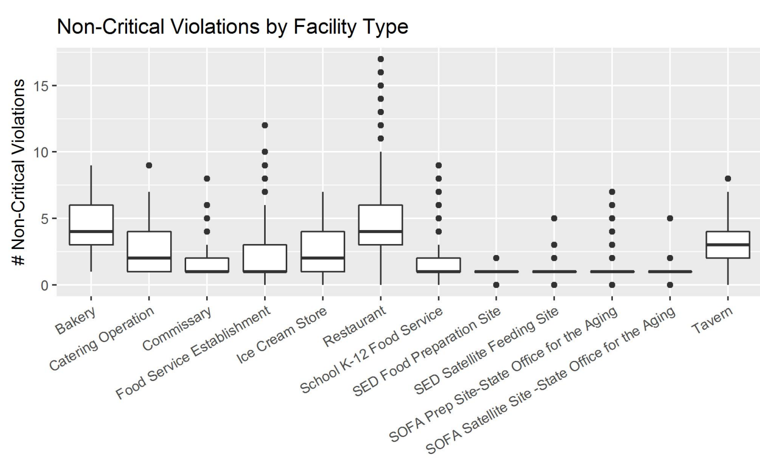

Figure 6: Boxplot showing the distribution of number of non-critical violations per regular inspection as a function of FES category. The upper and lower limits of the box correspond to the first and third quartiles. The whiskers extend from these quartiles to the lowest/highest values within 1.5 times the quartile. The dots correspond to outliers beyond these limits.

training data consisted of all inspections between 13 October 2005 and 31 December 2015. In a similar fashion as the Chicago study, a six month buffer period between training and test sets was used to safeguard against the Hawthorne effect (i.e., data influenced by subjects' knowledge of being observed). The testing data consisted of all inspections from 1 July 2016 to 10 October 2016. Only regular inspections were considered for training and testing (re-inspections were removed). Two analyses were performed: one that included all FSE categories ($N_{train} = 6,980$, $N_{test} = 240$), and one that only considered restaurants ($N_{train} = 3,093$, $N_{test} = 101$).

4.2 Baseline Model

A convenient baseline approach would be to base the prediction off of two indicators: the numbers of critical and non-critical violations from the previous inspection. The basic idea behind this approach would be that restaurants with a history of health code violations will continue to exhibit similar behavior. The model was implemented using logistic regression via the `glmnet` package in R, and outputs the posterior probability of an inspection having at least one critical violation.

4.3 Alternative Models

A variety of additional features were considered for enhancing the baseline model. Several of these features were also chosen for the Chicago model. Features dependent on tertiary data sources (e.g. crime, weather, 311 requests) were also used in the Chicago study, but our own analysis indicated they had little predictive power and were not considered for the Syracuse study described here. The features that were included in this analysis are as follows:

- Has an alcohol license (boolean variable)
- ZIP code (categorical variable)
- Number of critical violations from previous inspection (discrete variable)
- Number of non-critical violations from previous inspection (discrete variable)
- Days until permit expires (discrete variable)
- Days since last inspection (discrete variable)
- Average number of critical violations per 5 nearest FSEs (continuous variable)
- Average number of non-critical violations per 5 nearest FSEs (continuous variable)

All categorical variables were expanded into 2^k discrete binary variables, where k is the number of categories.

Various classification algorithms were considered for learning the alternative model. A random forest [4] was considered as a fully nonparametric approach to the problem, and logistic regression was considered as a parametric approach. Various types of regularization were considered for safeguarding against overfitting the logistic model: l_1-norm (i.e. the *lasso* [5], l_2-norm (i.e., *ridge regression* [6]), and an equal combination of both l_1-norm and l_2-norm (i.e., the *elastic net* [7]). The tuning parameter λ for controlling the amount of l_1-norm or l_2-norm regularization was set via 10-fold cross-validation on the training set.

5 Performance

The performance of the models for predicting whether a regular FSE inspection would yield at least one critical violation was measured using the ROC curve. A ROC curve represents how the two performance metrics *true positive rate* and *false positive rate* vary as the threshold applied to the classifier output is swept from zero to one. Any prediction with a score about the threshold is assumed to be at risk of yielding a critical violation. The true positive rate represents the fraction of inspections predicted to yield critical violations that actually had critical violations, and false positive rate is the fraction of inspections predicted to yield critical violations that actually did not. The AUC is a summary score for the entire ROC curve, in which 0.50 represents chance accuracy and 1.0 represents perfect accuracy. Table 3 summarizes the AUC of each model evaluated on all FSEs and on restaurants only. Discussion of both versions of performance is given in the following subsections.

Another performance metric that may be of more relevance to local health departments is the number of days earlier (or later) in which critical violations were found, compared to a "business as usual" inspection scheduling. To compute this metric, the classifier output scores are sorted in descending order and then associated with the original dates of the test inspections. It is assumed that inspections would have occurred on the same dates, with the same number of inspections per day, but the order in which FSEs are inspected is determined by the data-driven model. The inspection dates for the new schedule are compared to the dates of the old schedule, and the average difference in inspection date is computed. A positive number indicates that FSEs with critical violations are being inspected earlier, and a negative number indicates that they are being inspected later. Sorting the list randomly would result in zero days saved on average. This is the metric that was used to score performance in the Chicago study, where it was found that FSEs with critical violations were inspected 7.44 days earlier, on average.

A notable result is that the AUC and number of days saved is higher when the predictive model is applied to all FSEs. For all FSEs, the best model yielded an AUC of 0.7749 and critical violations

Table 3: Area under ROC curve. Best performance is shown in bold.

Model	AUC (All FSEs)	AUC (Restaurants Only)
Baseline (Previous Inspection)	0.7334	**0.7020**
Random Forest	0.6845	0.5956
Logistic (No Regularization)	0.7627	0.5932
Logistic (l_1 regularization)	0.7744	0.6301
Logistic (l_2 regularization)	0.7680	0.6299
Logistic ($l_1 + l_2$ regularization)	**0.7749**	0.6322

Table 4: Average number of days in which critical violations are found earlier. Best performance is shown in bold.

Model	Avg. # Days (All FSEs)	Avg. # Days (Restaurants Only)
Baseline (Previous Inspection)	14.80	**4.64**
Random Forest	11.31	-0.07
Logistic (No Regularization)	17.15	-0.02
Logistic (l_1 regularization)	17.51	1.58
Logistic (l_2 regularization)	17.48	1.54
Logistic ($l_1 + l_2$ regularization)	**17.52**	1.80

found 17.52 days earlier, on average. Meanwhile, for restaurants only, the AUC was reduced to 0.7020 and average time savings to 4.64 days. The main reason for the difference in performance is that by and large, critical violations tend to only be found in restaurants. Therefore, a simple way to find critical violations earlier would be to inspect restaurants first, and all other FSEs later. Specific discussions of the individual models' performance are included in the following subsections.

5.1 All Food Service Establishments

The ROC curves for all models applied to regular inspections of all FSEs in the test set are shown in Figure 7. The random forest has the worst ROC curve, followed by the baseline model. The various logistic regression methods have similar ROC curves, though the version employing both types of regularization had the highest AUC (0.7749). If the best logistic model were to be used for scheduling inspections, critical violations would have been found 17.52 days earlier, on average. These results indicate that the logistic models are able to leverage the additional features to provide some additional information that is useful for identifying inspections likely to yield a critical violation.

Table 5 summarizes the regression coefficients obtained by the logistic model using l_1 and l_2 regularization. Since the model is sparse, only the features with non-zero weight are shown. As

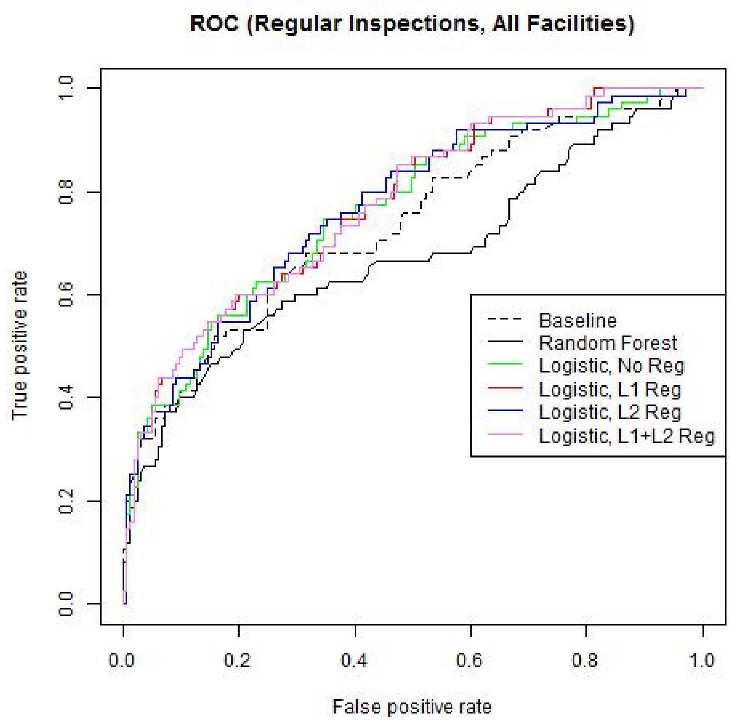

Figure 7: ROC curves for all models applied to regular inspections of all FSEs in the test set.

evidenced by the similar performance of the Baseline method, the number of critical violations in the previous inspection (`nCritical_prev`) has the highest weight (disregarding the intercept). The second-highest weight (in terms of absolute value), was the binary feature indicating whether the FSE was in the 13244 ZIP code (`zip13244`). This ZIP code includes parts of the Syracuse University campus, and the 558 inspections in that ZIP code include concession stands in the Carrier Dome and several on-campus dining facilities (neither of which are categorized as "restaurants" by the health department). Since this feature was given a negative weight, it indicates that FSEs in this ZIP code tend not to have many critical violations when inspected and may be considered low-priority for scheduling inspections.

Table 5: Non-zero logistic regression weights for all FSEs.

Feature	Weight
(Intercept)	0.2340
`nCritical_prev`	0.0719
`zip13244`	-0.0624
`alcLicenseTRUE`	0.0433
`avg_neighbor_num_critical`	0.0361
`nNonCritical_prev`	0.0307
`avg_neighbor_num_non_critical`	0.0058

5.2 Restaurants Only

The ROC curves for all models applied to regular inspections of only restaurants are shown in Figure 8. The ROC curves shown here are markedly worse than those shown in Figure 7, probably because including non-restaurants biases the prior probability of an inspection not having any critical violations; recall that most inspections of non-restaurants had no critical violations, as shown in Figure 5. Of the various models, the Baseline yielded the best ROC curve with an AUC of 0.7020. When scored based on the number of days earlier in which critical violations are found, rescheduling the inspections using the Baseline model would result in critical violations being found 4.64 days earlier, on average.

Th superior performance of the Baseline model indicates that additional features do *not* provide additional useful information when only predicting whether restaurants' inspections yield critical violations. The combination of a small data set and additional noisy features increases the risk of overfitting. The logistic models with regularization perform better than the random forest and regularization-free logistic model because regularization allows the model to ignore noisy features. Note that the performance of the logistic models is best (slightly) when both l_1 and l_2 regularization is employed.

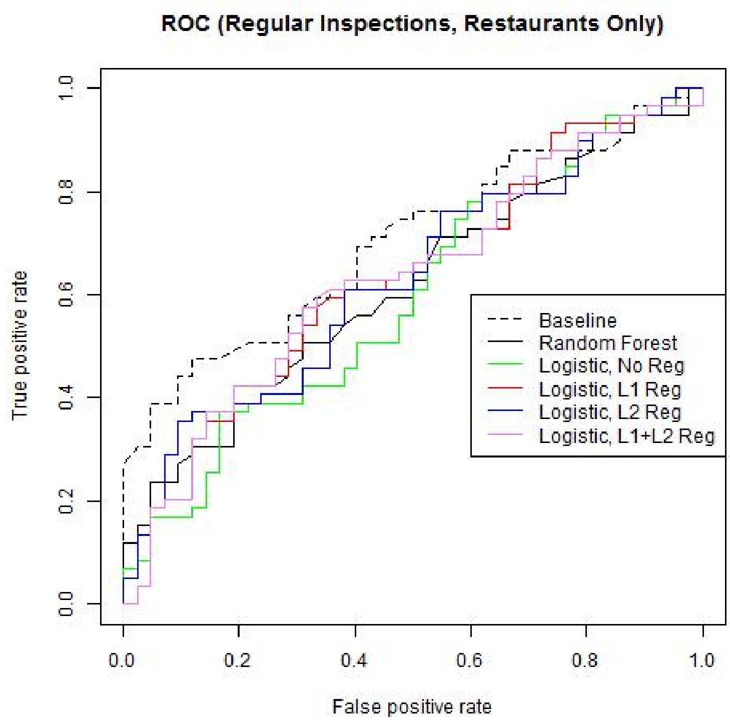

Figure 8: ROC curves for all models applied to regular inspections of only restaurants in the test set.

Table 6 summarizes the non-zero regression weights obtained by the logistic model using l_1 and l_2 regularization. It comes to no surprise that the `nCritical_prev` again has the highest weight, since the Baseline model performed so well. The second-highest weight was the binary feature indicating whether the FSE held a license for alcohol sales. Finally, the indicators for ZIP codes 13214 (`zip13214`) and 13219 (`zip13219`) were the only ZIP indicators with any weight. These ZIP codes correspond to the suburbs of DeWitt, Westvale, and Fairmount. Both of the weights were negative, indicating an inverse correlation with the presence of critical violations.

Table 6: Non-zero logistic regression weights for restaurants only.

Feature	Weight
(Intercept)	0.5292
`nCritical_prev`	0.0465
`alcLicenseTRUE`	0.0327
`zip13219`	-0.0144
`zip13214`	-0.0137
`nNonCritical_prev`	0.0002

6 Conclusions

This memorandum described analyses performed on Health Department inspection data for FSEs in the greater Syracuse, NY area. Following procedures suggested by a recent Chicago study, statistical models were developed for predicting whether a regular inspection (not a re-inspection) would yield at least one critical violation. Models were developed based on all categories of FSE as well as for restaurants only. Results suggest that critical violations can be predicted with above-chance accuracy (AUC $\approx$ 0.77 for all FSEs, AUC $\approx$ 0.70 for restaurants) and would result in critical health code violations being discovered several days earlier (17.5 days by rescheduling all FSE inspections, 4.6 days by only rescheduling restaurant inspections), significantly reducing the risk of food-borne illness. Incorporating features beyond the number of violations recorded in the previous inspection may provide additional predictive information, though it was observed to have a positive effect when re-scheduling inspections of all FSEs but a negative effect when the analysis was only constrained to restaurants. Since all of the information to derive these models were provided by the NYS Department of Health, and no third-party data was used to supplement it, the models proposed in this memo would be low-cost alternatives to the current state of practice of scheduling FSE health code inspections for FSEs.

References

[1] T. Schenk, G. Leynes, A. Solanki, S. Collins, G. Smart, B. Albright, and D. Crippin, "Forecasting restaurants with critical violations in chicago," City of Chicago - Chicago Open Source, Tech. Rep. [Online]. Available: https://chicago.github.io/food-inspections-evaluation/

[2] Food service establishment inspection data. [Online]. Available: https://health.data.ny.gov/Health/Food-Service-Establishment-Last-Inspection/cnih-y5dw

[3] T. DiNapoli, "Consumer protection: Restaurant inspections," Office of the New York State Comptroller, Tech. Rep. [Online]. Available: https://www.osc.state.ny.us/localgov/audits/swr/2012/restinspection/global.pdf

[4] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[6] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[7] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.