Data collection and modelling for food business compliance indicators

FS101214

17th May 2019

Author: Gordon Squire

decisionLab

smart models from creative thinkers

dL

Contents

1. Executive Summary

This project contributes to the Food Standards Agency's (FSA's) Regulating Our Future (ROF) programme. Within the current official control system, a Local Authority (LA) is required to inspect a newly registered establishment for compliance with food hygiene law within 28 days. Limited prioritisation is undertaken of which establishments need to be inspected first until after the first inspection.

The objectives for this project were firstly to collect a large data set relating to the business activities of a range of food establishments and secondly to develop a concept model for a risk engine that will segment new food establishments so that the most effective and proportionate initial intervention can be determined.

Data for 8,700 establishments across 13 Local Authorities was collected. The establishment data comes from data collected by LAs as part of the establishment inspection process. 4 Local Authority datasets were from app-based inspections, the others were from data manually captured from paper-based inspection forms.

Hygiene compliance prediction models were developed using this data. The models demonstrate predictive power substantially above the no predictive power benchmark model[1]. Figure 1 shows an example of the model's predictive results. The model is good at separating broadly compliant and non-broadly compliant establishments. 70% of the model's predictions are correct (see top left-hand cell and bottom right-hand cell) and the model correctly identifies 9 out of 10 (90%) non-compliant establishments (see bottom cells). This can be compared to 11% of predictions would be correct if all establishments are presumed to fail.

[1] Average precision of 48% compared to 11% for the benchmark uniform predictor.

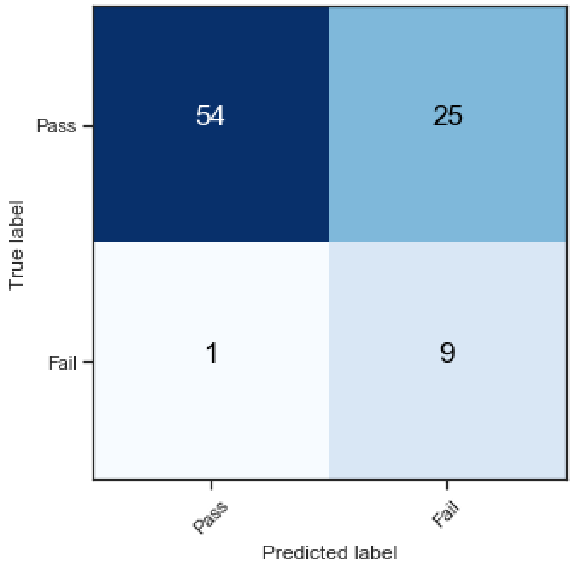

Figure 1 City of London results, number of establishments predicted to fail versus actual outcome. Fail indicates establishment is non-broadly compliant. Results are for low risk threshold.

Figure 2 shows how the results change if a higher risk threshold is selected. The proportion of correct predictions increases to 93% but only 4 of 10 non-compliant establishments are identified (see bottom 2 cells). The model provides a clear capability to prioritise inspections for higher risk establishments.

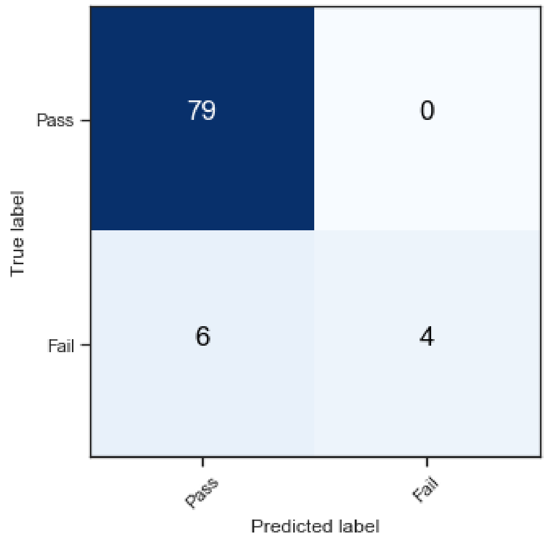

Figure 2 City of London results, number of establishments predicted to fail versus actual outcome. Fail indicates establishment is non-broadly compliant. Results are for high risk threshold.

The models developed are highly interpretable and give a clear explanation for why predictions are made. The model finds that the strongest predictors of non-compliance are processes and procedures around the food safety management system, training, personal hygiene, supplier assurance, whether the establishment is a take-away, and approaches to cleaning.

The project has demonstrated the potential for a food establishment risk segmentation model allowing pre-emptive identification of risk and improved intervention resourcing. It has the potential to enable improved regulatory controls that are fit for the future.

2. Introduction

2.1. Background

This project contributes to the Food Standards Agency's (FSA's) Regulating Our Future (ROF) programme. The FSA is improving the way regulatory controls are delivered by developing a modern, resilient system for ensuring that businesses meet their responsibilities. The new system will need to be proportionate to the type of food business and associated level of risk. It will take account of all available sources of information and be flexible enough to keep pace with technological change in the food industry, and able to adapt to the changing environment.

Under the current regulatory regime, all newly registered food businesses are subject to an initial physical inspection, regardless of the food safety risk they present. This 'one size fits all' approach is not sustainable or proportionate. The FSA is working closely with food business operators and local authorities to obtain behavioural insights to help inform a sustainable approach to food safety regulation, one that brings about business behaviour change to benefit consumers.

With the current official control system, all new food establishments need to register with, or be approved by, their Local Authority (LA). The LA is then required to inspect the establishment for compliance with food hygiene law within 28 days. Limited prioritisation is undertaken of which establishments need to be inspected first until after the first inspection, after which the inspection frequency is set using a system prescribed within the Food Law Code of Practice

2.2. Objective

The purpose of this project was to collect a detailed level of data on the food related activities of a large sample of food establishments and to use this data to develop models to forecast how compliant establishments with particular characteristics are likely to be with food safety law. The focus for this work was to establish the framework for prediction models for new establishments who have yet to have a food safety inspection and who do not have any enforcement history. The model(s) are intended to inform a new food business segmentation model which will determine, based on risk, how food businesses are regulated in the future.

The project has the following two core elements:

- Data collection
- Analysis of data and development of prediction models

Data collection

The first objective of the project was to gather a large data set relating to the business activities of a range of food establishments, including assessments of legal compliance with food law by the relevant inspecting Local Authority (LA). This data is routinely gathered by LAs through their official control work, such as inspections. The data is held by LAs, but not in a common or standard format. For the majority of local authorities this detailed level of information was expected to be held in paper reports, scanned documents or on databases. The data required included fields relating to:

- Establishment (Name, address, premises type, opening hours, seasonality, cuisine type);
- Food business operator (FBO) (franchise, sales activity);
- Food activities (food handling activities, food types handled, number of staff, method of processing, Food safety management system (FSMS) details, assurance, training, water supply, waste disposal contract, pest control contract);
- Import/Export of food;
- Food hygiene inspection results;
- Food standards inspection results.

The full list is available in Appendix E.

Development of model

The second objective of the project was to use this data to develop a concept model for a risk engine that will segment new food establishments so that the most effective and proportionate initial intervention can be determined. The requirement was to predict the following outcomes in order of priority:

- Seriously Non-Compliant (Food Hygiene Rating Scheme[2] (FHRS) equivalent ratings of 0, 1)
- Not Broadly Compliant (FHRS equivalent rating of 0, 1, 2)
- Fully compliant premises (FHRS equivalent rating of 5)

And to develop separate sub models such as for restaurants & catering.

3. Data Collection

Approach

The principal aims of the data collection exercise were to: collect data for a substantial number of establishments; to minimise the workload for LAs; and to collect a representative sample of data. Our approach to data collection was designed to meet these aims most effectively. We began by sending out a survey to all the LAs in England, Wales and Northern Ireland, to establish the content and format of their data. The survey responses were used to prioritise contact with local authorities with data in the most extractable formats (in order to maximise data volume), whilst ensuring a representative sample selection. LAs were contacted to find willing participants. Selected LAs then collected and sent over their digital data, spreadsheets

[2] The Food Hygiene Rating Scheme is a national hygiene rating scheme for food establishments, establishments are inspected by local authorities and given a rating. The FSA collates these ratings into a nation-wide data set and web API.

or collections of documents. Paper based inspection reports were sent to our data entry subcontractor, and these forms were manually typed up into spreadsheets. This data was then cleaned and formatted, ready for modelling.

LA survey

A digital survey was designed and sent out to all the LAs in England, Wales and Northern Ireland. 76 Local Authorities responded to the survey. The survey identified that the majority of hygiene inspection data is captured by hand, with 89% of LAs inspection reports being handwritten. 22% of local authorities store this data in paper records. 18% of LAs reported that at least some of their establishment data was held in database or spreadsheet format. The findings from the survey were used to prioritise data collection from LAs reporting to have a high proportion of digital data stored in the most extractable formats. The order of priority was: databases/spreadsheets > digital/scanned documents > paper records.

LA data collection

We contacted LAs in order to find those willing and able to participate and to find out more about the establishment data they hold. In general, the information that LAs hold on establishments includes: registration data, inspection reports/notes/aide-memoires, post inspection communications sent to an establishment, data for the Local Authority Enforcement Monitoring System (LAEMS), and data for the FSA's FHRS database, in addition to extra information for certain establishments. Beyond the data that LAs are mandated to collect we found that there is not a common approach to data fields collected (what information LAs are collecting about establishments), data formats (how are they capturing the data), or data storage approaches (where is the data being stored).

In almost all cases the bulk of the information that we were seeking to collect is contained within inspection reports/completed aide memoirs filled out during inspections. Registration data, provided by the food establishment at the point of registration, is generally limited to high level details of the business such as business name, address, type and FBO details. A summary of our findings from contacting LAs is given in Appendix A.

A list of LAs which kindly supplied us data is given in Table *1*, along with the data format and the number of establishment records provided.

Table 1 Summary of data collected

Local Authority	Data format	Number of records
Charnwood	Tabular data from inspection app	1265
Chiltern	Tabular data from inspection app	449
City of London	Tabular data from inspection app	462
Dartford	Tabular data from inspection app	131
Neath Port Talbot	Typed document	404
Croydon	Scanned document	487
Greenwich	Scanned document	702
Manchester City	Scanned document	1664

Sedgemoor	Scanned document	585
South Kesteven	Scanned document	1028
Stafford	Scanned document	530
Tunbridge Wells	Scanned document	585
Plymouth	Digital registration data	473
Total		**8765**

'Digital registration data' refers to Plymouth, which has a web-based registration system that collects a fuller than normal set of information about establishments (including: whether a kitchen is shared, number of staff, what high-risk processes are undertaken) and could be seen as a forerunner to the self-registration system envisioned by this research. We collected this data from Plymouth to investigate the predictive power of this self-registration data.
In total, we collected 8,765 establishment records. By prioritising the most extractable data and using efficient data entry on digital forms we were able to maximise the volume of data collected and minimise the burden on Local Authorities.

Ensuring Data Quality
The quality of data input from hand written records was checked in several stages: the data entry subcontractor has an internal quality assurance process, which includes sampling to check quality, and review by quality assurance manager. Further data checking was undertaken including manual review of the data, e.g. to check that fields are as expected, there are no clear errors or duplicates. Further cleaning and checking of the data was undertaken as part of the data cleaning process described in section 4.6. This was conducted using scripts written in Python, ensuring reproducibility and traceability.

Stratified Sample
Given the objective of maximising the volume of data collected and the constraint of LA cooperation, the data collected gives a reasonable representation of LAs in England and Wales with regards to LA type, geography, and rural/urban split. Full details are available in Appendix B.

4. Modelling

4.1. Overview

The aim for our models was to predict the hygiene compliance of a food establishment, prior to inspection. The models use data about an establishment such as the number of staff or whether it has a pest control contract in order to predict its compliance. The data fields the model uses to make its predictions are known as 'explanatory variables' or 'features'.
We investigated two approaches to modelling this problem:

1) Binary Classification: The objective is to make a binary prediction e.g. a business is broadly compliant or not broadly compliant. The output is a pass/fail prediction (a probability of failure can also be obtained).
2) Multi-class Classification: The objective is to predict the specific hygiene rating value. The output is a number 0,1,2,3,4 or 5.

The models are algorithms that aim to identify statistical relationships between the features (in this case the data about establishments) and the variable being predicted (in this case the hygiene compliance). The dataset is split into two parts, a training set and a testing set. The model finds the best fit on the training set and is then applied to the unseen test data to evaluate its performance on unseen data.

4.2. Evaluating performance

How do we evaluate how good a model is? This depends on whether the model is a binary classification model or a multi-class model.

Binary classification

We can consider that what a binary classification model is doing is to assign a probability of failure (e.g. being not broadly compliant) to an establishment. If the probability is over a certain threshold, the establishment is predicted to fail. This is depicted in Figure 3.

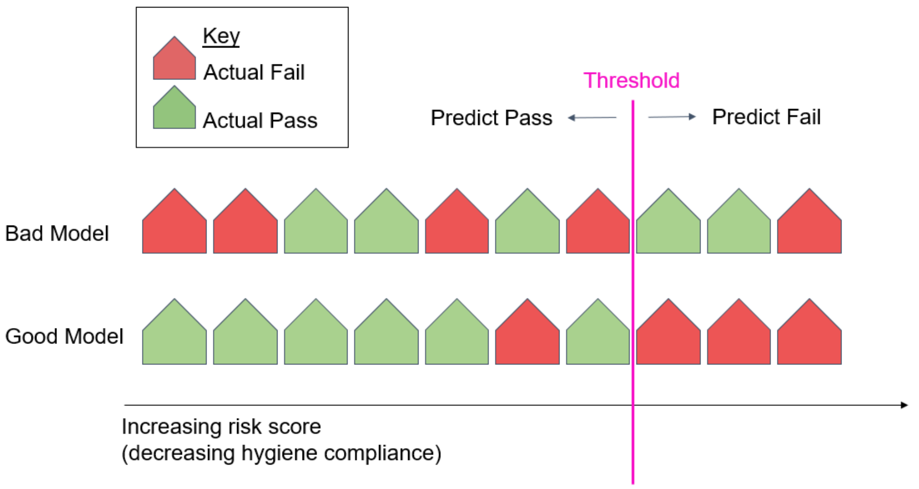

Figure 3 Depiction of binary model scoring. Each 'building' icon represents an establishment. Establishments are assigned a probability of failure. If the probability is above the threshold the establishment is predicted to fail.

The figure depicts the predictions of two models, a good and a bad model. The bad model predicts 4 establishments to pass that actually fail and predicts 2 establishments to fail that actually pass. The good model only makes one incorrect prediction. The better the model, the more it is able to assign failing establishments a high probability of failure. A perfect model would align all establishments that actually fail to the right of the risk threshold line.

Establishments with a risk score above the risk threshold are predicted as a fail, those that are below are predicted as a pass. Some predictions may be correct ("True") or incorrect ("False"). Figure 4 depicts a 'confusion matrix', which is a method for displaying these results in a table.

Actual result	Predicted result Pass[3]	Predicted result Fail
Pass	True Pass	False Pass
Fail	False Fail	True Fail

Figure 4 A confusion matrix, a method for evaluating the performance of binary classification models

[3] Typically 'positive' and 'negative' are used in machine learning literature. Pass and fail are used here for clarity of communication. For our purposes a positive is a fail i.e. the model makes a 'positive' prediction if it expects the establishment to fail.

Figure 5 displays the confusion matrices for the results in Figure 3.

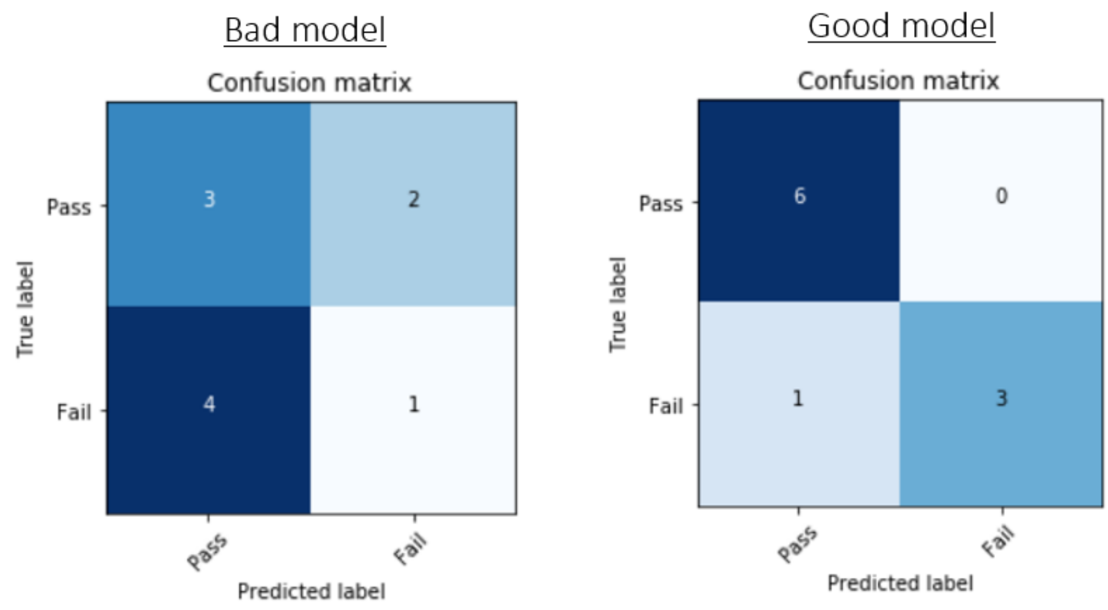

Figure 5 Confusion matrices for the models depicted in Figure 1

The good model effectively separates establishments into pass and fail categories: 9 predictions are correct (i.e. 6 establishments were predicted to pass and did pass (top left cell), 3 establishments were predicted to fail and did fail (bottom right cell)) and one is a prediction of a pass when the establishment is actually a fail (bottom left cell). (i.e. the model incorrectly predicted that the establishment would pass).

From the confusion matrix we can calculate the following metrics:

- **Accuracy**: The proportion of predictions which are correct. For the 'bad model' in Figure 5 Accuracy = (3+1)/(3+2+4+1) = 0.4 = 40%
- **Precision**: The proportion of 'fail' predictions that were correct. For the 'bad model' in Figure 5 Precision = 1/(1+2) = 0.33 = 33%.
  Recall: The proportion of failing establishments that were predicted to fail. For the 'bad model' in Figure 5 Recall = 1/(4+1) = 0.2 = 20%.

It should be clear that if the risk threshold is changed, the predictions of the model, and the resulting confusion matrix will change. The effect of changing the risk threshold for the bad model depicted in Figure 3 is depicted Figure 6.

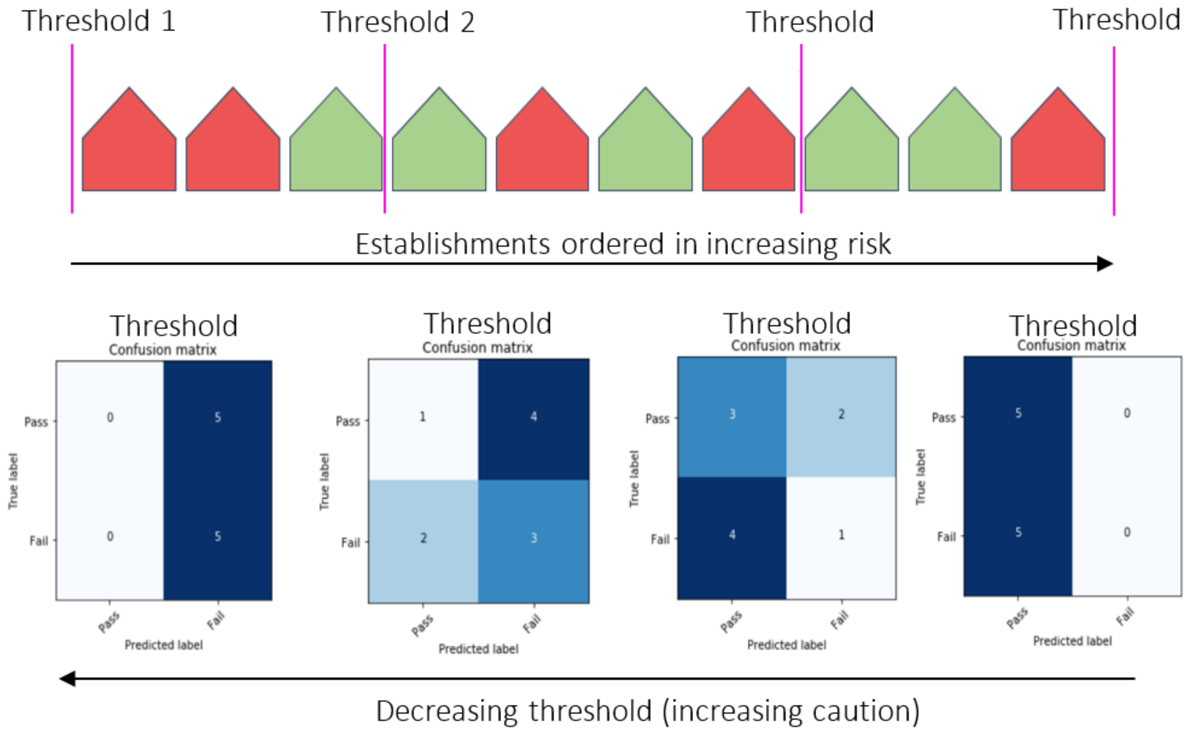

Figure 6 How results change with different thresholds. Top image corresponds to the bad model in Figure 4. Establishment is predicted to fail if it is to the right of the threshold, pass otherwise. A confusion matrix is plotted for each threshold.

In the context of this project, predicting an establishment to pass when in reality it would fail a hygiene inspection is not desirable, and the FSA and LAs are likely to prefer more cautious predictions. Mistakes in the lower left are worse than those in the top right of the confusion matrix[4]. Our results aim to recognise this, and models and thresholds are selected with this in mind.

A confusion matrix is a useful way to display a single set of results for a single risk threshold, however, rather than looking at many confusion matrices for all models, we need metrics that summarise the overall performance of the model to enable straightforward comparison between models.

One of the main measures we use is 'average precision'. Average precision is a measure of the model performance across different thresholds, it describes how well the model is able to assign low non-compliance risk to well performing establishments and high non-compliance risk to failing establishments. Essentially, it is the precision averaged across all risk thresholds. E.g. if the average precision is 0.4, we would expect that 40% of the establishments predicted to fail do indeed fail. If the risk threshold were lowered the value may be lower than this, if it were raised it may be higher, but on average the precision is 40%.

[4] In traditional machine learning terms: recall is preferable to precision.

Formal definition of average precision

Formally, average precision is defined as:

$$\text{AP} = \sum_n (R_n - R_{n-1}) P_n$$

Where R_n and P_n are the recall and precision at the nth threshold. All thresholds that change the prediction for any establishment are included in the set of ns. For example, in Figure there would be a threshold (pink line) between each establishment.

Multi-class classification

A confusion matrix can also be used to show the results of the multi-class model, where we are looking to predict the exact FHRS class of an establishment. An example is given in 7. It shows the predicted hygiene ratings made by the model compared to the true hygiene ratings.

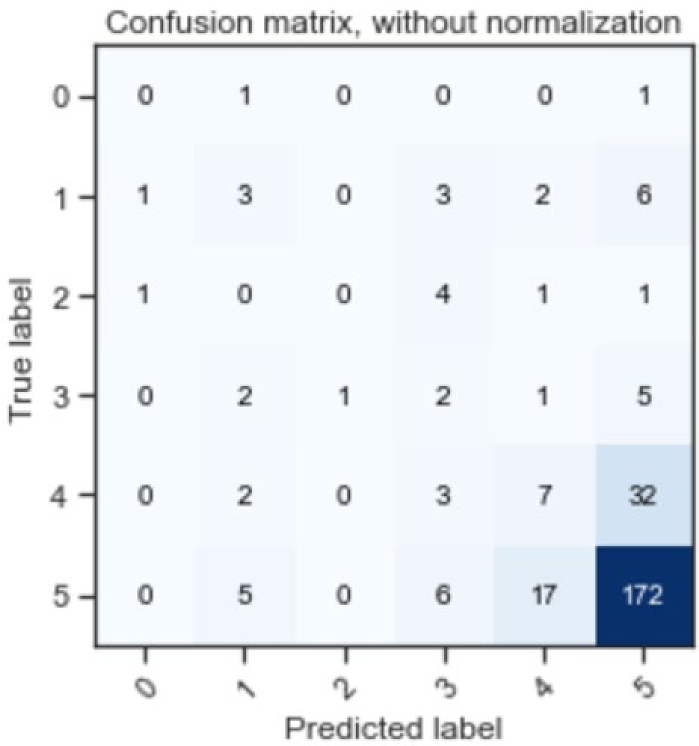

Figure 7 - Exampe multi-class confusion matrix

We use several measures to evaluate the overall performance of a multi-class model and compare between them, including mean absolute error (MAE). This is a measure of how far away results are on average from the correct value. The absolute value of the prediction error is calculated for each observation and the mean found. For example, the prediction in the top right corner of Figure 7 would have an absolute error of 5 (because the prediction is 5 and the true score is 0). A lower MAE score is better.

4.3. Models

The details of the various models tested are given below.

Logistic regression

Logistic regression is similar to linear regression (which the reader may be familiar with – e.g. fitting a straight line to a scatter plot). The model prediction is created from a weighted sum of the features of the form

$$z = w_0 + w_1 x_1 + \cdots + w_i x_i$$

where x_i is a feature (e.g. whether the establishment holds hot food for service) and w_i is the weight for that feature, this linear sum of weighted features is then transformed (by a logistic function, depicted in Figure 8) so that predictions are converted to be within the range 0 to 1.

$$y = f(w_0 + w_1 x_1 + \cdots + w_i x_i)$$

where

$$f(x) = \frac{1}{1 + e^{-x}}$$

The output value of the model (y) can be considered to be the probability of failure. A threshold is then applied, above which an establishment is predicted to fail.

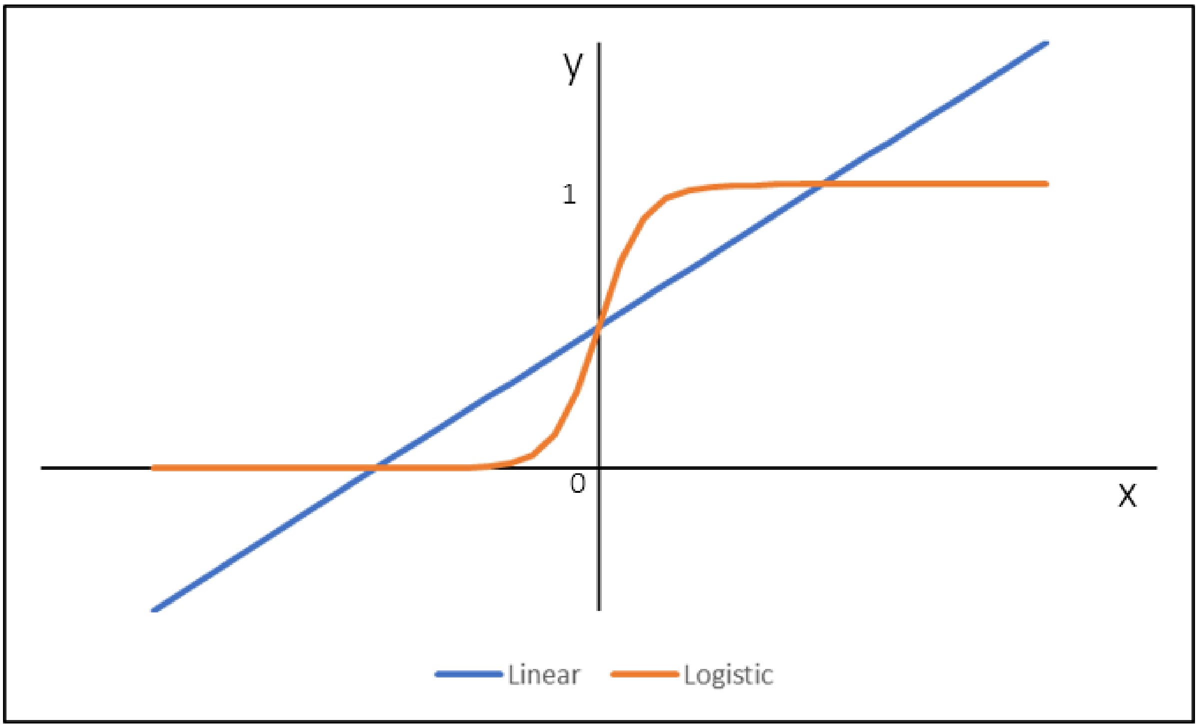

Figure 8 Illustrating the logistic transformation

The model is highly interpretable: feature weights specify how much that feature contributes to the overall prediction.

Decision trees

Decision trees work in a flow-diagram type fashion. Initially, the algorithm finds the feature which creates the best separation between each class (pass or fail). The data set is split into two groups and the process is repeated, until a stopping criterion is reached (e.g. maximum tree depth is reached). An illustration is given in Figure 9.

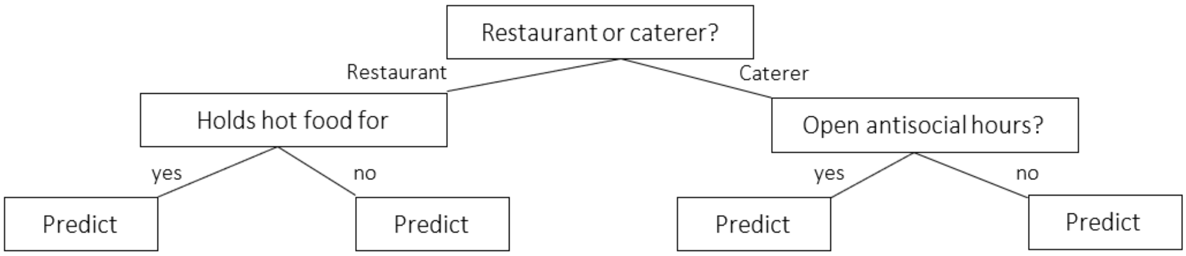

Figure 9 Illustration of a decision tree model

Decision trees are also highly interpretable, the tree can be understood in the same fashion as a flow diagram. In the example in Figure 9, if an establishment is a restaurant and it holds hot food for service it will be predicted to fail.

Random Forests

A random forest is an ensemble of decision trees. Varied trees are created by allowing a random subset of features at each split. The final prediction is an average of the predictions of all trees. While the logic of the decision is fairly easy to understand, interpreting a prediction can be quite involved and is as such much less interpretable than the previous two algorithms. However, performance is generally better than for individual decision trees.

XGBoost

XGBoost is a more complex algorithm, like random forests it is an ensemble approach. Trees are created, with successive trees predicting the errors of those previously. These are added together to make the final prediction. This is a technique called gradient boosting. XGBoost models are harder to interpret than the previous models but generally have very good performance.

Multi-class logistic regression

Logistic regression can be applied to multi-class problems by treating each class as a one-vs-all problem and selecting the class with the highest probability. In our case, our models are predicting an establishment to be in one of 6 classes, i.e. to have a rating of 0,1,2,3,4 or 5. The

6 binary models are trained to predict if the establishment is in one of these classes or not, the establishment is then predicted to belong to the class with the highest probability.

Ordinal logistic regression

A disadvantage of one-vs-all multi-class logistic regression is that it doesn't capture the order of classes i.e. the fact that predicting a score of 5 when the true score is 0 is worse than predicting a score of 1. All incorrect predictions are scored the same. Ordinal models take this ordering into account. For the multi-class problem, we use two variants of the ordinal logistic model. Both take the same approach as the binary logistic regression model, producing a 'risk score' for an establishment from a linear sum of weighted features (which is then transformed by the logistic function) but in the multi-class ordinal case, the set of possible risk scores is segmented into bands. If an establishment gets a risk score in a certain band it is assigned to the corresponding class (FHRS score). E.g. if the risk score is between 80% and 100% the establishment is assigned a hygiene score of zero, 80%-70% a score of 1 etc. The two variants of this algorithm that we test are the All-threshold and the Immediate-threshold variants[5].

Summary

Each of the models has advantages and disadvantages, one of the key trade-offs is interpretability versus performance. In general, it is easier to understand and interpret the results of simpler models such as decision trees and logistic regression, but more sophisticated models make better predictions. For this project, it is important for the FSA and LAs to be able understand why a particular prediction has been made. We therefore focus principally on the more interpretable models (logistic regression and decision trees), but investigate the others for benchmarking purposes.

The previous section gave a brief overview of the various models, there are many resources online for the interested reader to learn more about each of these algorithms.

4.4. Insights from FHRS Data Set

The FSA has data containing the FHRS ratings of all applicable establishments nationally. The data is available through a web-based API. The dataset contains few data fields (features) about the business other than name, address and establishment type, however it can be used to make some useful observations that inform our modelling.

[5] We use the mord package in python for implementation: https://pythonhosted.org/mord/index.html. Details of both models are available here: http://qwone.com/~jason/writing/olr.pdf. In the All-threshold version, loss is incurred for all thresholds, not only the neighbouring ones, as in the Immediate-threshold version.

The data is highly unbalanced

As can be seen in Table 2 the data is highly imbalanced. Only 2.7% of establishments are seriously non-compliant (rating 0 or 1) and only 5% are broadly non-compliant (rating 0, 1 or 2). This class imbalance presents a modelling challenge; to illustrate, consider that a 'dumb' model would be 95% accurate by predicting that all establishments are broadly compliant, the performance gains to be made are in just 5% of the data.

Table 2 Distribution of FHRS ratings given to establishments nationally (excludes Scotland)

Rating	Percentage of establishments
5	69.9%
4	17.4%
3	7.8%
2	2.3%
1	2.4%
0	0.3%

Certain business types have lower compliance than others

The FHRS data shows that compliance varies by establishment type. Figure 10 shows that takeaways have particularly low rates of compliance compared to other business types.

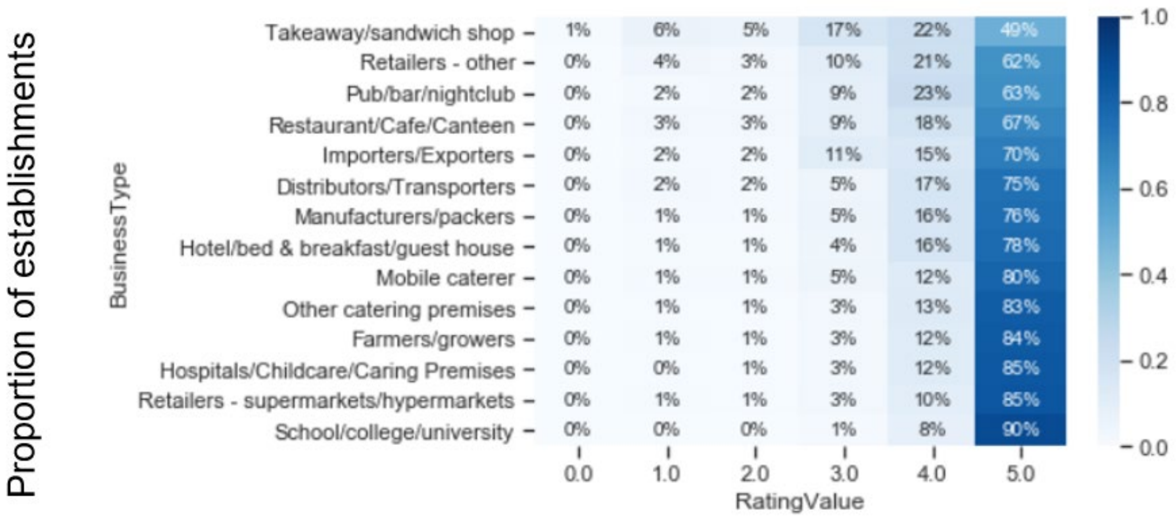

Figure 10 FHRS rating value by business type, percentage of establishments

4.5. Feature Engineering

Business Chains

There is useful information contained in the establishment name. We hypothesise that large chain establishments may have a better rating than average. Using the top occurring establishments in the FHRS dataset we identified a list of 48 chains with over 100 establishments. Figure 11 shows that the hypothesis is correct for most of the identified chains. This list of chains and means contains useful information, and is included as an additional feature in the model. If an establishment name is identified as matching one in the chain list it is given the feature 'ChainMean' equal to the mean rating for that chain in the FHRS dataset as an input to the model, otherwise it is given the non-chain mean value.

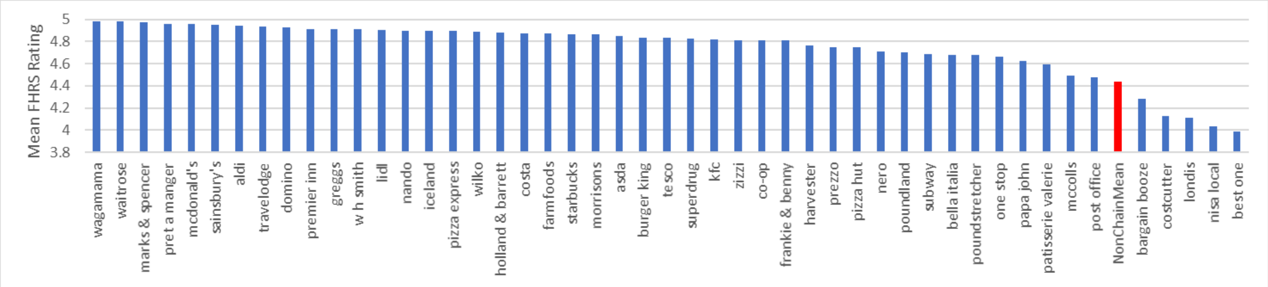

Figure 11 Mean FHRS Rating for chains, compared to non-chain mean

Names

An establishment's name can also tell us about the nature of the business, for example, if it contains the word 'cake' or 'bakery' we know that it is very likely to be a bakery. We define 13 categories and corresponding key words, if the business name contains that key word the business is assigned to that category. This is depicted in Figure 12.

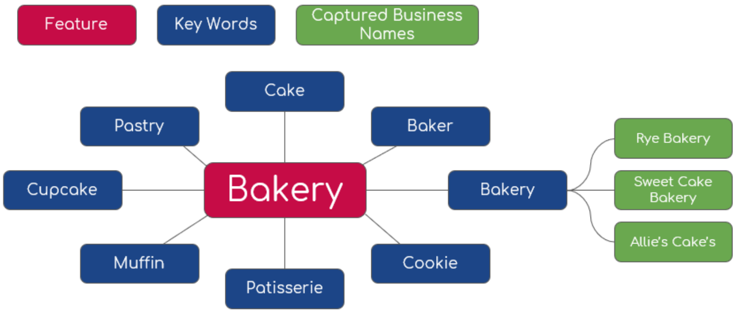

Figure 12 How establishments are encoded to categories, based on the establishment name

Figure 13 shows that rating varies with assigned category. With 'chicken' establishments (containing the word 'chicken' or 'peri peri') having a mean rating one level below no category establishments. In the same way as the chain information, this is included as a feature in the model.

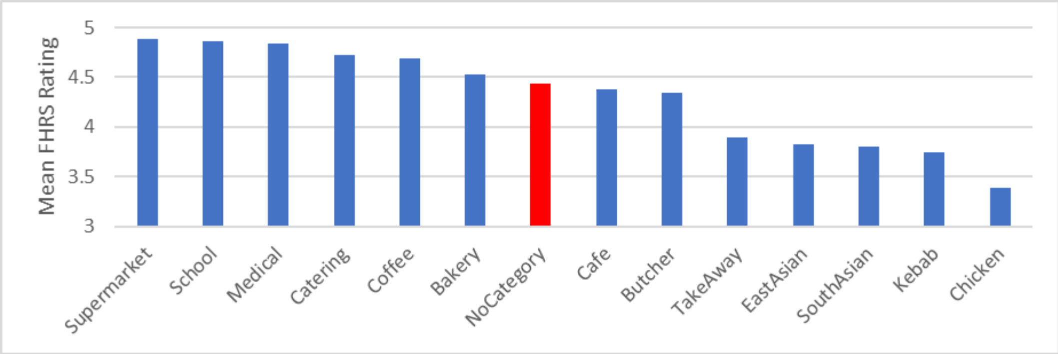

Figure 13 Mean FHRS rating by business name category

4.6. Pipeline

Figure 14 depicts the modelling pipeline.

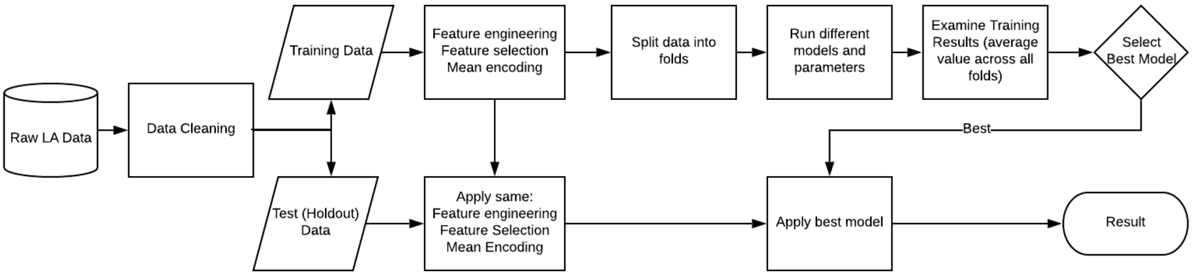

Figure 14 Illustration of the modelling pipeline

Each of the stages is described below:

Data Cleaning: This includes converting data fields to numerical format, removing establishments with critical data missing, dropping unusable data fields and removing duplicate entries.

Splitting the data: The data is split into a training and test (also known as holdout) set. In this way the model is evaluated on unseen data and is therefore a good representation of expected real-world performance. For the results presented below the holdout proportion is 20% of the data set.

Feature engineering: Features are added to the training set for chains and name categories, as described in the previous section.

Feature selection: Highly correlated features (data fields) and features with very few data points are removed. An iterative process of feature selection is undertaken, selecting intuitively relevant features, or removing features with numerically low feature importance in model results.

Mean encoding: Missing data values are filled with mean values for that feature.

Splitting the data into folds: The test dataset is split into 'folds' i.e. a subset of different train/test splits. Model performance is evaluated across each of these different model runs. This allows us to account for the variability of results with different data samples.

Run different models and parameters: Each of the modelling algorithms is run with various parameter settings.
Examine training results and select best model: The results for each of the models is examined and the model and parameter settings that has the best (highest mean) results across all of the samples is selected.
Applying the model to the test set: The same pre-processing steps that were applied to the training set are applied to the test set. The best model is applied to the test set and results obtained.

The model and pipeline were developed using Python, a widely-used, free, open-source programming language for data science. Our approach to data quality assurance is given in section 3 above. Quality assurance logs for modelling are supplied with the supporting documentation.

5. Results

5.1. Approach

The lack of commonality between fields collected by different LAs meant that it was not possible to create a high-quality common dataset across LAs. We therefore developed a modelling approach that was applicable to each LA dataset separately. We begin by presenting detailed results for the City of London as an example, before summarising the results for all data sets. For the exemplar City of London dataset, we present results for the non-broadly compliant premises model, the restaurant only model, the fully compliant premises model and the multi-class model. The results for the registration data model on the Plymouth dataset are then presented. Finally, we summarise results for all data sets.

Lack of commonality between LA data sets

LAs are only mandated to collect a small number of common data fields at inspections, such as the business name, address, business type and the food hygiene inspection scores. Although LAs are expected to follow the Food Law Code of Practice, they are free to take different interpretations of what metrics are important, this results in a lack of commonality between local authority data sets. There were very few data fields that were the same across the majority of LAs. It should also be noted that amongst LAs using an inspection app, data fields were not common as each LA defines its own survey questions within the app. Lack of commonality between LAs meant that we could not build a model for one LA and test on another.

For an example of the lack of commonality in data fields between LAs we look at the data collected on risky processes.

City of London collects data on key risky processes including:

- Lightly processed risky foods, e.g. less than thoroughly cooked burgers, sushi, tartare, meat/fish carpaccio
- Vacuum packing raw and ready to eat foods
- Sous vide processing
- Other processing with potential for public health risk, e.g. cold smoking, fermentation/ air drying of meat

The closest data fields for Charnwood are:

- Whether certain complex equipment is used on site, e.g. mincer, slicer, vacuum packers
- Other equipment/setup details such as whether there is separate equipment for raw and cooked food
- Significant food risk hazards e.g. low risk foods, frozen foods, wrapped high risk, open RTE food, open raw food
- Raw foods e.g. chicken, burgers, fish, sausages, vegetables

It can be seen that creating a common mapping for this single question and for just two LAs would be challenging if at all possible. For tens of data fields and for many LAs, the mapping to common fields was not possible.

For this reason, we were unable to create a common dataset containing a useful number of data fields. This meant that there were too few examples for some models to be built including seriously non-compliant premises and for business groups other than restaurants.

Components of hygiene compliance

Investigations into developing separate models for each component of the FHRS rating ("Compliance Hygiene", "Compliance Structural, "Confidence in management") indicated that predictive features for each were very similar, due to the strong correlation between these variables. The final models were not expected to deliver substantial insight into the differences in drivers of each of these components. For this reason, it was agreed with that development effort would be directed at the other models.

5.2. City of London

Data Overview

The City of London data was collected by inspectors using a smartphones/tablet app. The data was collected between 04/04/2018 and 15/01/2019. There are 445 establishments in the dataset after cleaning.

The number of establishments of each type is given in Table 3.

Table 3 Establishments by type, City of London data

Establishment Type	Number of establishments
Restaurant Cafe Canteen	277
Take away	71
Pub Club	38
Other Caterer	22
Retailer Small	11
Hotel-Establishment	8
Mobile food unit	6
Supermarket inc. all chains	5
School College	3
Caring establishment	2
Retailer other	2

Table 4 gives the proportion of establishments by rating value.

Table 4 Percentage of establishments by rating value, City of London data

Rating Value	Percentage of establishments
5	71.0%
4	14.6%
3	4.9%
2	3.6%
1	4.9%
0	0.9%

The fields in the data are:

- **Business information** (Premise name, address, appropriately registered, primary authority partnership, does it have a head office, opening times, establishment type, number of customers served, distribution regional/national/international)
- **Activities of the business** (supplies food to other establishments, import, produces high risk foods for vulnerable groups, uses 3rd party delivery companies, does transport or delivery)
- **Food processes** (pre-packed good made on site, Risky processes (inc. Sous-vide, vac packing etc)
- **Processes** (traceability for food purchases, 3rd party accreditation for processes, Food Safety Management (FSMS) system type, means to ensure FSMS is reviewed, accreditation type)
- **Inspection Scores**

5.3. Tuning the model

The parameters were tuned for each of the model types to optimise performance. For instance, we want to avoid overfitting the model to the characteristics of a particular data sample. Figure 15 shows the effect of changing the regularisation parameter for the logistic regression model. This parameter affects how many features are included in the model. When it is very low, it leads to simple models that do not fit well. When it is very high, it leads to very complex models. They tend to fit the training data so well that they are no longer good predictors of the test data. Somewhere in between, the precision of the model can be maximised. The plot shows that tuning just one model parameter can make a big difference on model performance. This process was repeated for various parameters across multiple folds (subsets) of the training set to optimise each model individually. The figure shows box plots of the average precision for different runs of the training and test set[6].

[6] Note that the test set is one fold of the overall training set, not the holdout set

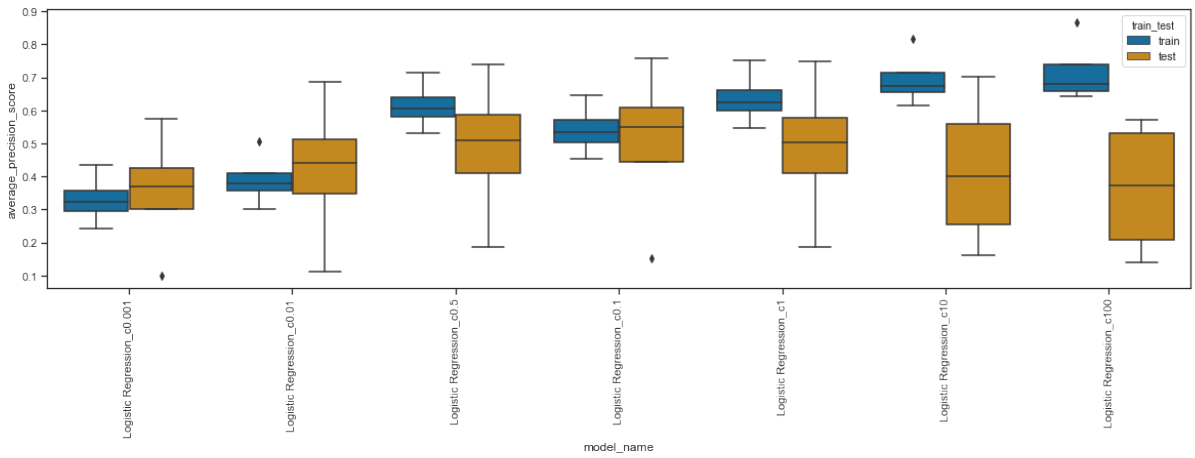

Figure 15 Results for different values of regularization parameter for logistic regression model, City of London data

5.4. Non-broadly compliant model

We developed a model for non-broadly compliant premises, those with FHRS ratings of 2 or less.

Model Selection

Figure 16 shows the performance of different algorithms (after the best parameters have been selected) against benchmarks. The performance measure is average precision. The mean of this score is taken across each data fold in the training set.

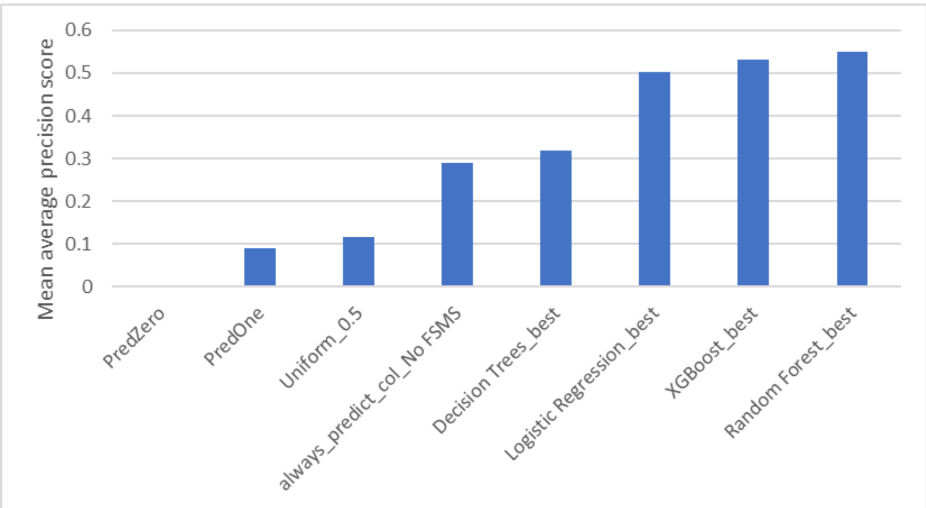

Figure 16 Performance of best tuned algorithms (mean of cross-validated results in training set) and benchmarks, broadly compliant binary model

The benchmarks are as follows:

- **PredZero**: Always predict that an establishment will pass.
- **PredOne**: Always predict that an establishment will fail.
- **Uniform_0.5**: Assign a risk score with uniform probability. This is a benchmark representing a model with no predictive power. As an example, in Figure, establishments

would be randomly distributed along the risk line, the model has no power to discriminate between failing and passing establishments.

- **Always_predict_col_No FSMS**: Always predict that an establishment will fail if it has no food safety management system.

All models outperform the benchmarks. PredOne represents the current approach which employs the precautionary principle, i.e. there is no supporting information or data on the levels of compliance with a new business, so an initial inspection is required. This is a low precision approach, of those establishments inspected, many pass. Increasing the average precision means that inspections are more focussed, inspections are more likely to identify failing establishments. The tuned logistic regression model increases the score by 0.4 above the PredOne benchmark. Decision trees demonstrate substantially worse performance than logistic regression. The less interpretable algorithms Random Forest and XGBoost have slightly better performance than logistic regression, however the performance gain is not substantial. Given the benefits to the FSA of interpretability, we select Logistic Regression as the model of choice. Similar results to those presented for City of London are also observed in other LA data sets.

Results, non-broadly compliant model, City of London

Figure 17 shows the results of the model on the unseen test data set for different thresholds. The model demonstrates good performance, with good ability to classify failing establishments as fails, with a low level of false predictions. If used in practice, the model would be a useful tool for prioritising high risk establishments. The average precision is 0.48 in the training set and 0.61 in the test set.

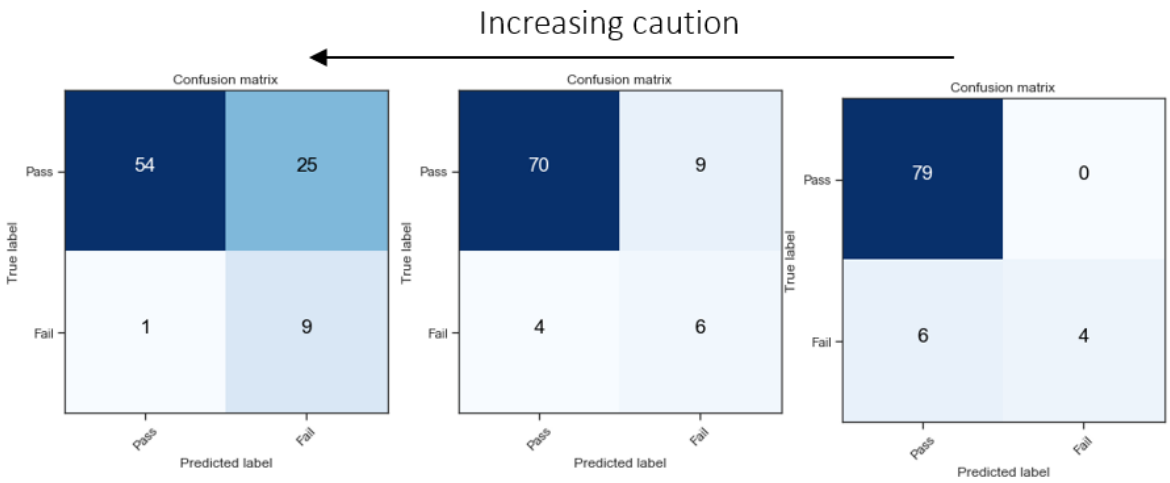

Figure 17 Best logistic regression model results on unseen test data, non-broadly compliant prediction model. The central confusion matrix maximises the f2 score, (weighted harmonic mean of precision and recall) the other two are selected to be illustrative of higher and lower risk thresholds.

Figure 18 shows the results of 3 benchmarks on the same data set for comparison.

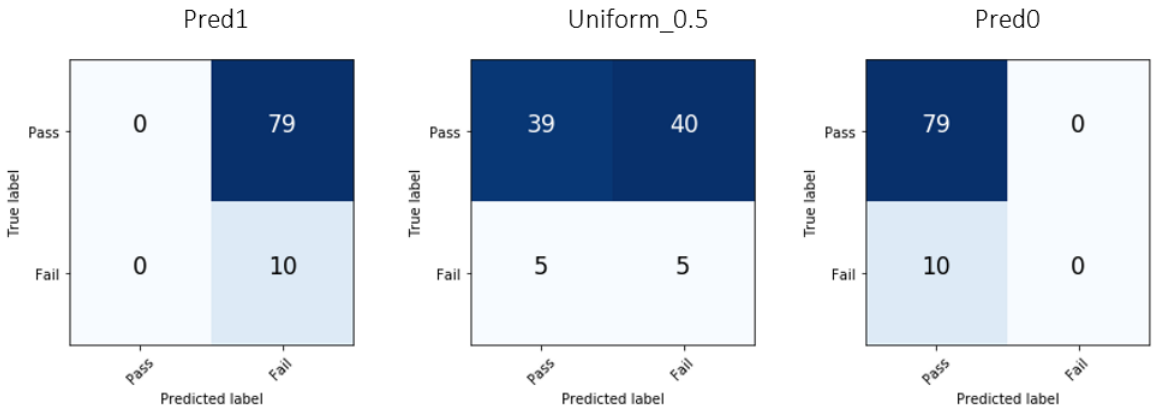

Figure 18 Benchmark results on test data, non-broadly compliant prediction model

Figure 19 shows the feature importance for each of the selected features within the model. Full details of the calculation of these values and how to interpret them are available in Appendix C. Roughly speaking, for yes/no features, the value estimates how much the risk probability increases when this feature is present (versus a reference scenario). Therefore, having no Food Safety Management System might increase an establishment's risk by about 8 percentage points.

The features that stand out as being particularly important within the model are process related features, for example: whether the establishment has a food safety management system, whether that system has a means to ensure it is reviewed, whether they have a supplier assurance scheme, whether they have traceability for food products purchased.

Feature Name	logstic(coefficient)-0.5
No FSMS	0.08
RiskyProcessesActivities_VacPacRawAndRTE	0.05
Does this business supply food to other food establishments?	0.04
RiskyProcessesActivities_LightlyProcessedRiskyFoodsEgSushiLTTCBurgersTartareCarpaccio	0.03
Take away-Establishment providing convenience foods primarily for consumption off the premises	0.03
Restaurant Cafe Canteen	0.03
This business serves the local area only or less than 1000 customers per day	0.02
AccreditationOrAssurances_Third party	0.01
RiskyProcessesActivities_SousVide	0.01
Other Caterer (Vessels, Home caterers,selling direct to consumers, Community Centres, WoCo's)	0.01
FSMS_Consultant designed	0.01
Open Intermittent (e.g. Event catering)	0.00
Hotel-Establishment providing catering only to customers staying with them	-0.01
Is the premises appropriately registered	-0.01
Are pre-packed goods made on site and sold elsewhere?	-0.01
AccreditationOrAssurances_Other	-0.01
Are High Risk foods produced for vulnerable groups?	-0.02
Retailer Small	-0.02
Open Daytime	-0.02
The business use a third party(ies) to accredite or assure it's processes and procedures	-0.02
Open Evening	-0.03
ChainMean	-0.03
AccreditationOrAssurances_Primary Authority	-0.03
FSMS_Other	-0.03
CategoryMeanRating	-0.04
This business serves a substantial number of customers (1000+), inc. sig. prop. from outside the local area	-0.04
Open Night time Economy	-0.04
FSMS_SFBB	-0.05
Pub Club	-0.05
Primary Authority: Does this business have a Primary Authority Partnership?	-0.07
FSMS_In house / Company own HACCP	-0.07
Can the business demonstrate adequate traceability for the food products it purchases?	-0.10
Does the business have a head office	-0.11
Does the business have Supplier assurance scheme (e.g. BRC, SALSA), Supplier auditing, Specifications for purchasing	-0.14
Does the food safety management system have a means to ensure it is reviewed?	-0.17

Figure 19 Model features by importance, higher numbers contribute to higher risk (more likely to fail)

5.5. Restaurant only model

Using the City of London dataset, we implement a restaurant/café/canteen only model, by stripping out other establishment types, 277 restaurants remain.

The performance of the model is shown in Figure 20. The average precision is 0.54 in train, 0.60 in test. This is a similar level of performance to the all-establishment model.

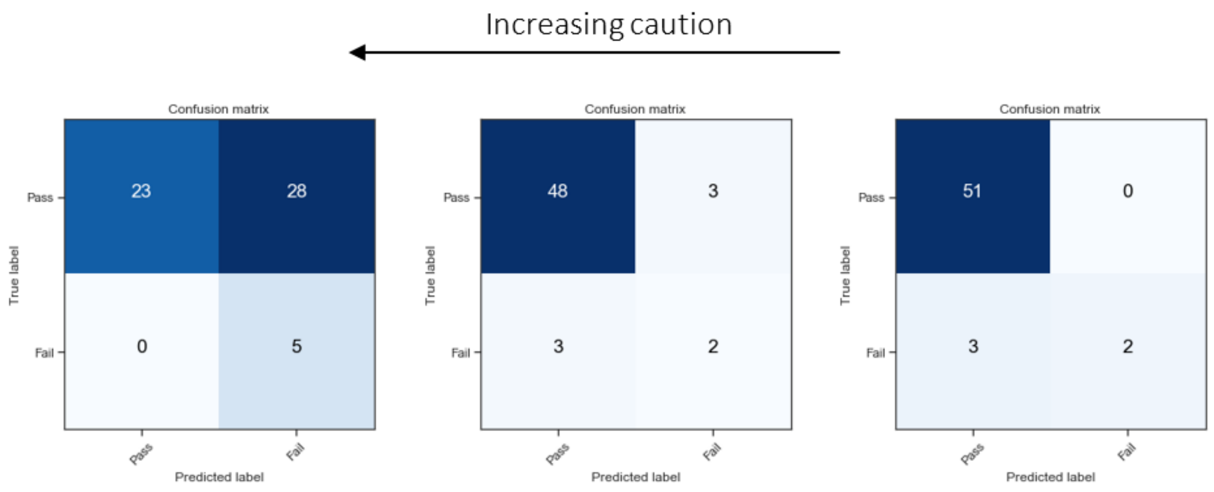

Figure 20 Results for restaurant only model, City of London broadly compliant prediction model. The central confusion matrix maximises the f2 score, (weighted harmonic mean of precision and recall) the other two are selected to be illustrative of higher and lower risk thresholds.

The features for the restaurant only model are given in Figure 21. They are very similar to the all-establishment model. This is perhaps expected as over half of the establishments in the 'all establishment types' model are restaurants, another 25% of the establishments are pub/club or Take-away. Only 13% of the establishments are more distinctive establishment types. Substantial differences in compliance may be discernible between establishment types with larger samples of other establishment types.

Feature Name	logistic(coefficient)-0.5
No FSMS	0.07
RiskyProcessesActivities_LightlyProcessedRiskyFoodsEgSushiLTTCBurgersTartareCarpaccio	0.03
RiskyProcessesActivities_VacPacRawAndRTE	0.03
RiskyProcessesActivities_SousVide	0.01
This business serves the local area only or less than 1000 customers per day	0.01
Open Intermittent (e.g. Event catering)	0.01
Open Evening	0.01
Does this business supply food to other food establishments?	0.00
Open Daytime	0.00
Are High Risk foods produced for vulnerable groups?	0.00
Are pre-packed goods made on site and sold elsewhere?	-0.01
FSMS_Other	-0.01
AccreditationOrAssurances_Primary Authority	-0.01
AccreditationOrAssurances_Other	-0.02
ChainMean	-0.02
AccreditationOrAssurances_Third party	-0.02
Open Night time Economy	-0.02
This business serves a substantial number of customers (1000+), including a significant pr	-0.02
FSMS_SFBB	-0.02
CategoryMeanRating	-0.03
FSMS_Consultant designed	-0.03
Is the premises appropriately registered	-0.04
The business use a third party(ies) to accredite or assure it's processes and procedures	-0.04
Primary Authority: Does this business have a Primary Authority Partnership?	-0.04
FSMS_In house / Company own HACCP	-0.05
Does the business have Supplier assurance scheme (e.g. BRC, SALSA), Supplier auditing, Sp	-0.09
Does the business have a head office	-0.09
Can the business demonstrate adequate traceability for the food products it purchases?	-0.09
Does the food safety management system have a means to ensure it is reviewed?	-0.16

Figure 21 Model features by importance, ‘restaurant only’ ‘broadly compliant’ model

Aside on understanding feature importance

The importance that a model puts on particular features depends on the whole data set that it is given and the interplay between features. For example, if a dataset does not contain information about whether an establishment is a take-away it may identify using delivery vehicles as a proxy for this (even if other establishment types use delivery vehicles too). Likewise, if having FSMS documentation is generally an important feature, but it is missing in a particular data set, some other features may be selected as important. Remember that the models do not make logical judgement about the meanings of particular features, but rather identify statistical trends in the data. This should be considered when assessing the top features for each model: each dataset will tell a slightly different story about what is important.

5.6. Fully compliant premises model

Our model can be applied to other segmentation problems, for example to identify whether premises will be fully compliant or not. The model demonstrates good performance and out-performs the benchmarks, as shown in Figure 22. The benchmarks are higher for this problem as the number of non-fully compliant premises is higher than the number of non-broadly compliant premises, making it easier for a 'dumb' model to make high precision predictions.

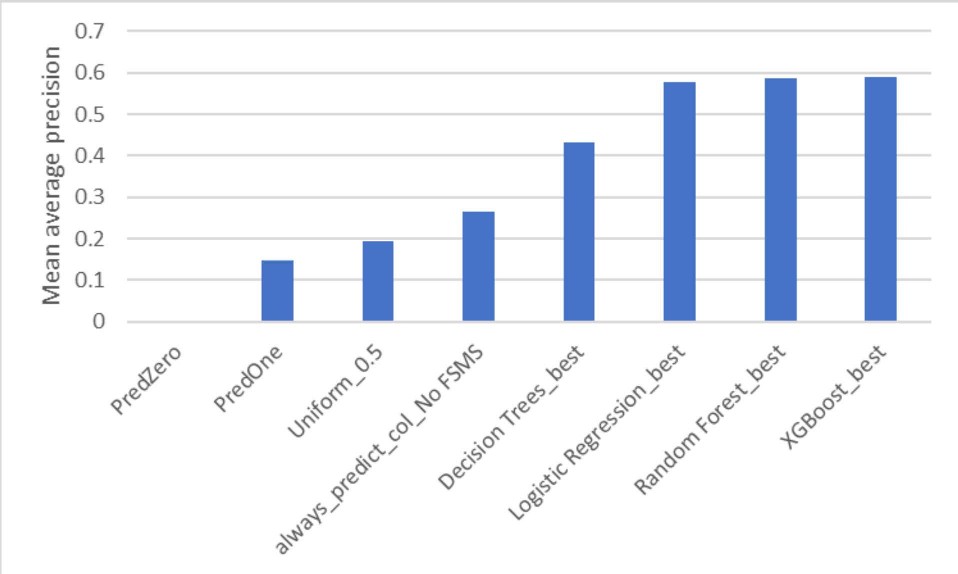

Figure 22 Performance of model (mean of cross-validated test results in training set) against benchmarks, fully compliant binary model

The results for the model are given in Figure 23. Average precision is 0.61 in the training dataset and 0.57 in the test set.

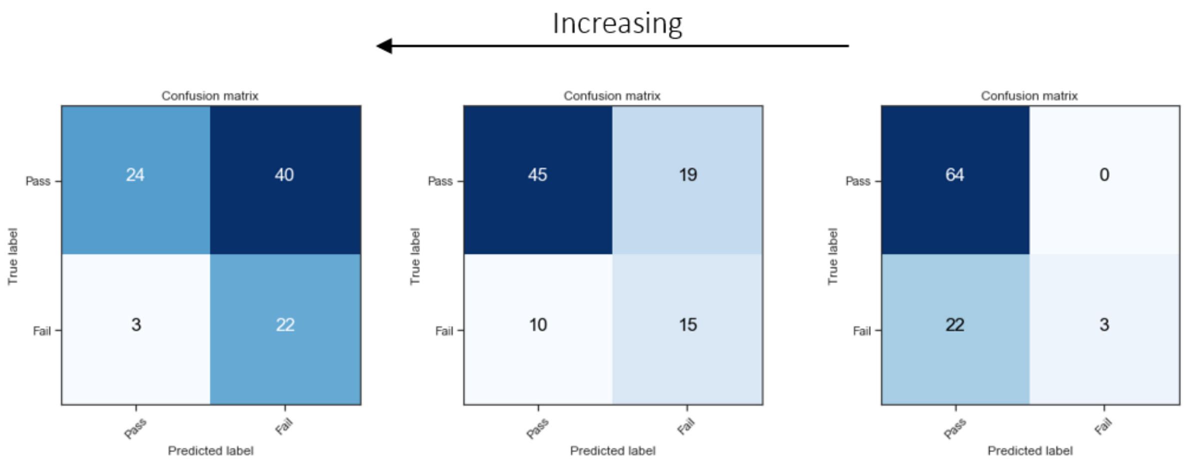

Figure 23 Results on unseen test set at different thresholds, fully compliant prediction model. The central confusion matrix maximises the f2 score, (weighted harmonic mean of precision and recall) the other two are selected to be illustrative of higher and lower risk thresholds.

Figure 24 shows the feature importance for the fully compliant model. It also gives a comparison to the feature importance values in the broadly compliant model. The strongest predictors are similar: having a food safety management system review process and supplier auditing scheme are still the strongest predictors. There are some changes in other variables, but overall changes are not substantial.

Feature name	logistic(coefficient)-0.5		Increase over broadly compliant model	
Open Evening	0.07		0.10	
Take away-Establishment providing convenience foods primarily for consumption off the premises	0.07		0.04	
Open Daytime	0.06		0.08	
This business serves the local area only or less than 1000 customers per day	0.06		0.04	
Does this business supply food to other food establishments?	0.05		0.02	
No FSMS	0.04		-0.04	
Restaurant Cafe Canteen	0.04		0.01	
RiskyProcessesActivities_VacPacRawAndRTE	0.03		-0.02	
AccreditationOrAssurances_Other	0.03		0.04	
FSMS_Consultant designed	0.03		0.02	
FSMS_SFBB	0.03		0.08	
Pub Club	0.03		0.08	
RiskyProcessesActivities_LightlyProcessedRiskyFoodsEgSushiLTTCBurgersTartareCarpaccio	0.02		-0.02	
Can the business demonstrate adequate traceability for the food products it purchases?	0.00		0.09	
Is the premises appropriately registered	0.00		0.01	
RiskyProcessesActivities_SousVide	-0.01		-0.02	
Hotel-Establishment providing catering only to customers staying with them	-0.02		-0.01	
CategoryMeanRating	-0.02		0.03	
ChainMean	-0.02		0.01	
FSMS_Other	-0.02		0.01	
Are High Risk foods produced for vulnerable groups?	-0.02		-0.01	
Other Caterer (Vessels, Home caterers,selling direct to consumers, Community Centres, WoCo's)	-0.02		-0.03	
This business serves a substantial number of customers (1000+)	-0.03		0.01	
Are pre-packed goods made on site and sold elsewhere?	-0.03		-0.02	
Open Intermittent (e.g. Event catering)	-0.03		-0.03	
Retailer Small	-0.04		-0.02	
AccreditationOrAssurances_Third party	-0.04		-0.05	
The business use a third party(ies) to accredite or assure it's processes and procedures	-0.05		-0.03	
Open Night time Economy	-0.06		-0.01	
AccreditationOrAssurances_Primary Authority	-0.06		-0.03	
FSMS_In house / Company own HACCP	-0.07		0.00	
Does the business have a head office	-0.08		0.02	
Primary Authority: Does this business have a Primary Authority Partnership?	-0.11		-0.04	
Does the business have Supplier assurance scheme (e.g. BRC, SALSA), Supplier auditing	-0.13		0.01	
Does the food safety management system have a means to ensure it is reviewed?	-0.17		0.00	

Figure 24 Model features by importance, ‘fully compliant’ model. Second column gives the increase in the feature value over the non-broadly compliant model.

The model is useful for predicting different FHRS thresholds and could therefore be used by LAs in a way that suits their particular intervention preferences.

5.7. Multi-class model

As an additional view on the hygiene compliance problem we developed a multi-class model to predict the exact FHRS rating of an establishment, rather than separating the problem into binary problems (such as predicting compliant or not compliant). Figure 25 compares the performance of different multi-class algorithms against benchmarks. The benchmarks are as follows:

- **Always_predict_col_no FSMS**: Always predicts 0 if the model has no food safety management system and 5 otherwise;
- **PredFive**: Always predict 5;
- **PredZero**: Always predict zero (not shown on chart as MAE is close to 1 and removes detail from chart).

The Mean Absolute Error (MAE) measures how far away predicted ratings are on average from the correct value. Compared to always predicting that an establishment gets a rating of 5, the best model is an improvement of 0.1. Remember that the dataset is very unbalanced, most establishments get a rating of 5, so the benchmark model has low error. A perfect model would have a MAE of zero. The model is a 17% improvement over the benchmark. There is a small improvement in performance from using the 'all threshold' variant of ordinal logistic regression over the non-ordinal one-versus-all logistic regression.

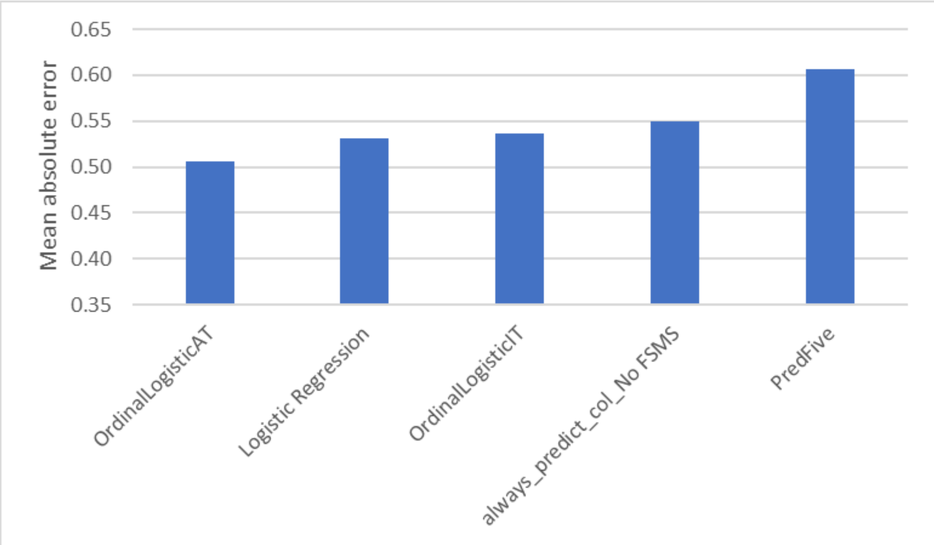

Figure 25 Multi-class model performance comparison, mean over test runs in training dataset, City of London

Figure 26 shows the performance of the model in the unseen test set, benchmarks of 'always predict zero' or 'always predict one' are given for comparison. Predictions along the top-left to bottom-right diagonal are perfect predictions, towards the top-right are too-high-classifications (model under-rates the risk), towards the bottom left are too-low-classifications (model over-rates the risk).

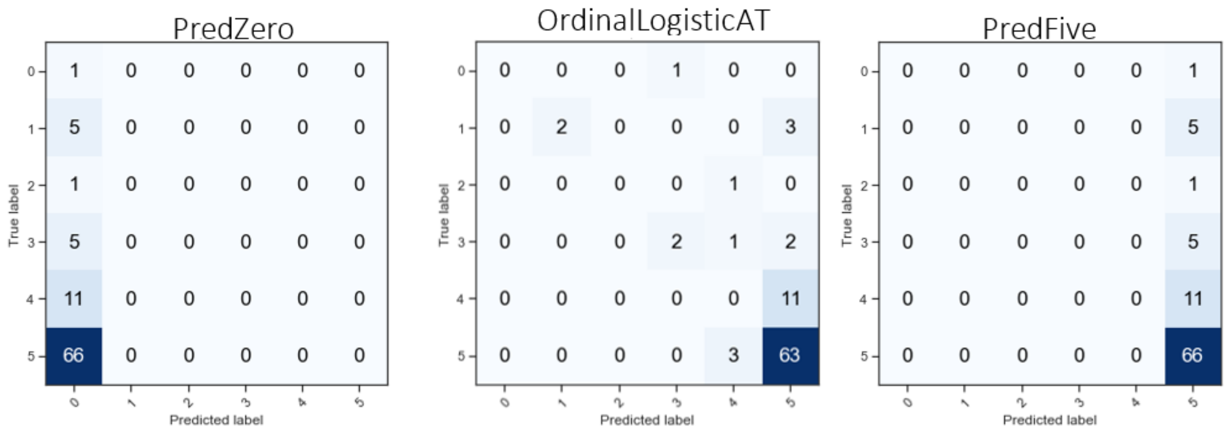

Figure 26 Performance of selected model (Ordinal Logistic AT) against benchmarks in unseen test set.

The model correctly predicts a third of the establishments with ratings lower than 4. It gives the FSA/LAs a more detailed picture of which establishments are likely to be risky. However, a binary model that is trained to focus on the distinctions between the two classes e.g. broadly compliant/non-compliant is better at finding the drivers of this separation than the model that is trained to categorise establishments more generally.

5.8. Registration data only model - Plymouth

We used the data registration data provided by Plymouth, to examine whether this self-reported data could be used to build a useful prediction model.
The registration data contained the following features:

- **Business information**: Premise name, address, primary authority partnership, does it have a head office, establishment type, distribution regional/national/international;
- **Activities of the business**: supplies food to other establishments;
- **Food processes**: Risky processes (inc. Sous-vide, vac packing etc);
- **Processes**: training of staff and management.

There were a very substantial number of establishments in the Plymouth registration dataset that weren't in the dataset of scored establishments, perhaps because registered and not yet inspected or excluded from inspection. This reduced the data set to 182 establishments, which is at the bottom end for a feasible data set size, there are only 8 establishments with a FHRS rating of 2 or less. In order to increase the balance in the data set, in addition to the non-broadly compliant model we developed a non-fully compliant model.

Unfortunately, neither model produced results that had good predictive power, they were not significantly better than benchmarks. We believe this in part likely to be due to the small data set size, some of the fields had only a couple of examples. Other fields which were important in the City of London model are also missing, e.g. information about processes such as the food safety management system. There may also be some disadvantage arising from the fact that the dataset is self-reported by establishments.

5.9. Other LA Datasets

We applied our (binary logistic regression) model for non-broadly compliant premises (rating of 2 or less) on each of the LA datasets in order to examine the applicability more broadly and to identify the key drivers of non-compliance across LAs. We also tested the other model types on these datasets: our logistic regression model was among the top performing models in each of the datasets, indicating its general applicability. The full results for each of the datasets, including the top features for each model are given in Appendix D. A summary of performance is given in Figure 27.

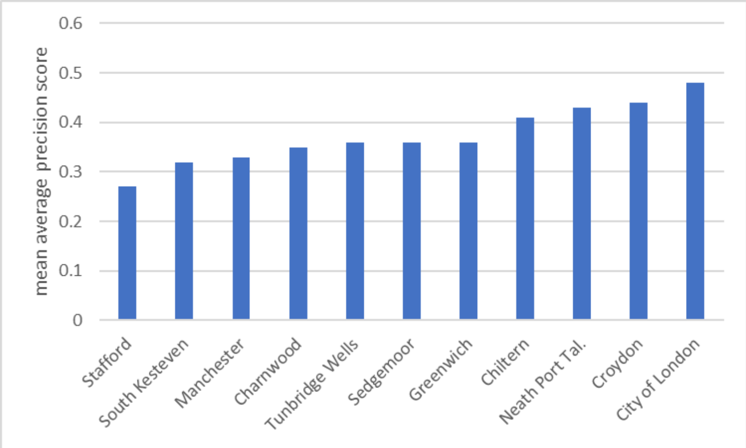

Figure 27 Model performance by LA, mean of cross validated training set results

All the models demonstrate good performance; they are substantially above the 'no predictive power' benchmark (which varies but is around 0.13). The performance ranges from an average precision of 0.27 to 0.48, meaning that across different thresholds (different levels of caution) between 27% and 48% of fail predictions will be correct (the precision will generally be higher than this for a less cautious threshold and vice versa).

Looking at the results across each of the L.A.s (See Appendix D), there is not a large difference between mean average precision for train and test, results are not overfit to the training data i.e. the model generalizes well to unseen data[7].

There is a wide variety in the data in each of the LA datasets. Despite the fact that this meant we were unable to create one large merged dataset with many samples, it meant that, effectively, each LA acts as an independent experiment, providing insight into which data is important.

A significant reason for the variation in performance between the LAs is differences in the way that data is captured. Data sets from inspection apps or from proformas with structured numerical/check-box style inputs have better performance as more information is usable by the models; City of London and Croydon are examples. Stafford and Manchester are examples of less modelling friendly inspection proformas, as they have fewer quantitative fields. Some datasets had a large number of inspector judgement style fields (e.g. "Is the probe thermometer being used correctly?", compared with "is a probe in use") which would not be self-reportable by establishments, and so were discarded for this modelling (however, the data would be interesting for other analysis, such as understanding drivers of inspection failures).

[7] In several instances the mean average precision for the test set is a little higher than the train value. This is due to the relatively small test sample sizes, and relatively high variance in the test set results. In some instances, the test set is able to perform particularly well, increasing the mean test value.

Looking across the LA models gives us insight into what are the important drivers of compliance.

Table *5* shows the most important features identified by the models and how frequently they were among the top predictors. The data behind the numbers are given in Appendix D. The table clearly shows that information around the food safety management system (such as whether it is documented or has a means to ensure it is reviewed) is the most important driver of compliance, indeed for 8 for the models it is the most important predictor. Following this are staff training, personal hygiene, supplier assurance, whether the establishment is a take-away and information about surface cleaning. Other important predictors include the use of a probe, allergen precautions, whether the business is responsible for the structure, how equipment is washed and whether the business undertakes cook-chill activities.

Table 5 High importance features, count by LA model

Feature	**Number of LAs out of 13**
FSMS: documented, means to ensure reviewed, records kept	10
Staff/Manager training	5
Personal hygiene: hand washing/drying/facilities/protective clothing	5
Supplier assurance scheme/ Food traceability	4
Take-away/Delivery	4
Surface cleaning: disposable cloths/disinfectant/schedule/two stage cleaning	4
Probe: use/cleanliness	3
Allergen info: staff can identify allergens/notices for customers	3
Structure: responsible for structure/repair	2
Equipment washing: dishwasher/by hand	2
Chilling/cook-chill activities	2

6. Concept Use Case

As an additional piece of work over and above the project specification, we developed a concept demonstrator[8] user interface to illustrate how the model might be used in practice. A hypothetical concept use case is as follows:

Following the successful proof of concept demonstrated in this report, the FSA implement a trial roll-out of the system with a single LA. Working in partnership with the LA, new data fields (e.g. details of the FSMS, details of their supplier assurance scheme, etc) are added to their existing

[8] Using Power BI

online registration form. The model is deployed locally within the LA. Inspectors in the food hygiene team have access to the dashboard shown in

Figure 28 - When a new business registers through the online portal, the data is sent to the model, stored in a database and displayed on the user interface. The manager of the team can log in and see the following:

1) A list of establishments prioritised by risk and whether they are expected to be non-compliant;
2) A map of uninspected establishments, showing expected risk;
3) The importance the model places on different data fields when making a risk estimate;
4) A breakdown of the prediction for a selected establishment.

Using this information, the manager can plan and prioritise inspections, diverting resources from low risk establishments and targeting high risk ones, making the best use of resources and minimising risk.
Beyond this, future developments could integrate data from app-based inspections, to verify registration information and to make predictions about follow-up visits.

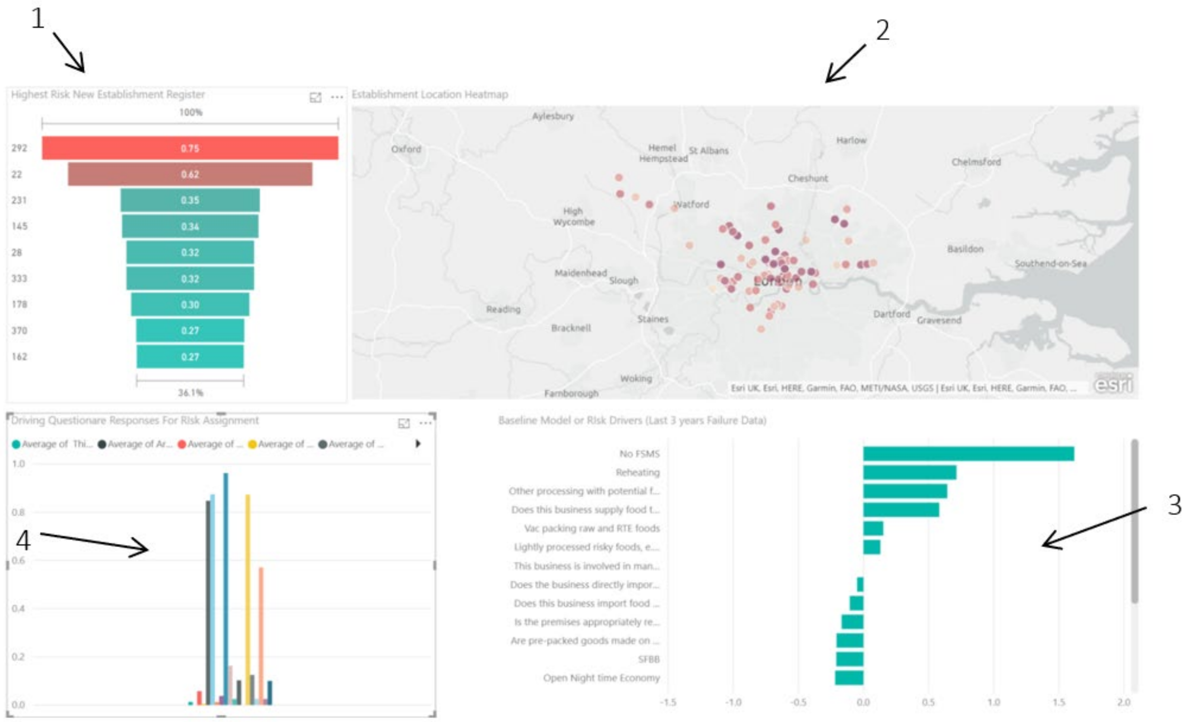

Figure 28 Concept demonstrator user interface

7. Strengths and Weaknesses

Strengths of the work are as follows:

- A large data set was collected: 8,765 establishments from 13 LAs. The data is a reasonable representation of LAs in England and Wales with regards to LA type, geography, and rural/urban split. It contains inspection data collected via both inspection apps and hand-written inspection notes.
- The developed models have good predictive power. The project demonstrates the potential for hygiene compliance risk prediction modelling in the future of new food business regulation.
- The selected logistic regression models are highly interpretable, giving a clear explanation for why a particular prediction has been made and a clear indication of the importance of different data fields.
- The modelling pipeline compares the developed logistic regression model to other model types. There is not a substantial performance penalty compared to other powerful but less interpretable model types.
- Each LA data set is distinctive and contains data fields particular to that LA. The range of different data gives a broad view of potentially useful establishment data.
- The model was applied to each of the distinct LA datasets and demonstrated good predictive power on each, demonstrating its general applicability.
- Across the different LAs, certain data fields are consistently identified as important.
- Data fields used in the model have the potential to be self-reported by establishments.
- This model was developed for food hygiene compliance due to the availability of good data, but demonstrates the potential for predictive modelling in other areas such as food standards compliance.

Weaknesses/limitations of the work are as follows:

- Establishment data is collected by inspectors, not self-reported by establishments. In the envisaged deployed version of this model data would be either self-reported or drawn from public available data sets where possible.
- Due to the inconsistency in data sets between LAs it was not possible to develop a single large data set. Results are correspondingly on smaller samples of data.
- Due to the smaller size of data sets it was not possible to develop models containing only very few examples of non-compliant establishments, such as seriously non-compliant, or manufacturer only models.
- The level of consistency and accuracy in interpreting and answering questions by inspectors is unknown.
- Variation between LAs and officer interpretation of in wording for similar questions might make some better predictors than others.

8. Next Steps

Our suggestions for further development are as follows:

1) Through this project, a rich dataset of establishment inspection data has been collected. The focus for this project was to develop hygiene compliance prediction models for an establishment's first inspection. However, the data is likely to be very insightful for many other applications, for example: developing a data-driven picture of why establishments fail, which could be used to inform legislation or inspection protocols or developing prediction models for follow-up inspections.
2) This project has demonstrated the potential for a model to predict hygiene compliance. Further work could establish the potential benefits of such a system: given the model accuracy, how beneficial would rolling out such a system be in terms of risk reduction and improved intervention scheduling? This work could also focus on the best way to implement such a system, e.g. what should the risk threshold be? How does varying this threshold affect the costs and benefits, or in which situation should a model detecting 'non-broad compliance' be used compared to a 'fully compliant' prediction model?
3) Implement a trial roll-out of the prediction model with a single LA, as described in the previous section.
4) Undertake an analysis to understand how easy it is for newly registering businesses to provide information about themselves required to implement this system or whether alternatively this information could be sourced from available data.

9. Conclusions

The purpose of this project was to collect a detailed level of data on the food related activities of a large sample of food establishments and to use this data to develop models to forecast how compliant establishments with particular characteristics are likely to be with food safety law. The focus for the work was to establish the framework for prediction models for new establishments who have yet to have a food safety inspection and who do not have any enforcement history. The project has successfully met these objectives.

We have collected a balanced data set of 8,700 establishments from 13 local authorities. It contains detailed, varied and informative information on the food hygiene attributes of these businesses, collected during food hygiene inspections. Using the collected data, we have developed models which predict the hygiene compliance of food establishments with good precision. The model uses logistic regression and is highly interpretable; it gives a clear explanation as to why a prediction has been made and which factors are important for compliance. We applied the same modelling methodology to the problems of predicting:

- Non-broadly compliant premises (rating of 2 or less);
- Fully compliant premises (rating of 5);
- Non-broad compliance for restaurants only.

In each of these cases the developed models demonstrate a high precision at different levels of risk aversion and produce practically useful results.

For example, consider the performance of the non-broadly compliant (NBC) model for the City of London data set (see figure 18). The unseen test data contains 89 establishments, of which 10 are NBC.

Using the less risk averse model, where the focus is on ensuring that any prediction that an establishment will be NBC is correct (at the cost of some NBC establishments being predicted as broadly compliant), the model:

- correctly predicts that 4 of the 10 NBC establishments will be NBC when inspected.
- incorrectly predicts that the other 6 NBC establishments will be broadly compliant (BC) when inspected.
- correctly predicts the inspection outcome of the 79 BC establishments.

Using the more risk averse model, where the focus is on ensuring that any NBC establishments are predicted to be NBC (at the cost of some BC establishments being predicted as NBC), the model:

- correctly predicts that 9 of the 10 NBC establishments will be NBC at inspection.
- incorrectly predicts 1 NBC establishment will be broadly compliant (undesirable, but less of a risk compared to the model above).
- incorrectly predicts that 25 of the 79 broadly compliant establishments will be NBC i.e. over predicts the risk of these 25 establishments.

The models demonstrate good ability to distinguish compliant establishments from non-compliant establishments and could be used to effectively prioritize inspections.

In comparison with other tested models, we found that logistic regression outperformed decision trees and suffered only a small performance penalty compared to more complex but less interpretable models including XGboost and random forests. Optimal model parameters for the logistic regression models were identified using cross-validated tuning. The developed models are provided in the supporting model files. The most important predictors for each LA dataset are provided in Appendix D. Overall, we find that the strongest predictors of non-compliance are processes and procedures around the food safety management system, training, personal hygiene, supplier assurance, whether the establishment is a take-away, and approaches to cleaning.

Appendix A – Data Collection Findings

All 11 Northern Ireland Unitary Authorities did not participate. We directly contacted 43 LAs and were able to speak to relevant officers at 33 LAs about their data.

The inspection data was held in the formats shown in Table 6.

Table 6 Inspection data format, contacted LAs

Inspection Data Format	Number of LAs
Tabular data from inspection app	4
Typed digital document	4
Scanned document	23
Paper	2
Total	**33**

Supporting the findings of the survey, we found that the majority of inspections are undertaken with inspectors filling out a document template on paper, by hand. The majority of LAs scan these and upload them to a management information system or digital file system. We were able to identify 4 LAs actively using smartphone/tablet apps for inspections. These 4 LAs were using an app called iAuditor. Having surveyed 76 LAs, spoken to 33 LAs, spoken to staff at the FSA, and undertaken market research, we do not believe that there are likely to be a significant number of other LAs actively using apps for inspections.

Aside on the advantages of data from inspection apps

App based inspection data is extractable in spreadsheet format, and thus quickly provides structured, tabular inspection data, circumventing the need for manual data entry. A further advantage is that data is 'digital-first', meaning that at the point of entry an inspecting officer is more often required to enter data in a digital/structured way, e.g. by selecting from a drop-down list, rather than writing in a free-form text box. This is not inherent to a digital system, many paper aide-memoires have check boxes, but compared to writing on paper, the medium is more suited to 'click'-style interactions, over free form writing. Additionally, rules can be enforced, such as 'only select one option' or, 'don't move on until section completed' or 'total score must equal sum of constituents'. These make datasets cleaner, more accurate, and more consistent.

Of the 33 LAs, 20 were unsuitable or unwilling to participate for the reasons shown in Table 7.

Table 7 Reason for LA non-participation

Reason for LA non-participation	**Number of LAs**
Data not suitable (e.g. non-consistent inspection template)	3
Not willing or not able to participate (e.g. too busy)	7
Paper based records (excess effort to extract data)	2
Difficult to extract (e.g. massive amount of human time to collate inspection forms)	4
Low LA engagement	3
Chose not to pursue (as too close to another LA, biasing geographic spread)	1
Total	**20**

The format of the inspection data for the 13 LAs we did collect is shown in Table 8.

Table 8 Data format, selected LAs

Data format	**Number of LAs**
Tabular data from inspection app	4
Typed document	1
Scanned document	7
Digital registration data	1
Total	**13**

Appendix B – Stratified Sample Analysis

The aim was to collect a representative sample of data from Local Authorities, getting a suitable split by: LA type, geographical and rural/urban/mixed whilst maximising the sample size. The sample split compared to the national (the population) split is given in the following tables.

Table 9 Number and percentage of LAs in sample compared to population, by LA type

Local authority type	Population	Sample	Population	Sample	Difference
District Council	195	7	56%	54%	-2%
London Borough	32	3	9%	23%	14%
Metropolitan Borough Council	35	1	10%	8%	-2%
Unitary Authority (England)	55	1	16%	8%	-8%
Welsh Unitary Authority (Wales)	22	1	6%	8%	1%
NI Unitary Authority*	11	0	3%	0%	-3%
TOTAL	**350**	**13**	**100%**	**100%**	

Table 9 shows the split by LA type. The sample contains one LA of each type and a reasonably representative spread across types. London Boroughs are slightly over represented, whilst unitary authorities in England were slightly under represented.

Table 10 Number and percentage of LAs in sample compared to population, by region

Local authority type	Population	Sample	Population	Sample	Difference
East Midlands	40	2	11%	15%	4%
East of England	45	0	13%	0%	-13%
London	32	3	9%	23%	14%
North East	11	0	3%	0%	-3%
North West	37	1	11%	8%	-3%
South East	65	3	19%	23%	5%
South West	37	2	11%	15%	5%
Wales	22	1	6%	8%	1%
West Midlands	29	1	8%	8%	-1%
Yorkshire and The Humber	21	0	6%	0%	-6%
Northern Ireland*	11	0	3%	0%	-3%
TOTAL	**350**	**13**	**100%**	**100%**	

Table 10 show the split by geography. The sample covers a reasonable representation of the geographic spread of LAs, with London slightly over represented and East of England slightly under represented.

* All Northern Irish LAs declined to participate

Table 11 Percentage of LAs in sample compared to population, by rural/urban/mixed[10]

Type of LA	Population	Sample	Difference
Rural	27%	18%	-9%
Urban	56%	55%	-1%
Mixed	17%	27%	10%
Totals	**100%**	**100%**	**N/A**

Table 11 shows the split by rural/urban/mixed. The sample covers each of the categories, and captures a reasonable representation of each, with mixed slightly over-represented and rural slightly under represented.

Given that the sample selection was subject to the variable participation of LAs, and sought to find LAs with data quantities and formats that would maximise the quantity and quality of data collected, overall, the sample covers a reasonable spread for each of the categories.

[10] Using 2011 rural/urban classification from ONS. https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification

Appendix C – Understanding Feature Importance

The feature importance values in this report are the logistic regression coefficient value transformed by the logistic function[11]. This can be considered to be the amount the risk probability increases with a one unit increase in the feature value, when all other variables are controlled for, e.g. not having a Food Safety Management System increases an establishment's risk by 8% (when all other features are zero). Note that the reader should not interpret the given value as summable (as it is in linear regression coefficients). For example, for an establishment with No FSMS of business type Pub Club, the change in risk score is not equal to 0.08 – 0.05. The given values should be considered as depicting the relative importance of features.

Positive feature importance values are those that increase the risk above the default 50% risk position: when all features are zero the risk score is 50%, positive values increase this risk, negative values decrease it.

[11] 0.5 is subtracted from this value to give the change from the default value when all features are zero.

Appendix D – Model Results for Other LAs

Neath Port Talbot

There are 265 establishments in the Neath Port Talbot dataset after cleaning. The inspection proforma is typed document with many check style boxes. The performance is good, with training set[12] average precision of 0.43, and test set average precision of 0.57, this is substantially better than the no-predictive power benchmark (uniform predictor, which had a test set value of 0.17). The performance of the model on unseen test data is shown in Figure 29.

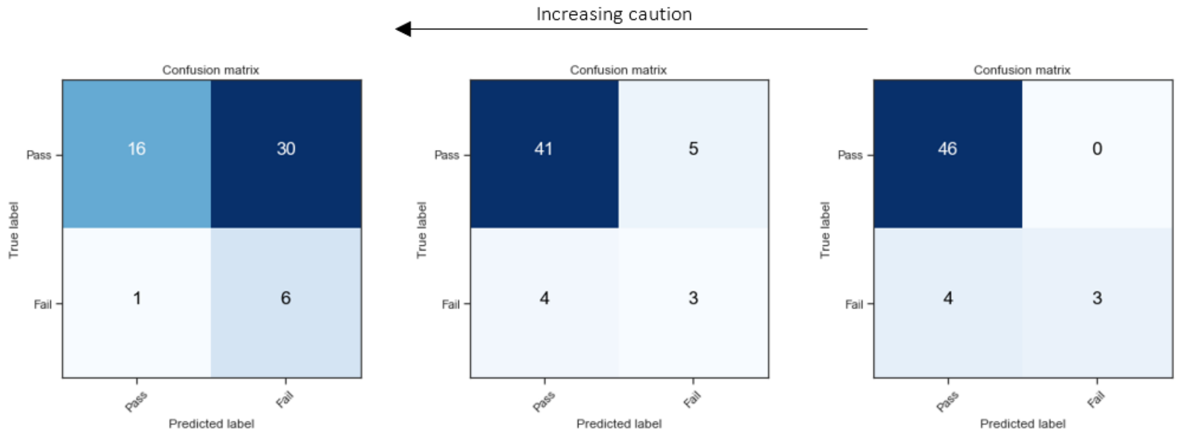

Figure 29 Performance of Neath Port Talbot model. The central confusion matrix maximises the f2 score, (weighted harmonic mean of precision and recall) the other two are selected to be illustrative of higher and lower risk thresholds.

The top predictive features in the model are shown in Table 12. Similar to the City of London model, the top predictors include food safety management system, and other processes. The Neath Port Talbot dataset also includes information on whether the establishments keep records, which is a strong predictor. Although this may not be reportable by new businesses, proxies may be available at registration such as their intention to keep records or quiz questions about their understanding the importance of record keeping. Manager training is the most important feature in the model.

Table 12 Top features - Neath Port Talbot model, negative features reduce the risk of non-compliance.

Feature	**Feature weight[13]**
Adequate training for managers in HACCP training	-0.11
FSMS system cover all foods produced	-0.09
Storage_Records	-0.09
Calibration of probe	-0.08

[12] K-fold cross validation test results in training set.
[13] Logistic(coefficient)-0.5 (same as displayed in previous tables).

Customer allergen info	-0.08
Handling/Prep_Personal hygiene	-0.07
Goods_In_Records	-0.07
Two stage cleaning utilised	-0.06

Charnwood

The Charnwood dataset is from an inspection app. There are 514 establishments in the dataset after cleaning. The model gets an average precision score in the training set of 0.35, above the zero predictive power model of 0.14. In the test set it scores 0.30. the performance is worse than for the Neath Port Talbot and City of London models, but still demonstrates predictive power. The performance may be worse due to the data fields collected by Charnwood, e.g. it does not have a data on whether establishments have a supplier assurance scheme, or data on whether FSMS has means to ensure it is reviewed which were important for the City of London model.

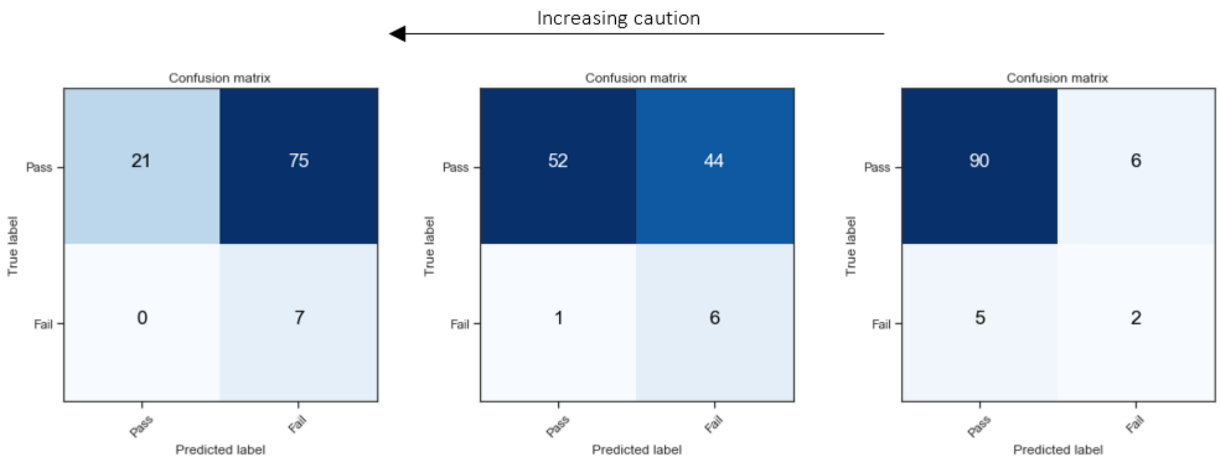

Figure 30 Performance of Charnwood model

Table 13 shows the top features for the model. Similar to the City of London and Neath Port Talbot models the Food Safety Management System is important. Food traceability was also important in the City of London model. The Charnwood model also suggested that probe use and personal hygiene were important.

Table 13 Top features - Charnwood model

Feature	**Feature weight**[14]
Is there a documented food safety system	-0.018
Are staff trained / instructed on effective hand washing	-0.018
Is there traceability on site for food items	-0.018
Probe in use	-0.018
Responsible for structure	-0.017

[14] Weights are lower due to higher regularization

Soap[15]	-0.017
Is there any complex equipment on site	-0.016
FirstLanguageEnglish	-0.015
CleaningCloths_DisposableNotReusable	-0.015

Chiltern

The Chiltern data is from an inspection app. A much larger proportion of the data fields contain questions asking for an inspector's assessment, and contain terms such as "adequate" or "suitable". For example, one field asks the question "Is the probe thermometer working, clean and being used correctly (including calibration)?", unlike in the previous LAs where there is a more self-reportable field: "is a probe in use". These fields are discarded for our purposes as they would not be self-reportable by a business. The results on this dataset are good, on par with City of London and Charnwood, with a mean average precision in the training set of 0.41 (compared to a uniform random "no predictive power" model of 0.22) and test set precision of 0.4. It also outperforms the benchmark of always predicting a fail if the establishment has no FSMS (the results for this are on par with the no predictive power model).
The top features for the model are given in Table 14. Allergen identification, disenfection and FSMS are strong predictors, along with whether a dishwasher is used. Re-usable cloths and having offsite or outside facilties increase the risk.

Table 14 Top features - Chiltern model

Feature	**Feature weight**
Cleaning and structure_Is a dishwasher used	-0.10
Food Inspection - 2_Can staff identify the 14 key allergens when used as ingredients in menu/food items?	-0.10
Cleaning and structure_Is the business using BS EN 1276 and/or EN 13697 compliant disinfectant /sanitiser?	-0.09
WhatFSMS_Own Documented FSMS	-0.09
Cleaning and structure_Is chemical disinfection used to disinfect surfaces or equipment used for raw and ready-to-eat food preparation?	-0.07
Business details_Nature and size of business_Offsite/Outside Facilities	0.04
Business details_Traceability_Traceability records in place?	-0.04
CleaningClothsUsed_Re-usable	0.03

South Kesteven

There are 811 establishments in the South Kesteven dataset after cleaning. The performance is worse than for the previous datasets. The average precision in the training set is 0.32 and 0.37 in the test set. The inspection proforma is less structured for South Kesteven, it does not have

[15] No further information about what this specifically refers to is available

clearly structured check boxes, the data captured is more free form and less clear cut/objective than the previous LAs, we believe this is why results are less good.

The top features are shown in Table 15. Surface cleanliness, along with cleaning processes are important predictors. However, there are some data omissions in the dataset, we believe these are less reliable than previous LA results.

Table 15 Top features – South Kesteven model

Feature	Feature weight
SurfacesStructures_Cleanliness	-0.16
PersonalHygiene_Protective Clothing	-0.11
SurfacesStructures_Repair	-0.10
WashHandBasins_Soap	-0.07
CleaningDisinfection_Cleaning Schedule	-0.07

Tunbridge Wells

There are 586 establishments in the data set after cleaning. The form is hand written but very well structured. Average precision: training: 0.36, benchmark uniform predictor: 0.07, test: 0.38.

Feature	Feature weight
Documented system based on HACCP?	-0.09
Cooking/preparation	-0.07
Nature of cuisine_other	0.07
Staff training record r.e. SFBB	-0.07
Is equipment washed by hand?	0.07
Chilling	-0.06
Adequate hand washing facilities	-0.06
Cleanliness of probe_dirty	0.06

Stafford

The Stafford forms are shorter and have fewer numerical inputs than most of the others. Average precision: training: 0.27, benchmark uniform predictor: 0.14, test: 0.25.

Feature	Feature weight
HACCP_HACCP	-0.22
CustomerBase_Delivery	0.19
Training_AnyoneHasLevel2	-0.18
CustomerBase_Catering	-0.17
Type of premise_Retail	-0.17

Manchester

Average precision: training: 0.33, benchmark uniform predictor: 0.17, test: 0.46.

Feature	Feature weight
FSMS/HACCP procedures_Are monitoring records kept	-0.16
Business type.1_Take Away	0.12
Open on any weekend days	0.09
Business type.1_Restaurant / Cafe / Canteen	0.09
FSMS/HACCP procedures_any system in use	-0.08
Serve high risk food	-0.07
FSMS/HACCP procedures_Are regular reviews undertaken	-0.07

Sedgemoor

Average precision: training: 0.36, benchmark uniform predictor: 0.15, test: 0.47.

Feature	Feature weight
Is HACCP/FSMS documentation in place	-0.15
Cook-chill activities	0.15
Training records/cert's available	-0.13
Is the FSMS in place commensurate with nature and size of business?	-0.12
Type of premises_takeaway	0.10

Croydon

Average precision: training: 0.44, benchmark uniform predictor: 0.19, test: 0.5.

Feature	Feature weight
Food Safety Management System Documented	-0.11
Antibacterial sanitiser used	-0.09
Allergens_Notice for customers	-0.08
Premise type_home caterer	-0.08
Hand drying.1	-0.0
Use of SFBB_Yes	-0.07

Greenwich

Average precision: training: 0.36, benchmark uniform predictor: 0.15, test: 0.47.

Feature	Feature weight
HACCP/FSMS Assessment_Is documentation in place	-0.15
Cook-chill activities	0.15
Training records/cert's available	-0.13
Is the FSMS in place commensurate with nature and size of business?	-0.12
Type of premises_takeaway	0.10

Appendix E – Original Data Specification

Reference	Questions	Possible answers
0	Local authority name	Plain text
	Establishment data	
1.1	**Establishment Name**	Plain text
1.2	**Unique Establishment Identifier**	Plain text
1.3	**Establishment address**	Plain text
1.4	**Establishment postcode**	Plain text
1.5	**Premises type**	Commercial, Domestic, Mobile, Public building, Other
1.6	**Establishment type (LAEMS classification)**	Primary producers, Manufacturers & Packers, Importers/Exporters, Distributors/Transporters, Supermarket/Hypermarket, Smaller retailers, Retailer - others, Restaurant/Café/Canteen, Hotel/Guest house, Pub/Club, Take away, Caring establishment, School/College, Mobile food unit, Other Restaurants and caterers
1.7	**Establishment type detail descriptor**	Farm-Fruit and vegetable grower, Farm-Livestock, Farm-Arable, Beekeeper, Honey maker, Hunting and trapping, Egg processor, Egg producer, Fishing vessel, Farmed fishing, Processing and preserving of meat, Abattoirs, Processing and preserving of fish, crustaceans and molluscs, Purification centres for shellfish, Processing and preserving of potatoes, Manufacture of fruit and vegetable juice, Other processing and preserving of fruit and vegetables, Manufacture of oils and fats, Operation of dairies and cheese making, Manufacture of ice cream, Commercial bakery, Manufacture of prepared meals and dishes, Manufacture of homogenised food preparations and dietetic food, Manufacture of other food products n.e.c., Manufacture-Alcoholic drinks, Manufacture-Soft drinks, mineral waters and other bottled waters, Packer-Mineral waters, Packer-Contract packers, Food delivery service - Deliver food to consumers, Food delivery service - Process food order only for consumers, e.g. Just

		Eat, Storage provider, Food broker, Wholesaler, Cash and carry, Haulage company, Internet only food, Supermarket, Convenience store/mini-market/corner shop, Farm - Gate sales, Farm shop, Confectionery/sweet shop, Butcher (retail only), Fishmonger, Greengrocer/fruiterer, Health food shop, Bakers shop (retail only), Newsagent, post office, Market stalls, Off licence, Petrol station/Garage, Delicatessen, Chemist, Retailers, Vending machine, Restaurant/Cafe/Canteen/Fast food restaurant, Hostel or B&B, Hotels, Pubs - Meals, Pubs - snacks and drinks only, Take away - no food consumed on site, Nursing, care homes, day centres, Hospitals, Childminder, Childcare, Nursery; pre/after school care; play group etc-Meals, Educational establishments (schools, colleges, university), Mobile retailer, Mobile caterer, Movable food establishment, Contract caterer, Home caterer, Meat cutting plant/catering butcher, Auction hall (fish)
1.8	**Do any other businesses trade from the same premises?**	Yes, No
	Typical opening hours	
1.9	Open weekdays	Yes, No
1.10	Open weekends	Yes, No
1.11.1	Open daytime	Yes, No
1.11.2	Open unsociable hours	Yes, No
1.11.3	Open 24 hours	Yes, No
1.12	**Does the business trade seasonally?**	Yes, No
	If yes, for how many months per year?	number 1-11
1.13	**Primary cuisine type**	African, British, Caribbean, Chinese, Eastern European, French, Greek, Indian, Italian, Japanese, Mediterranean, Mexican, South American , Spanish, Thai, Turkish, Vietnamese , American, Asian, Other, Middle Eastern , Seafood
	Food business operator	**Possible answers**
2.1	**Type of FBO**	Sole trader, Partnership, Limited company, Charity, Unincorporated bodies, Public body, Other
2.2	**Is this a franchise?**	Yes, No

2.3	**Sales activity**	
	Direct to final consumer	Yes, No
	Business to business sales	Yes, No
	Wholesale	Yes, No
	Direct to final consumer - Internet sales only	Yes, No
	Registration	**Possible answers**
3.1	**Date of registration**	dd/mm/yy
	Food activities	**Possible answers**
4.1	**Food handling activities (select all that apply)**	
	Manufactures food	Yes, No
	Handles/prepares open low risk food	Yes, No
	Handles/prepares open high risk food	Yes, No
	Hot holds food for service	Yes, No
	No direct food handling	Yes, No
	Re-wrap/re-pack food and apply their own labels	Yes, No
	Sale of Prepacked Foods	Yes, No
	Bulk Transport of prepacked foods	Yes, No
	Harvest primary products	Yes, No
4.2	**Food types handled (select all that apply)**	
	Ready to eat foods	Yes, No
	Ambient/ shelf stable foods	Yes, No
	Frozen foods	Yes, No
	Chilled foods	Yes, No
4.3	**Number of food handling staff**	Whole number
4.4	**Method of processing (select all that apply)**	
	Canning/aseptic packing low acid foods	Yes, No
	Vacuum packing	Yes, No
	Sous-vide cooking	Yes, No
	Manufacture of cook/chill food	Yes, No
	Fermentation of meats and other foods	Yes, No
	Air drying	Yes, No
	Freeze drying	Yes, No
	Addition of salt/other preserving agents	Yes, No
	Cooking and cooling of meats prior to service	Yes, No
	Manufacture or preparation of uncooked or lightly cooked ready to eat food of animal origin	Yes, No
	Serve high risk uncooked or lightly cooked ready to eat food of animal origin	Yes, No
	Pasteurisation	Yes, No

	Purification of LBM	Yes, No
	Pack food and apply own labelling.	Yes, No
	Supply surplus food for animal feed	Yes, No
4.6	**If the establishment has a FSMS, which type is it?**	SFBB, My HACCP, In house / Company own HACCP, Other FSMS system, Safe catering, Consultant designed system, No FSMS
4.7	**If the establishment has a form of assurance, which type is it?**	In house, Second party, Third party, Primary Authority, Trade Association, Other, No assurance
4.8	**Training**	
4.8.1	Training - managers	CIEH level 1, CIEH level 2, CIEH level 3, CIEH level 4, In house, other, None
4.8.2	Training - staff/food handlers	CIEH level 1, CIEH level 2, CIEH level 3, In house, other, None
4.9	**Water supply**	Mains, Private water supply
4.10	**Waste disposal contract**	Yes, No
4.11	**Pest control contract**	Yes, No
	Import/Export	**Possible answers**
5.1	**Do they import food?**	Yes, No
5.2	**Do they export food?**	Yes, No
	Hygiene inspection data	**Possible answers**
6.1.1	**Date - Intervention 1 (most recent intervention)**	dd/mm/yy
6.1.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
6.1.3	**Potential hazard**	
	Type of food and method of handling	40, 30, 10, 5
	Method of processing	20, 0
	Consumers at risk	15, 10, 5, 0
	Vulnerable group	22, 0
6.1.4	**Compliance**	
	Hygiene Compliance	25, 20, 15, 10, 5, 0
	Structure compliance	25, 20, 15, 10, 5, 0
	Confidence in management	30, 20, 10, 5, 0
	Significant risk	20, 0
6.2.1	**Date - Intervention 2**	dd/mm/yy
6.2.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
6.2.3	**Potential hazard**	
	Type of food and method of handling	40, 30, 10, 5
	Method of processing	20, 0
	Consumers at risk	15, 10, 5, 0
	Vulnerable group	22, 0
6.2.4	**Compliance**	

	Hygiene Compliance	25, 20, 15, 10, 5, 0
	Structure compliance	25, 20, 15, 10, 5, 0
	Confidence in management	30, 20, 10, 5, 0
	Significant risk	20, 0
6.3.1	**Date - Intervention 3**	dd/mm/yy
6.3.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
6.3.3	**Potential hazard**	
	Type of food and method of handling	40, 30, 10, 5
	Method of processing	20, 0
	Consumers at risk	15, 10, 5, 0
	Vulnerable group	22, 0
6.3.4	**Compliance**	
	Hygiene Compliance	25, 20, 15, 10, 5, 0
	Structure compliance	25, 20, 15, 10, 5, 0
	Confidence in management	30, 20, 10, 5, 0
	Significant risk	20, 0
	Hygiene non-compliance - Only answer if not broadly compliant at Intervention 1	**Possible answers**
7.1	**Hygiene**	
	Cross contamination	Yes, No
	Personal hygiene	Yes, No
	Temperature control	Yes, No
	Safe food preparation	Yes, No
7.2	**Structural**	
	Pest activity	Yes, No
	Waste provisions	Yes, No
	Water/lighting/ventilation/drainage	Yes, No
	Design/layout	Yes, No
	Equipment	Yes, No
	Hand washing	Yes, No
	Structure - cleanliness/repair	Yes, No
7.3	**Confidence in management**	Yes, No
	FSMS	Yes, No
	Track record of FBO	Yes, No
	Understanding of hazards	Yes, No
	Training	Yes, No
	Standards	**Possible answers**
8.1	**Which risk assessment model do you use for food standards risks?**	FLCOP, LACORS, NTS, Do not do standards
	Stds inspection data (FLCOP)	
	"This section is only completed if risk assessment model is FLCOP"	

8.1.1	**Date - Intervention 1 (most recent intervention)**	dd/mm/yy
8.1.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
8.1.3	**Potential risk**	
	Risks to consumers & other businesses	30, 20, 10, 0
	Extent to which the activities of the business affect any hazard	30, 20, 10, 0
	Ease of compliance	30, 20, 10, 0
	Consumers at risk	20, 10, 5, 0
8.1.4	**Level of (current) compliance**	40, 10, 0
8.1.5	**Confidence in Management/Control Systems**	30, 20, 10, 0
8.1.6	**Food Standards Risk Category**	A, B, C
8.2.1	**Date - Intervention 2**	dd/mm/yy
8.2.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
8.2.3	**Potential risk**	
	Risks to consumers & other businesses	30, 20, 10, 0
	Extent to which the activities of the business affect any hazard	30, 20, 10, 0
	Ease of compliance	30, 20, 10, 0
	Consumers at risk	20, 10, 5, 0
8.2.4	**Level of (current) compliance**	40, 10, 0
8.2.5	**Confidence in Management/Control Systems**	30, 20, 10, 0
8.2.6	**Food Standards Category**	A, B, C
8.3.1	**Date - Intervention 3**	dd/mm/yy
8.3.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
8.3.3	**Potential risk**	
	Risks to consumers & other businesses	30, 20, 10, 0
	Extent to which the activities of the business affect any hazard	30, 20, 10, 0
	Ease of compliance	30, 20, 10, 0
	Consumers at risk	20, 10, 5, 0
8.3.4	**Level of (current) compliance**	40, 10, 0
8.3.5	**Confidence in Management/Control Systems**	30, 20, 10, 0
8.3.6	**Food Standards Category**	A, B, C
	Stds inspection data (LACORS)	**Possible answers**
	"This section is only completed if risk assessment model is LACORS"	
9.1.1	**Date - Intervention 1 (most recent intervention)**	dd/mm/yy

9.1.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
9.1.3	**National element**	
	Maximum potential risk to the public posed by the business	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5
	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5
	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
9.1.4	**Local Element**	
	What confidence do you have in the business's control systems based on levels of previous and current compliance and knowledge of management's systems of control?	30, 20, 10, 5
9.1.5	**Risk Category**	High, Medium, Low, No risk
9.2.1	**Date - Intervention 2**	dd/mm/yy
9.2.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
9.2.3	**National element**	
	Maximum potential risk to the public posed by the business	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5
	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5
	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
9.2.4	**Local Element**	
	What confidence do you have in the business's control systems based on levels of previous and current compliance and knowledge of management's systems of control?	30, 20, 10, 5
9.2.5	**Risk Category**	High, Medium, Low, No risk
9.3.1	**Date - Intervention 3**	dd/mm/yy
9.3.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
9.3.3	**National element**	
	Maximum potential risk to the public posed by the business	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5

	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5
	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
9.3.4	**Local Element**	
	What confidence do you have in the business's control systems based on levels of previous and current compliance and knowledge of management's systems of control?	30, 20, 10, 5
9.3.5	**Risk Category**	High, Medium, Low, No risk
	Stds inspection data (NTS)	**Possible answers**
	"This section is only completed if risk assessment model is NTS"	
10.1.1	**Date - Intervention 1 (most recent intervention)**	dd/mm/yy
10.1.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
10.1.3	**Hazard Element**	
	What is the maximum potential harm to the public posed by the business?	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5
	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5
	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
10.1.4	**Likelihood of Compliance**	
	What confidence does the assessor have in the business's control systems based on levels of previous & current compliance and knowledge of the management's systems of control?	80, 60, 40, 20, 0
10.1.5	**Risk Category**	High, Upper Medium, Lower Medium, Low, Unrated
10.2.1	**Date - Intervention 2**	dd/mm/yy
10.2.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
10.2.3	**Hazard Element**	
	What is the maximum potential harm to the public posed by the business?	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5
	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5

	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
10.2.4	**Likelihood of Compliance**	
	What confidence does the assessor have in the business's control systems based on levels of previous & current compliance and knowledge of the management's systems of control?	80, 60, 40, 20, 0
10.2.5	**Risk Category**	High, Upper Medium, Lower Medium, Low, Unrated
10.3.1	**Date - Intervention 3**	dd/mm/yy
10.3.2	**Intervention type**	Inspection or audit, Other types of official controls, AES
10.3.3	**Hazard Element**	
	What is the maximum potential harm to the public posed by the business?	30, 20, 10, 5
	To what extent do the activities of the business affect the hazard?	30, 20, 10, 5
	What volume & complexity of legislation does the business need to comply with?	20, 15, 10, 5
	How many consumers are likely to be affected by the business failing to comply?	20, 10, 5, 0
10.3.4	**Likelihood of Compliance**	
	What confidence does the assessor have in the business's control systems based on levels of previous & current compliance and knowledge of the management's systems of control?	80, 60, 40, 20, 0
10.3.5	**Risk Category**	High, Upper Medium, Lower Medium, Low, Unrated
	Standards non-compliance - Only answer if not broadly compliant at last inspection	**Possible answers**
11.1	**Product/Presentation**	
	Chemical Contamination	Yes, No
	Claims	Yes, No
	Composition	Yes, No
	Food fraud/Food crime	Yes, No
	Labelling - FIRs	Yes, No
	Labelling - Other	Yes, No
	Management Controls	Yes, No
	Weights & Measures	Yes, No
11.2	**CIM**	
	Management System	Yes, No
	Leadership	Yes, No

	Competence	Yes, No
	Hazards	Yes, No
	Current Compliance	Yes, No
	Track Record	Yes, No