

RECOMMENDATION

To improve the predictive power of the model, we can improve certain type of features in the database. Some of these are available public data and some may have to be web scraped from websites. But I believe as a government agency, Department of health can directly access some 3rd party data from the hosting companies without going through web scraping process. Most of these features are a mix of categorical and numerical types and have been listed below in the descending order of importance.

1. Yelp Reviews

In reference to the research paper [1] (**Walter 2013**) it shows that sentiment analysis on yelp reviews regarding a particular restaurant can be used to predict the inspection score. The basic idea is that customers who have eaten at a specific restaurant can best describe their service. Although every customer doesn't give reviews but there tend to be a decent number of reviews (both negative and positive) for a restaurant opened for at least a year. The only problem with this dataset is that it is not a verifiable source as it is not collected through any formal process. Data type – Categorical

2. Daily High Temperature

This paper mentions about temperature being a deciding factor regarding hygiene of food served. (**Kannan et al. 2016**) We know that each type of food has a specific temperature to be stored at otherwise it becomes unsuitable for consumption. Outside high temperatures can influence the refrigeration efficiency and it also leads to infestation by rodents. Data type – Numerical, float (in degrees C or F)

3. Food/Cuisine type

A particular food type can have strong correlation with health grade (**Walter 2013**). Seafoods and raw meat tend to go stale much earlier than vegetables. Restaurants serving wide variety of cuisines may also have to take extra steps to ensure food safety. Bakery which contains bread and other dairy items are at highest risk of food spoilage. Data type – Categorical, Nominal

4. Sanitarian Clusters

According to research (**Kannan et al. 2016**), sanitation of a neighborhood has influence on possibility of infestation or sewage problems. It can be a categorical feature with 3 categories – Clean, Moderate and Unclean. The source of this data can be cleanliness complains or cleanliness inspection carried out by public agencies. Data type – Categorical, Nominal

5. Income of the employees

Salary earned by an employee has direct correlation with the organizational commitments (**Khalizani Khalid et al. 2012**). So, if the employees are paid more, they would go an extra mile to maintain cleanliness and follow rules and regulations enforced by FDA. Data type – Numerical, float

6. Net profit earned by the restaurant

Profits earned by the restaurant may have significant effect on the policies they tend to adapt for the future. A low earning restaurant would focus more on marketing and increasing the ambience

of their establishment rather than taking care of food quality and cleanliness of the kitchen. Data type – Numerical, float

7. Crime rate in the area

Crime rate in the area determines the poverty level of the people which can affect the general attitude of a restaurant towards health and safety. Data type – Numerical, float

8. Hours of operations

It can be another deciding factor in the cleanliness or food quality as establishments with longer open hours (say 24/7) would not have well maintained schedule for cleaning in between shifts. Data type – Numerical, Integer

9. Real estate expense in the area

Real estate expense or cost of operation of an establishment can be another factor to determine cost of operation of an establishment. It can be numerical float feature which can be feature scaled to determine its correlation with inspection grade. Although there can be negative or a positive correlation with inspection score. E.g- Expensive real estate price affecting operation cost decreasing the profits while expensive real estate can also mean better upmarket area which means lesser chance of collecting violation codes. Data type – Numerical, float

BIBLIOGRAPHY

[1] Walter (2013). Prediction of NYC Restaurant Health Inspection Results.

<https://cs229.stanford.edu/proj2013/Walter-PredictionOfNYCRestaurantHealthInspectionResults%20pdf.pdf>

[2] Kannan, Vinesh & Shapiro, Matthew & Bilgic, Mustafa. (2019). Hindsight Analysis of the Chicago Food Inspection Forecasting Model. <https://arxiv.org/pdf/1910.04906.pdf>

[3] Anand, Purohit, Kaur & Goomer. (2018). Predicting Restaurant Health using Yelp data and Government Inspections. <https://rafaelsilva.com/files/teaching/inf-553-fall-2018/007-Predicting-Restaurant-Health.pdf>

[4] Khalizani Khalid and Khalisanni Khalid. (2012). The role of wage and benefit in engaging employee commitment. <https://rafaelsilva.com/files/teaching/inf-553-fall-2018/007-Predicting-Restaurant-Health.pdf>