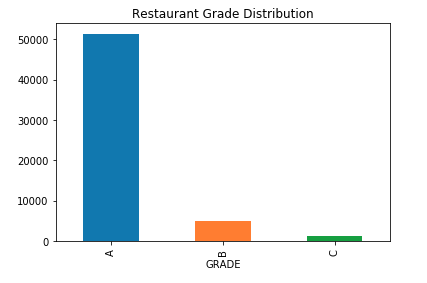
M I trained a Random Forest Classifier to predict NYC Restaurant Health Grades (A,B, or C). I used data from NYC Open Data (“DOHMH NYC Restaurant Inspection Results”), Yelp, and the IRS SOI Tax Stats by Zipcode for 2016.

Features include the yelp rating & # of reviews, price, inspection month and year, cuisine type, and income data by zipcode.

Here is the target class distribution.



The Random Forest model achieved the following results on the test data:

|  |  |  |  |
| --- | --- | --- | --- |
|  | A | B | C |
| Accuracy | 0.889 |  |  |
| Precision | 0.898 | 0.223 | 0.150 |
| Recall | 0.990 | 0.028 | 0.012 |
| F1 | 0.942 | 0.049 | 0.023 |

Due to the class imbalance, it is difficult to predict the B and C grades. My next step will be to use some of the techniques to account for class imbalance as well as to try alternative models.