#Yelp-Ratings-Model

Done by – Naimesh Pramanik (17CS30023)

Introduction

This finds out the correspondance between reviews and ratings on the Yelp database used. The main aim of this project is to find out the words that are mostly found in the reviews with ratings 5 or 1. This is not a multiclass depiction, here its only shown for 2 classes namely rating-5 and rating-1.

Environment used.

Python 3.7

packages/libraries used

numpy

 pandas

 sklearn

command line

python yelp_rating_predict.py

yelp_rating_predict.py - Result

```text
-----------------------------------------------
 Accuracy of  MultinomialNB  -  0.909001956947


 Confusion matrix
 [[119  76]
 [ 17 810]]


 Area Under Curve of  MultinomialNB  -  0.941909279757

-----------------------------------------------

-----------------------------------------------
 Accuracy of  LogisticRegression  -  0.93542074364


 Confusion matrix
 [[147  48]
 [ 18 809]]


 Area Under Curve of  LogisticRegression  -  0.969609028617

-----------------------------------------------
 Finding top rating and non-rating words
 Total Features:  16800
 Total observations in each class  [  554.  2510.]
 -------------------- Top 20 five star rating words --------------------
```

	five_star	one_star	five_star_ratio
token			
fantastic	206	2	22.733865
perfect	235	4	12.967131

flavors	106	2	11.698008
outstanding	53	1	11.698008
yum	52	1	11.477291
favorite	322	7	10.152988
ribs	45	1	9.932271
gluten	44	1	9.711554
mozzarella	42	1	9.270120
bianco	42	1	9.270120
gem	38	1	8.387251
hubby	36	1	7.945817
pasty	35	1	7.725100
amazing	477	14	7.520148
delish	34	1	7.504382
organic	34	1	7.504382
waffles	34	1	7.504382
authentic	66	2	7.283665
superb	33	1	7.283665
die	65	2	7.173307

-------------------- Top 20 one star rating words --------------------

token	five_star	one_star	five_star_ratio
disgusting	1	30	0.007357
remove	1	11	0.020065
unprofessional	1	10	0.022072
rude	6	58	0.022833
pointing	1	9	0.024524
inedible	1	9	0.024524
flag	1	9	0.024524
horrible	8	71	0.024870
refused	2	17	0.025967
hubster	1	8	0.027590

acknowledged	1	8	0.027590
ants	1	8	0.027590
fedex	1	8	0.027590
ugh	2	16	0.027590
fuse	1	8	0.027590
boca	1	8	0.027590
unacceptable	1	8	0.027590
ignored	2	15	0.029429
yuck	2	15	0.029429

```