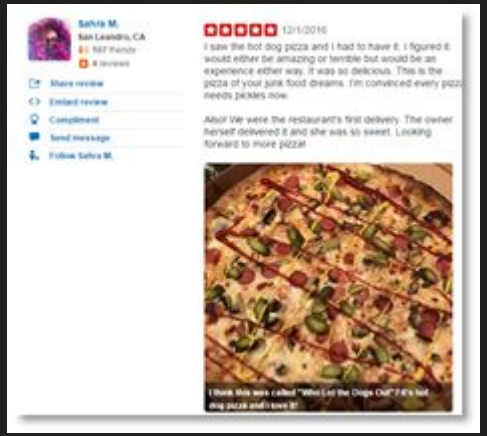
Yelp Text Review Rating Predictions

Yi, Ashley, Umber

Project Goals

- Apply NLP techniques to predict reviews star rating on yelp
- Building off of non NLP ML models
- Accuracy analysis



Background & Motivation

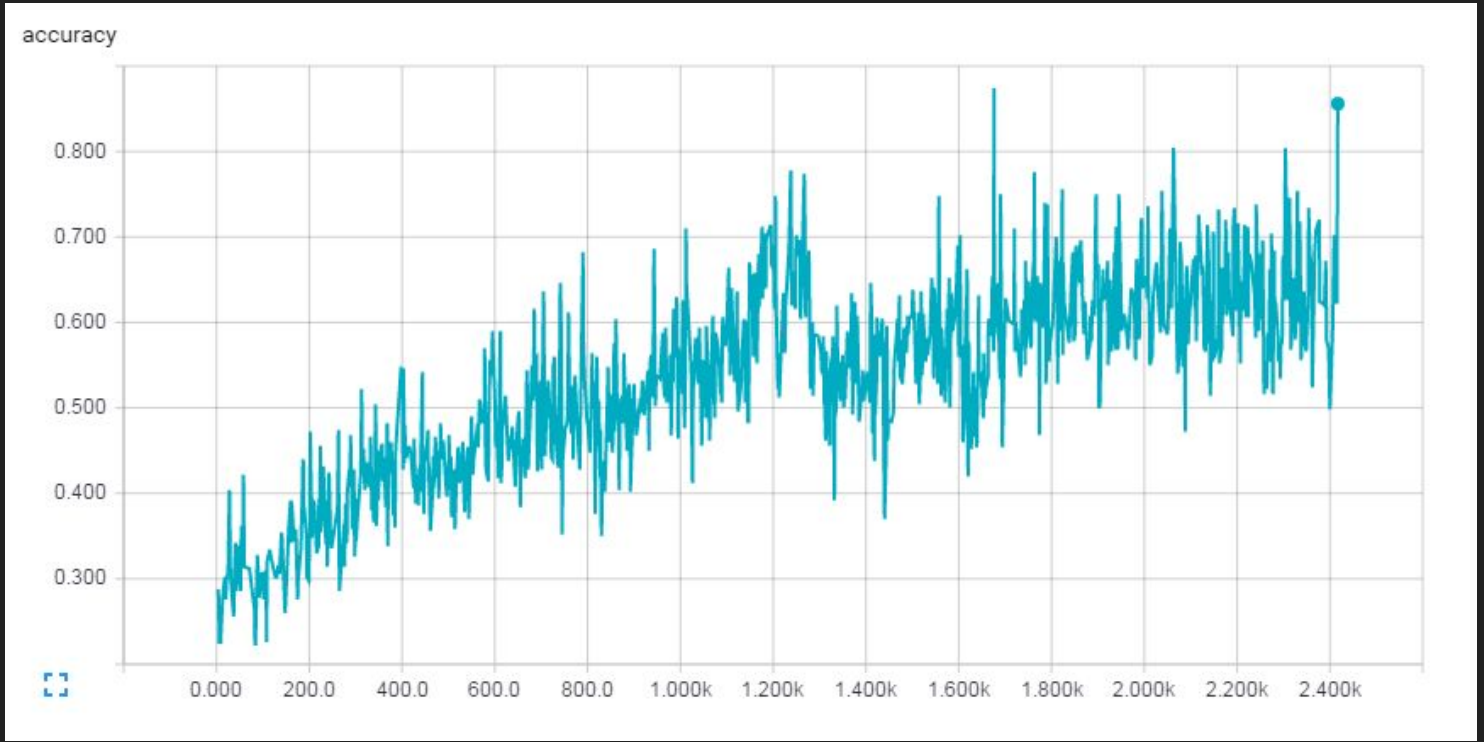
- Yelp dataset Challenge
  - 2.7 M reviews
  - 687 K users
  - 85 K businesses
- Analysis of words/methods to predict review stars

```
{
    'type': 'review',
    'business_id': (encrypted business id),
    'user_id': (encrypted user id),
    'stars': (star rating, rounded to half-stars),
    'text': (review text),
    'date': (date, formatted like '2012-03-14'),
    'votes': {(vote type): (count)},
}
```

Design & Implementation

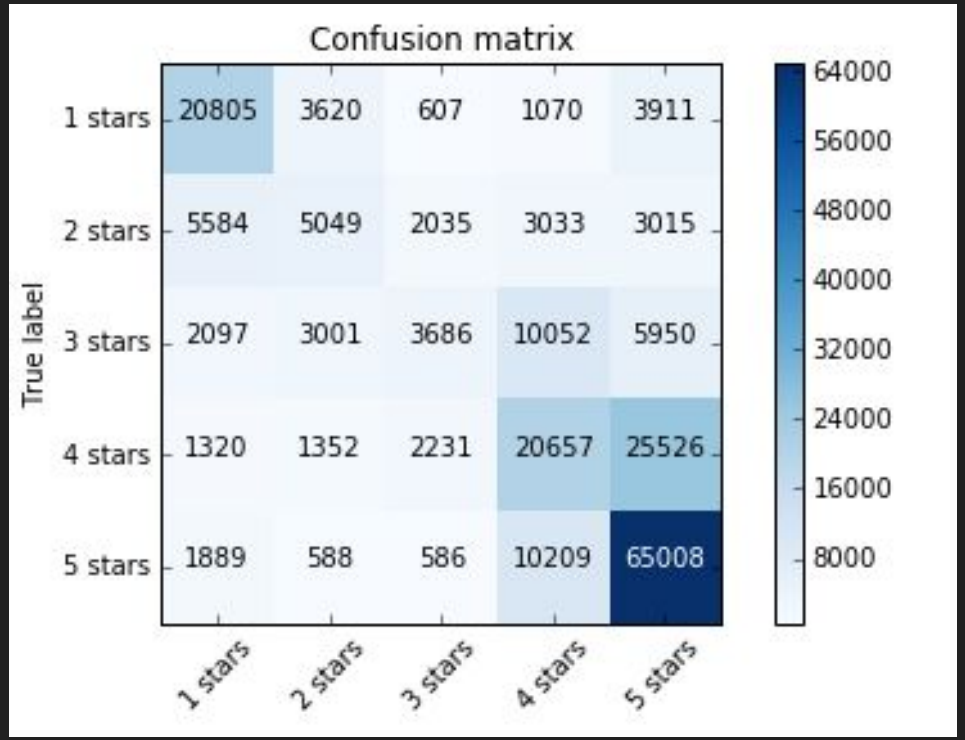
- Baseline
  - Bag of words, Multinomial Naive Bayes
- CNN
  - CNN focuses on word segments, automatic feature extraction
- Parameter tunings
  - Vocabulary size
  - Number of filters
  - Embedding size
  - Filter sizes
- Data pre-processing
- Dataset
  - Training: 1.7 M
  - Dev: 89 K
  - Test: 890 K

Results & Conclusions

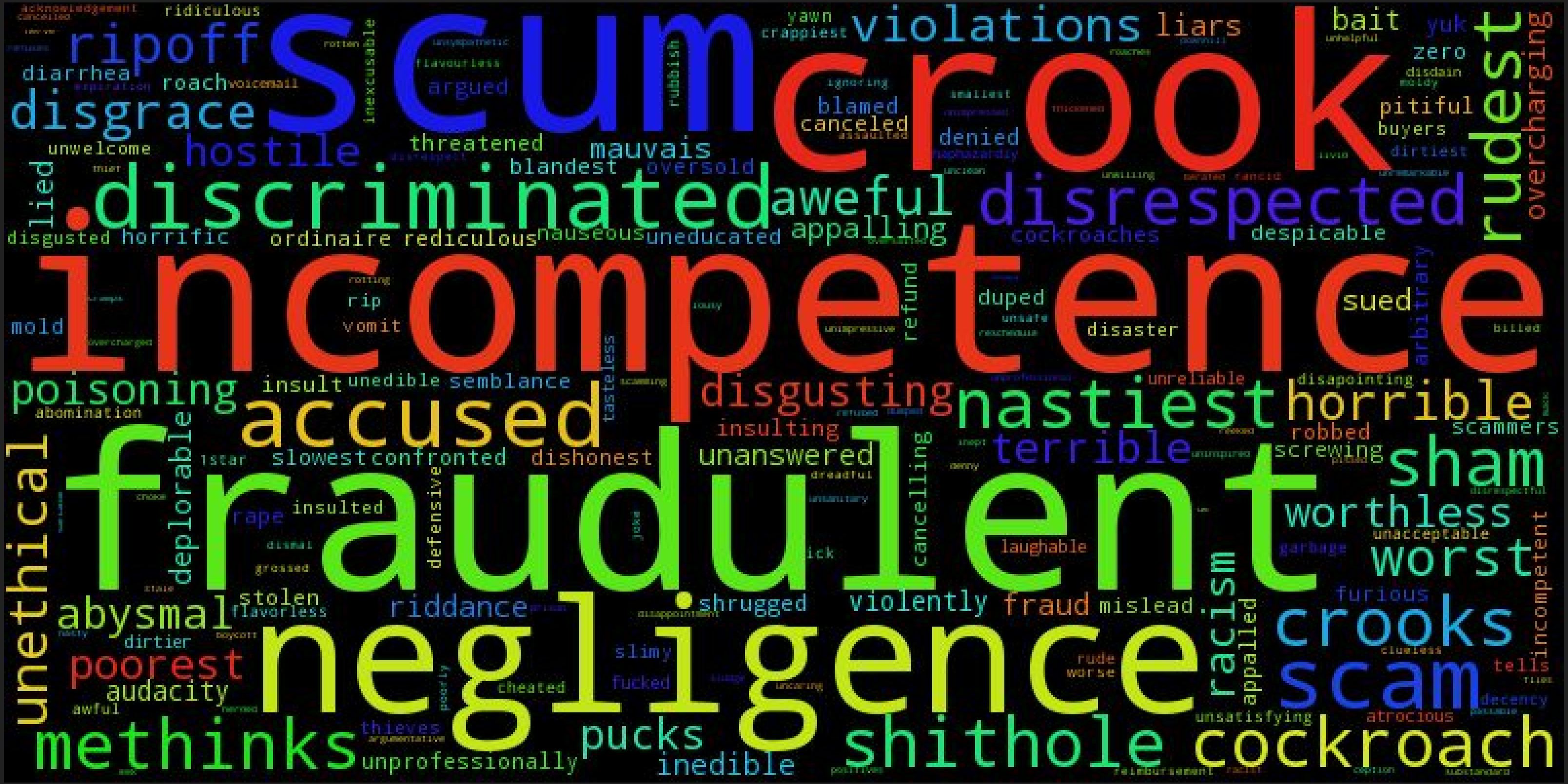


Accuracy in training (batch_size:500)

Results & Conclusions

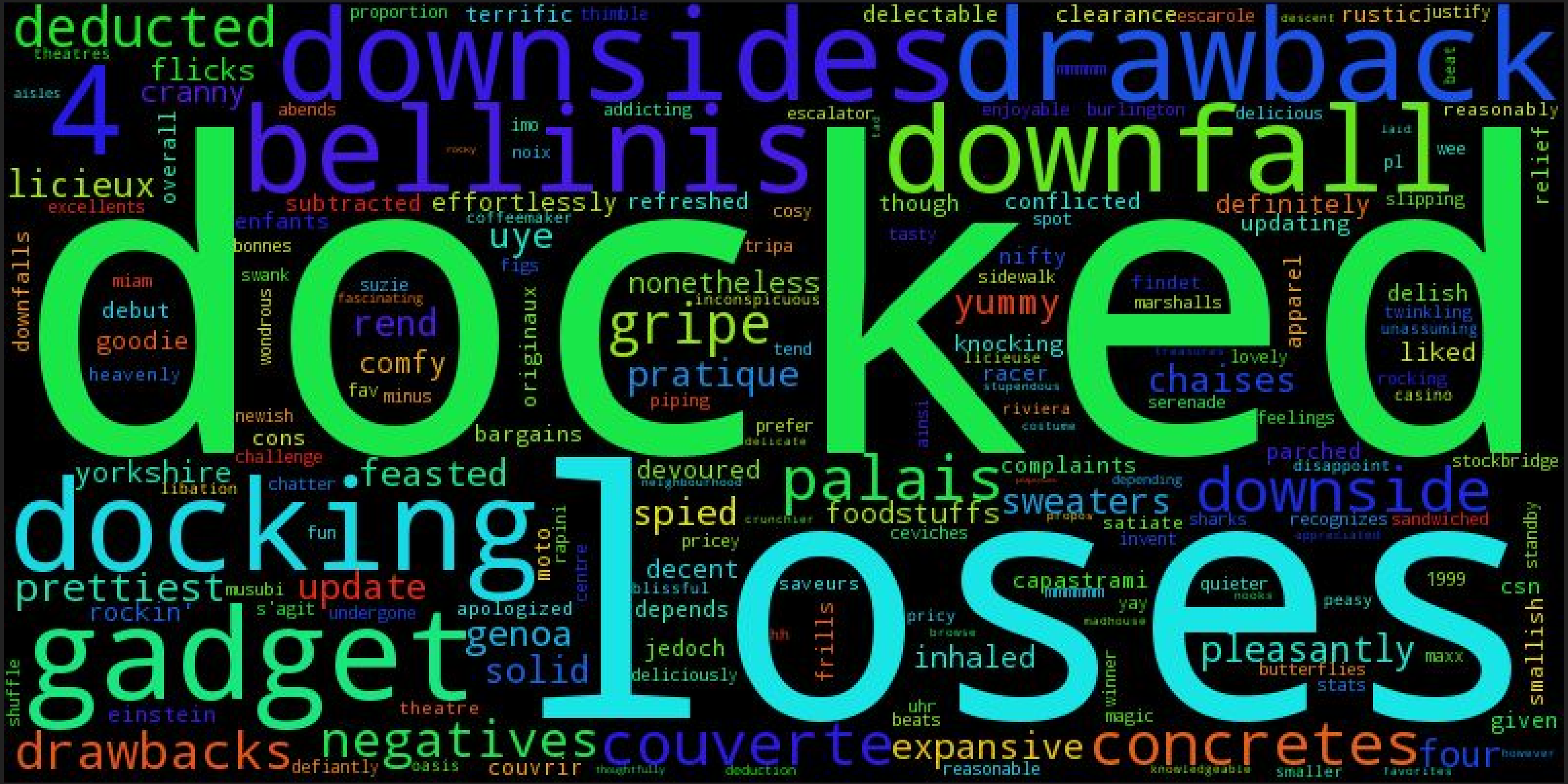


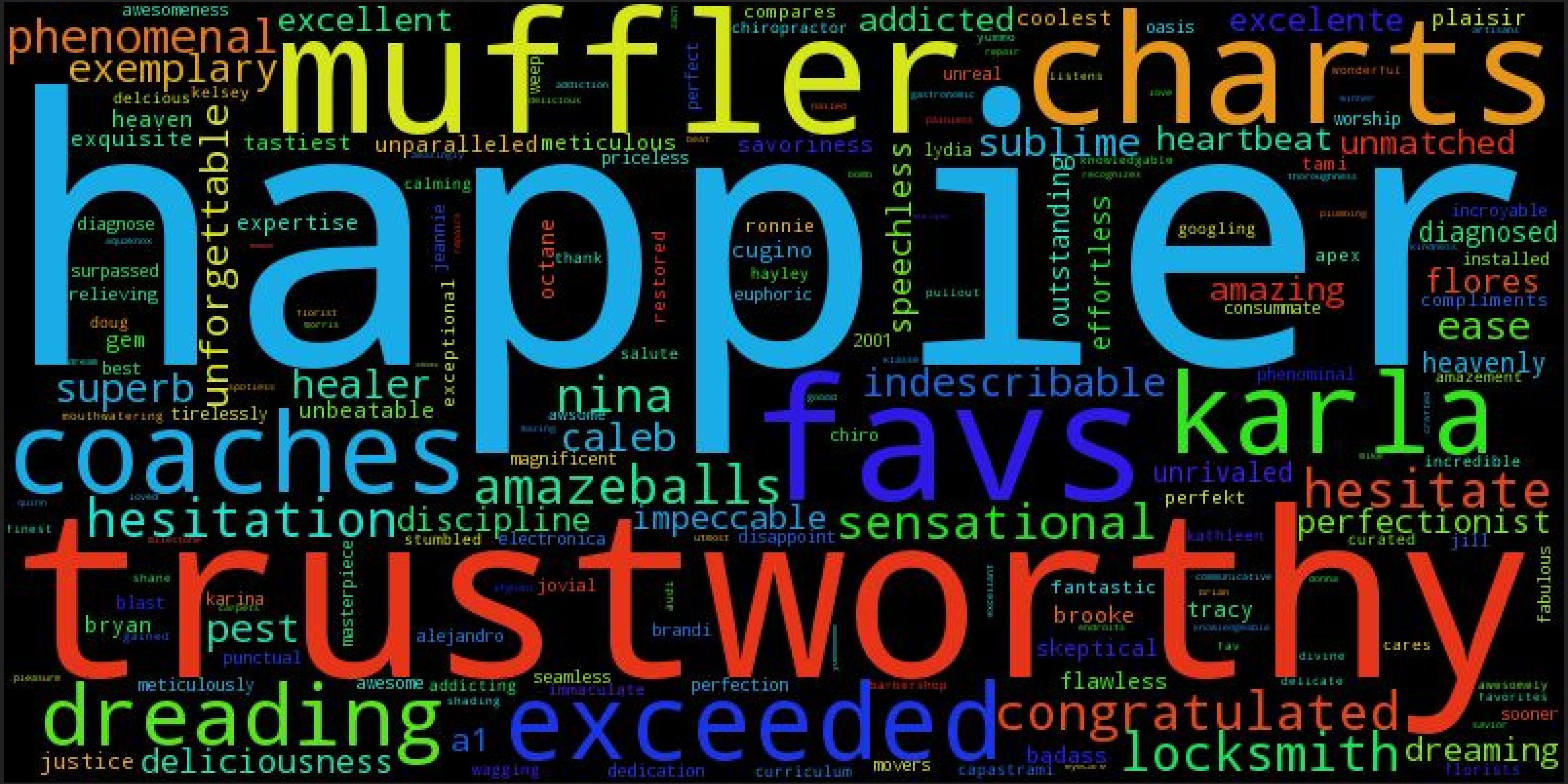
Results on 200K test set


scum
crook
incompetence
fraudulent
negligence
discriminated
disrespected
aweful
ripoff
violations
liars
bait
yuk
rudest
overcharging
disgrace
hostile
accused
nastiest
horrible
disgusting
poisoning
unethical
deplorable
abysmal
poorest
methinks
pucks
shithole
cockroach
scam
crooks
sham
worthless
worst
terrible
racism
appalled
incompetent
diarrhea
roach
argued
blamed
canceled
denied
pitiful
buyers
dirtiest
zero
disdain
threatened
mauvais
unwelcome
blandest
oversold
lied
disgusted
horrific
ordinaire
rediculous
nauseous
uneducated
appalling
cockroaches
despicable
arbitrary
refund
duped
sued
disaster
mold
rip
vomit
tasteless
insult
unedible
semblance
unreliable
disapointing
abomination
insulting
robbed
scammers
1star
slowest
confronted
dishonest
unanswered
screwing
cancelling
insulted
rape
defensive
laughable
unacceptable
stolen
flavorless
riddance
shrugged
violently
fraud
mislead
furious
dirtier
slimy
fucked
audacity
cheated
awful
thieves
unprofessionally
inedible
unsatisfying
atrocious
tells
decency
ridiculous
acknowledgement
yawn
crappiest
inexcusable
rubbish
voicemail
unhelpful
dreadful
garbage
rude
worse
dismal
grossed
billed
ignoring
smallest
unsafe


yawn
meh
lackluster
underwhelmed
semblance
unimpressed
slowest
duped
passable
moldy
'ok'
uneducated
cockroach
disdain
lacked
overrated
slimy
ehhhh
worthless
lukewarm
atrocious
ordinaire
depressing
unfortunately
negligence
tasteless
pucks
mismanaged
uninspired
rubbish
marginal
shrunk
ick
letdown
grossed
inconvenience
oversalted
disapointing
damit
argued
overpriced
stale
flies
unpleasant
mediocre
shrugged
liars
herded
riddance
inexperienced
wack
implying
indifferent
unwelcome
defensive
leider
belittled
mauvais
diminished
yell
substandard
downhill
underwhelming
ception
dismal
unsafe
lacking
disgusting
unimpressive
unacceptable
flavourless
appalling
denny
saddens
dirtier
flawed
disturbing
inexcusable
whopper
fraud
blamed
unripe
flavorless
insulted
unspectacular
violently
desired
abysmal
insult
flagged
rancid
absurd
disappointing
overcharging
oversold
insulting
chipped
dirtiest
hopes
bland
expiration
thieves
poorly
spotty
inedible
horrific
crook
deplorable
worse
decency
soupy
smallest
forgettable
unremarkable
redeeming
inept
ruining
haphazardly
unwilling
acknowledgement
confronted
disappointment
violations

ok
ehhhh
updating
cependant
meh
underwhelmed
lacking
corrected
leider
desired
lacked
flagged
docked
soso
marshalls
downsides
frustrating
uninterested
mediocre
forgettable
unspectacular
underwhelming
unimpressive
unimpressed
overrated
misses
marginal
however
passable
bummer
overpriced
letdown
unfortunately
hyped
inconvenience
lackluster
decent
inconsistent
feasted
underdone
shrunk
glorified
disdain
disapointing
moldy
saddens
whelmed
inefficient
disappointing
downhill
downfalls
slowest
crumbling
unmemorable
serviceable
lukewarm
dissapointing
subtracted
refundable
loudest
toutefois
diminished
spotty
smallish
forbid
jedoch
damit
unripe
denny
inattentive
depressing
unacceptable
justify
ception
flavorless
skimped
inconsiderate
improve
negatives
uye
loses
ensalada
inconsistency
dismal
potential
absurd
eh
ugh
yell
hence
cafeteria
pretends
inexperienced
outskirts
unremarkable
cougars
pared
watery
whopper
subpar
aloof
pricey
factor
lacks
blech
wouldve
unsatisfied
flavourless
malheureusement
adequate
oversalted
haphazardly
belittled
dirtiest
underage
registers
mismanaged
uncomfortably
frou
mish
cons
novelty
goodwill
update
errors
bitchy
yuck
flawed
disturbing
improves
bland
hopes
luke
smallest
marred
ugliest
pricy
depends
uncalled
gossiping
yawn
downfall
cursory
aight
pisces
centre
pinch
blandness
lackadaisical
maxx
deuxi
annoyance
maybe


docked
loses
downsides
drawback
bellinis
downfall
docking
gadget
deducted
4
palais
downside
concretes
couverte
negatives
drawbacks
expansive
four
gripe
pratique
pleasantly
chaises
sweaters
foodstuffs
spied
devoured
feasted
yorkshire
prettiest
update
genoa
solid
nonetheless
yummy
rend
comfy
uye
flicks
cranny
licieux
effortlessly
refreshed
subtracted
enfants
conflicted
definitely
updating
though
delectable
clearance
rustic
justify
proportion
terrific
escarole
reasonably
delicious
enjoyable
burlington
addicting
escalator
abends
imo
noix
overall
relief
wee
pl
slipping
cosy
tripa
tasty
spot
nifty
sidewalk
findet
marshalls
apparel
delish
twinkling
unassuming
liked
rocking
casino
lovely
knocking
racer
serenade
feelings
riviera
piping
tend
inconspicuous
bonnes
swank
suzie
figs
originaux
wondrous
miam
debut
goodie
heavenly
downfalls
fav
minus
newish
cons
bargains
challenge
prefer
excellents
coffeemaker
fascinating
complaints
parched
disappoint
stockbridge
recognizes
sandwiched
satiate
sharks
invent
ceviches
pricey
decent
blissful
depends
jedoch
deliciously
saveurs
frills
inhaled
capastrami
quieter
peasy
1999
csn
smallish
standby
yay
pricy
maxx
butterflies
stats
given
winner
magic
uhr
beats
theatre
einstein
rockin'
s'agit
undergone
apologized
musubi
fun
moto
rapini
centre
chatter
libation
shuffle
defiantly
oasis
couvrir
deduction
reasonable
smaller
aisles
theatres
thimble
ainsi
depending


happier
trustworthy
muffler
charts
favs
karla
coaches
exceeded
dreading
locksmith
congratulated
hesitate
hesitation
amazeballs
sensational
phenomenal
exemplary
excellent
addicted
coolest
excelente
plaisir
unforgettable
sublime
heartbeat
unmatched
indescribable
superb
healer
nina
caleb
impeccable
discipline
perfectionist
unrivaled
speechless
outstanding
effortless
amazing
flores
ease
heavenly
diagnosed
dreaming
deliciousness
pest
a1
justice
compares
chiropractor
awesomeness
oasis
worship
exquisite
heaven
tastiest
unparalleled
meticulous
savoriness
lydia
priceless
delcious
kelsey
diagnose
expertise
surpassed
relieving
gem
doug
best
unbeatable
tirelessly
exceptional
jeannie
octane
thank
restored
ronnie
cugino
hayley
euphoric
googling
apex
installed
incroyable
compliments
consummate
salute
2001
phenominal
magnificent
chiro
perfekt
kathleen
jill
curated
incredible
fabulous
stumbled
electronica
disappoint
masterpiece
jovial
karina
blast
bryan
alejandro
brandi
punctual
meticulously
awesome
addicting
seamless
immaculate
perfection
flawless
fantastic
brooke
tracy
skeptical
badass
movers
curriculum
dedication
wagging
capastrami
sooner
cares
divine
weep
perfect
unreal
listens
tami
calming
mouthwatering
shane

Resources & References

Github: https://github.com/ashley-dsci/266-Yelp

References:

- https://www.yelp.com/dataset_challenge
- http://cseweb.ucsd.edu/~jmcauley/cse255/reports/fa15/031.pdf
- http://cs229.stanford.edu/proj2014/Chen%20Li,%20Jin%20Zhang,%20Prediction%20of%20Yelp%20Review%20Star%20Rating%20using%20Sentiment%20Analysis.pdf