

# Sentiment Analysis on Yelp Reviews

March 23, 2022

## 1 Importing Libraries

```
[5]: import warnings
warnings.filterwarnings('ignore')
```

```
[1]: import requests
from bs4 import BeautifulSoup as bs
```

```
[2]: from pprint import pprint
import random
import json
```

```
[6]: from nltk.sentiment import SentimentAnalyzer
import nltk.sentiment.util
from nltk.classify import NaiveBayesClassifier
```

## 2 Fetching Reviews through Yelp's API

```
[7]: accessToken = '---'
```

```
[8]: r = requests.get('https://api.yelp.com/v3/businesses/search?
↳location=Toronto&limit=50', headers={'Authorization': 'Bearer {}'.
↳format(accessToken)})
```

```
[9]: print(r.status_code) # status code '200' implies the request was successful
```

200

```
[36]: pprint(r.json()['businesses'][:5])
```

```
[{'alias': 'pai-northern-thai-kitchen-toronto-5',
 'categories': [{'alias': 'thai', 'title': 'Thai'}],
 'coordinates': {'latitude': 43.64784, 'longitude': -79.38872},
 'display_phone': '+1 416-901-4724',
 'distance': 3009.3347434869856,
 'id': 'r_BrIgzYcwo1NAuG9dLbpg',
 'image_url':
```

```

'https://s3-media1.fl.yelpcdn.com/bphoto/qEbOKh9rDmTrmHdaUDDnNg/o.jpg',
  'is_closed': False,
  'location': {'address1': '18 Duncan Street',
                'address2': '',
                'address3': '',
                'city': 'Toronto',
                'country': 'CA',
                'display_address': ['18 Duncan Street',
                                    'Toronto, ON M5H 3G8',
                                    'Canada'],
                'state': 'ON',
                'zip_code': 'M5H 3G8'},
  'name': 'Pai Northern Thai Kitchen',
  'phone': '+14169014724',
  'price': '$$',
  'rating': 4.5,
  'review_count': 3010,
  'transactions': [],
  'url': 'https://www.yelp.com/biz/pai-northern-thai-kitchen-toronto-5?adjust_creative=he8fAQ5dPodHGI8nV-U_jw&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'},
  {'alias': 'seven-lives-tacos-y-mariscos-toronto',
    'categories': [{'alias': 'mexican', 'title': 'Mexican'}],
    'coordinates': {'latitude': 43.6543411559068, 'longitude': -79.4004796072841},
    'display_phone': '+1 416-393-4636',
    'distance': 2236.76924282367,
    'id': 'iGEvDk6hsizigmXhDKs2Vg',
    'image_url':
'https://s3-media3.fl.yelpcdn.com/bphoto/TMB02pQtqOwzXMm_oomjJw/o.jpg',
    'is_closed': False,
    'location': {'address1': '72 Kensington Avenue',
                  'address2': '',
                  'address3': '',
                  'city': 'Toronto',
                  'country': 'CA',
                  'display_address': ['72 Kensington Avenue',
                                      'Toronto, ON M5T 2K1',
                                      'Canada'],
                  'state': 'ON',
                  'zip_code': 'M5T 2K1'},
    'name': 'Seven Lives Tacos y Mariscos',
    'phone': '+14163934636',
    'price': '$',
    'rating': 4.5,
    'review_count': 1358,
    'transactions': [],
    'url': 'https://www.yelp.com/biz/seven-lives-tacos-y-mariscos-toronto?adjust_creative=he8fAQ5dPodHGI8nV-U_jw&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'}]

```

```

um=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'},
  {'alias': 'kinka-izakaya-original-toronto',
   'categories': [{'alias': 'japanese', 'title': 'Japanese'},
                   {'alias': 'pubs', 'title': 'Pubs'}],
   'coordinates': {'latitude': 43.66045, 'longitude': -79.37915},
   'display_phone': '+1 416-977-0999',
   'distance': 2089.1715664300095,
   'id': 'RtUvSW0_UZ8V3Wpj0n077w',
   'image_url':
'https://s3-media3.fl.yelpcdn.com/bphoto/nHFVveLBqeiCGy12R2CnRA/o.jpg',
   'is_closed': False,
   'location': {'address1': '398 Church Street',
                 'address2': '',
                 'address3': '',
                 'city': 'Toronto',
                 'country': 'CA',
                 'display_address': ['398 Church Street',
                                     'Toronto, ON M5B 2A2',
                                     'Canada'],
                 'state': 'ON',
                 'zip_code': 'M5B 2A2'},
   'name': 'KINKA IZAKAYA ORIGINAL',
   'phone': '+14169770999',
   'price': '$$',
   'rating': 4.0,
   'review_count': 1671,
   'transactions': ['restaurant_reservation'],
   'url': 'https://www.yelp.com/biz/kinka-izakaya-original-
toronto?adjust_creative=he8fAQ5dPodHGI8nV-U_jw&utm_campaign=yelp_api_v3&utm_mediu
m=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'},
  {'alias': 'banh-mi-boys-toronto',
   'categories': [{'alias': 'vietnamese', 'title': 'Vietnamese'},
                   {'alias': 'sandwiches', 'title': 'Sandwiches'}],
   'coordinates': {'latitude': 43.64884, 'longitude': -79.39695},
   'display_phone': '+1 416-363-0588',
   'distance': 2834.97359074939,
   'id': 'N93EYZy9R0sdlEvubu94ig',
   'image_url':
'https://s3-media4.fl.yelpcdn.com/bphoto/PegGigxIZ7BfCbzAsXBtXw/o.jpg',
   'is_closed': False,
   'location': {'address1': '392 Queen Street W',
                 'address2': '',
                 'address3': '',
                 'city': 'Toronto',
                 'country': 'CA',
                 'display_address': ['392 Queen Street W',
                                     'Toronto, ON M5V 2A9',
                                     'Canada'],

```

```

        'state': 'ON',
        'zip_code': 'M5V 2A9'},
    'name': 'Banh Mi Boys',
    'phone': '+14163630588',
    'price': '$',
    'rating': 4.5,
    'review_count': 1148,
    'transactions': [],
    'url': 'https://www.yelp.com/biz/banh-mi-boys-
toronto?adjust_creative=he8fAQ5dPodHGI8nV-U_jw&utm_campaign=yelp_api_v3&utm_mediu
um=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'},
    {'alias': 'richmond-station-toronto',
     'categories': [{'alias': 'newcanadian', 'title': 'Canadian (New)'}],
     'coordinates': {'latitude': 43.6513729, 'longitude': -79.379285},
     'display_phone': '+1 647-748-1444',
     'distance': 2911.814072486624,
     'id': 'mZRKH9ngRY92bI_irrHq6w',
     'image_url': 'https://s3-media1.fl.yelpcdn.com/bphoto/NdPPUsq5L7gAJ-
mqr6t7-w/o.jpg',
     'is_closed': False,
     'location': {'address1': '1 Richmond Street W',
                  'address2': '',
                  'address3': '',
                  'city': 'Toronto',
                  'country': 'CA',
                  'display_address': ['1 Richmond Street W',
                                      'Toronto, ON M5H 3W4',
                                      'Canada'],
                  'state': 'ON',
                  'zip_code': 'M5H 3W4'},
     'name': 'Richmond Station',
     'phone': '+16477481444',
     'price': '$$$',
     'rating': 4.5,
     'review_count': 796,
     'transactions': [],
     'url': 'https://www.yelp.com/biz/richmond-station-
toronto?adjust_creative=he8fAQ5dPodHGI8nV-U_jw&utm_campaign=yelp_api_v3&utm_mediu
um=api_v3_business_search&utm_source=he8fAQ5dPodHGI8nV-U_jw'}}]

```

### 3 Scraping Reviews using BeautifulSoup

```

[12]: reviews = []
      for business in r.json()['businesses']:
          reviews_requests = requests.get('https://api.yelp.com/v3/businesses/{}/
↪reviews'.format(business['id']), headers={'Authorization': 'Bearer {}'.
↪format(access_token)})

```

```

review_url = reviews_requests['reviews'][0]['url'] # extracting URL for a
↳specific business
page = requests.get(review_url)
soup = bs(page.content, 'html.parser') # parsing page contents using HTML
↳parser
reviews_list = soup.find_all('div', class_ = lambda x: x and x.
↳startswith('review_')) # list of all reviews
for review in reviews_list:
    reviews.append((review.find('span', class_ = lambda x: x and x.
↳startswith('raw_')).text.rstrip('.'),
↳reviews_requests['reviews'][0]['rating'])) # creating tuple of review and
↳its rating

```

```

[13]: reviews[:5] # checking the first 5 reviews extracted

```

```

[13]: [("A friend recommended this restaurant and I would have waked right by
otherwise. Down a flight of stairs, once you walk in, it's a cozy hub of
bustling activity, feeling intimate, warm, and hip all at the same time. We had
the Thai chicken wings, gai satay, moo ping, and spring rolls to start. The
chicken wings were crispy with just enough sauce and the tamarind came through
well - we ordered a second serving immediately after getting the first. We opted
to dine family style, sharing the Khao Soi, Massaman curry, and the Chef Nuit
Pad Thai. These dishes were all well-executed, with the typical flavours of
lime, coconut milk, pepper, peanut, and tamarind balanced appropriately. Serving
sizes were also great for a group of four to share and get a taste of
everything. Service was quick, considering how incredibly busy it was for a
Saturday night. Would certainly recommend making reservations in advance, as
this place gets PACKED around dinnertime. Prices are only slightly higher than
other Thai restaurants and on-par for the location and quality of food",
5),
("Been here 4 times and each time is absolutely delicious.Taste, portion sizes,
beverages, vibe and wide selection are strengths.It's generally always busy, but
the place is huge and usually they can find you a spot (if you're just 2).I love
to bring friends here",
5),
("Finally got to try Pai after so long! I made reservations two weeks in
advance for a weekday birthday dinner for my friend. My friend got the vegan
khoa soi and she enjoyed it very much. I ordered the vegetarian pad thai (thai
spicy!) ($19) which came with egg and veggies - enoki mushrooms, cabbage,
carrots, bean sprouts, and leek. It came with a side of peanuts.Giving 4 stars
because it was good but I wasn't blown away. I asked for thai spicy and it
tasted quite mild to me. The portion size and taste was great. It also came out
piping hot! Service was also fast and warm. I would return to try their other
dishes- especially their curries. I definitely understand the hype",
5),
("Absolutely worth the hype! Came on Sunday at opening time without a
reservation and were seated right away. The decor inside is very welcoming and

```

bright, the music complemented to make a chill vibe. Our waiter informed us they only accept credit card and have a 90 minute table limit. The thai iced tea is delicious - perfect level of sweetness and crushed ice assures it doesn't get watered down. The pad see ew was flavorful and had a great ratio of noodle to greens to shrimp. The hot chili flakes on the side was a nice addition with a squeeze of lemon. Portion size was large. Pad thai was also excellent! The service is fast and attentive. Best restaurant experience I've had in Toronto so far",

5),

("You would never find this, if it wasn't so well recommended. \xa0Lining up down the stairs, I thought we would be waiting forever. \xa0But once we made it down, they took us in right away and proceeded to take us through this awesome grey bricked underground that opened up into what became a huge restaurant. \xa0The vibe is cool, from the atmosphere to the music to the great staff. \xa0Never mind the incredible northern Thai cuisine. \xa0Thai food like this doesn't appear often. \xa0Don't miss out",

5)]

## 4 Storing Results in JSON

```
[14]: with open('data.json', 'w') as file:
    json.dump(list(map(list, {x: 'Positive' if r > 3 else 'Negative' for (x, r)
    ↪in reviews}.items()))), file) # reviews with rating > 3 are defined to be
    ↪positive, else negative
```

## 5 Extracting Features

```
[15]: review_features = [(x.split(' '), 'Positive' if y > 3 else 'Negative') for (x,
    ↪y) in reviews]
```

## 6 Modeling

```
[25]: random.shuffle(review_features)
training_docs = review_features[:int(len(review_features)*0.25)]
test_docs = review_features[int(len(review_features)*0.25):]

print("Training: %d, Testing: %d" % (len(training_docs), len(test_docs)))

sentim_analyzer = SentimentAnalyzer()
```

Training: 122, Testing: 368

```
[26]: all_words_neg = sentim_analyzer.all_words([nlk.sentiment.util.
    ↪mark_negation(doc) for doc in training_docs])
```

```
[27]: unigram_feats = sentim_analyzer.unigram_word_feats(all_words_neg, min_freq=4)
      sentim_analyzer.add_feat_extractor(nltk.sentiment.util.extract_unigram_feats,
      ↪unigrams=unigram_feats)
```

```
[28]: training_set = sentim_analyzer.apply_features(training_docs)
      test_set = sentim_analyzer.apply_features(test_docs)
```

```
[29]: trainer = NaiveBayesClassifier.train
      classifier = sentim_analyzer.train(trainer, training_set)
      for key,value in sorted(sentim_analyzer.evaluate(test_set).items()):
          print('{0}: {1}'.format(key, value))
```

Training classifier

Evaluating NaiveBayesClassifier results...

Accuracy: 0.9728260869565217

F-measure [Negative]: None

F-measure [Positive]: 0.9862258953168043

Precision [Negative]: None

Precision [Positive]: 0.9728260869565217

Recall [Negative]: 0.0

Recall [Positive]: 1.0

The accuracy obtained through a Naive Bayes Classifier is 97.28%