

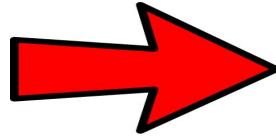


Stock Predictor Based on News Headlines

Michelle Lu, Michael Qu, Kristen Wang

Problem

Based on news headlines about a company one day, can we predict whether the stock goes up or down the next day?





Data Sets

- 1) List of S&P 500 Stocks
(<https://datahub.io/core/s-and-p-500-companies>)
- 2) Pulls news sources for specific companies and the respective dates
(<https://newsapi.org/docs>)
- 3) Historical daily prices and volumes of all current S&P 500 Stocks
(<https://www.kaggle.com/camnugent/sandp500>)



Approach

For each S&P 500 Stocks company:

1. Query the top articles from the News API over a 2 year period

```
In [27]: url = ('https://newsapi.org/v2/top-headlines?'
               'q=Facebook&'
               'from=2016-03-21&'
               'to=2018-03-21&'
               'sources=the-new-york-times&'
               'sortBy=popularity&'
               'language=en&'
               'apiKey=6b33fffeb4cb463f9908694d4be3a532')

response = requests.get(url)
r = response.json()

print ( json.dumps(r, indent=2, sort_keys=True))

{
  "articles": [
    {
      "author": "Cecilia Kang",
      "description": "The Federal Trade Commission is examining whether Facebook violated a 2011 agreement with the agency on data privacy, a person with knowledge of the inquiry said.",
      "publishedAt": "2018-03-21T01:47:18Z",
      "source": {
        "id": "the-new-york-times",
        "name": "The New York Times"
      },
      "title": "New Investigations Into Facebook Add New Pressures",
      "url": "https://www.nytimes.com/2018/03/20/business/ftc-facebook-privacy-investigation.html",
      "urlToImage": "https://static01.nyt.com/images/2018/03/21/business/21FACEBOOK/merlin_135756423_8ac3aeb6-7781-4912-b9e5-fd7210e35170-facebook-jumbo.jpg"
    },
    {
      "author": "The New York Times",
      "description": "The company's board said it was suspending the chief executive, Alexander Nix, with immediate effect, pending an independent investigation.",
      "publishedAt": "2018-03-20T19:45:59Z",
      "source": {
        "id": "the-new-york-times",
        "name": "The New York Times"
      }
    }
  ]
}
```



Approach

2. Run an sentiment analysis using nltk library on each news article headline and retrieve the negative, neutral, and positive values.

```
Great place to be when you are in Bangalore.
neg: 0.0, neu: 0.661, compound: 0.6249, pos: 0.339,

The place was being renovated when I visited so the seating was limited.
neg: 0.147, neu: 0.853, compound: -0.2263, pos: 0.0,

Loved the ambience, loved the food
neg: 0.0, neu: 0.339, compound: 0.8316, pos: 0.661,

The food is delicious but not over the top.
neg: 0.168, neu: 0.623, compound: 0.1184, pos: 0.209,

Service - Little slow, probably because too many people.
neg: 0.0, neu: 1.0, compound: 0.0, pos: 0.0,

The place is not easy to locate
neg: 0.286, neu: 0.714, compound: -0.3412, pos: 0.0,

Mushroom fried rice was tasty
neg: 0.0, neu: 1.0, compound: 0.0, pos: 0.0,
```

The compound value here conveys the overall positive or negative user experience.



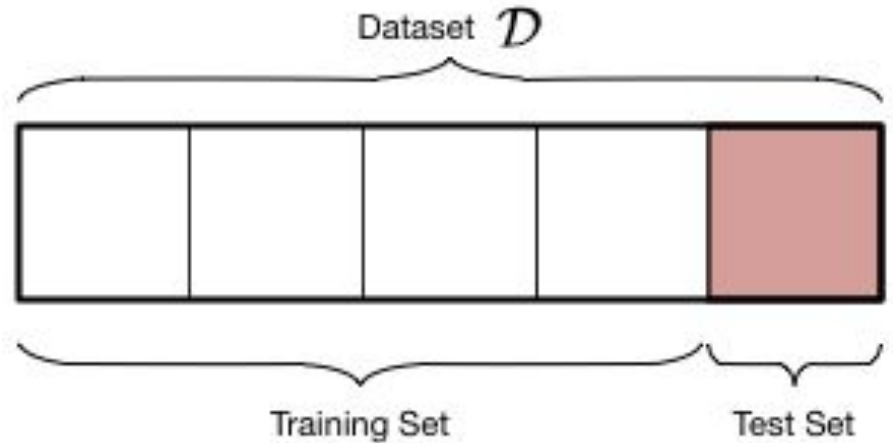
Approach

3. Retrieve stock price change from the day before and after the article was posted

1	Date	Open	High	Low	Close	Volume	OpenInt
2	9/7/84	0.42388	0.42902	0.41874	0.42388	23220030	0
3	9/10/84	0.42388	0.42516	0.41366	0.42134	18022532	0
4	9/11/84	0.42516	0.43668	0.42516	0.42902	42498199	0
5	9/12/84	0.42902	0.43157	0.41618	0.41618	37125801	0
6	9/13/84	0.43927	0.44052	0.43927	0.43927	57822062	0
7	9/14/84	0.44052	0.45589	0.44052	0.44566	68847968	0
8	9/17/84	0.45718	0.46357	0.45718	0.45718	53755262	0
9	9/18/84	0.45718	0.46103	0.44052	0.44052	27136886	0
10	9/19/84	0.44052	0.44566	0.43157	0.43157	29641922	0
11	9/20/84	0.43286	0.43668	0.43286	0.43286	18453585	0
12	9/21/84	0.43286	0.44566	0.42388	0.42902	27842780	0
13	9/24/84	0.42902	0.43157	0.42516	0.42516	22033109	0
14	9/25/84	0.42388	0.42388	0.41618	0.41618	46515020	0
15	9/26/84	0.41618	0.4354	0.41111	0.41111	30947546	0
16	9/27/84	0.41111	0.41366	0.41111	0.41111	29541971	0
17	9/28/84	0.41111	0.41111	0.39316	0.40081	65093531	0
18	10/1/84	0.39956	0.39956	0.39186	0.39186	27268068	0
19	10/2/84	0.39443	0.40853	0.39443	0.39443	32977801	0
20	10/3/84	0.40081	0.40724	0.40081	0.40081	33583772	0

Approach

4. Split data into 80% training set and 20% test set





Approach

5. Run a regression using scikit-learn on the sentiment values and stock price changes. Try different regressions to improve the performance.

X = sentiment values (negative, neutral, positive)

Y = stock price changes



What are the features?

Negative: between 0 and 1, indicates degree of negative connotation in the headline

Neutral: between 0 and 1, indicates degree of neutral connotation in the headline

Positive: between 0 and 1, indicates degree of positive connotation in the headline



Approach

6. Predict stock price changes based on the sentiment analysis of new news articles.



Possible Modifications

- Adding in compound sentiment feature (produced by nltk sentiment analysis, indicates overall positive/negative connotation on a scale of -1 to 1)
- Manually perform sentiment analysis with unigrams, bigrams, etc.
- Adjust number of headlines/articles included
 - E.g. only allowing articles where the company name is in the headline vs. allowing articles where the company name is mentioned





Questions?