The stock market has a major influence on the national economy. Prediction models can decrease volatility of investments and dangers of economic recessions.

Purpose

- Stock market witnesses $77.662 trillion dollars traded worldwide yearly [1]
- Social Media is an endless supply of public data
- Emotion and sentiment are large influences on investors and stocks

Progress

- URL content is extracted by textblob (Natural Language Processing tool)
  - Textblob provides polarity and subjectivity score from 0 to 1
- URL score is combined with twit sentiment score for prediction model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Old Model Accuracy</th>
<th>New Accuracy</th>
<th>Base Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFLX</td>
<td>0.4</td>
<td>0.56667</td>
<td>0.45</td>
</tr>
<tr>
<td>AMZN</td>
<td>0.45</td>
<td>0.51667</td>
<td>0.45</td>
</tr>
<tr>
<td>AAPL</td>
<td>0.35</td>
<td>0.53333</td>
<td>0.3</td>
</tr>
<tr>
<td>TSLA</td>
<td>0.5</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>BAC</td>
<td>0.35</td>
<td>0.58333</td>
<td>0.35</td>
</tr>
<tr>
<td>GS</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>INTC</td>
<td>0.45</td>
<td>0.53333</td>
<td>0.5</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.45</td>
<td>0.45</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Future Work

- Use twitter data in the model with python package Tweepy
- Test different methods/models of using the URL data
- Use more URLs from different users
- Bring in independent news articles as more data
- Use other models developed last year with the URL data
- Create a sentiment analyzer specifically catered to economic vocabulary
  - Use the Loughran and McDonald Dictionary

Acknowledgements

I would like to thank Professor Praveen Madiraju for mentoring me, Joseph Coelho for his technical help and support, and Scott Coyne for the project’s foundation.

This work was sponsored in part by a National Science Foundation REU Supplement to STEM+C grant #CNS-1640217, ‘PUMP-CS: Preparing Urban Milwaukee for Principles of Computer Science’.

Paper Title | Social Media | Other Data
---|---|---
Modular Neural Networks | Not Used | Technical/economic indexes
Tensor Based Model | Term Vectors (“Bag of words”), TF-IDF | Firm-specific and Event-specific
Correlating S&P500 Data | Tweet Counting | Not Used
Twitter Mood Model | OpinionFinder, GPOMS | Not Used
Stock Market Analysis | Multiclass Support Vector Machine | Not Used
Investment Recommendation | Aggregate Sentiment Indexes | Not Used
Social Media Mining Tech. | Net+ sentiment dictionary | Not Used
Microblob Sentiment | Bag of Words, Word Embeddings | Not Used

Reference:

CONTACT
Name: Dawson d’Almeida
Email: dawsond8@gmail.com
Phone: 360-499-1852