Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis

Stuart Colianni, Stephanie Rosales, Michael Signorotti
Stanford Center for Professional Development, Stanford University

Project Goals

The primary goal for this project is to explore how Twitter data can be utilized to develop advantageous crypto coin trading strategies. The objectives are as follows:

- Outline machine learning pipelines with the objective of identifying cryptocurrency market movement through supervised learning techniques.
- Apply supervised learning algorithms such as logistic regression, Naive Bayes, and support vector machines with unique and novel feature vectors.
- Achieve a test set accuracy greater than 80% for predicting when digital currency prices will go up or down.

Method

**Step 1: Collecting the Data**

- Keywords (“bitcoin”, “litecoin”, “ltc”, “lpy”, “xrp”) are searched in Tweets
- Data is pulled via the Tweepy API
- Prices for cryptocurrency collected every hour via Cryptocoin API
- All placed in separate text files

**Step 2: Cleaning the Data**

- Excess whitespace is removed from tweets
- Text is changed to lowercase
- Duplicate tweets, non-alphanumeric characters, and invalid words are removed
- Porter Stemming algorithm used for remaining words in tweets

**Step 3: Creating Models**

- Multinomial Naive Bayes
- Bernoulli Naive Bayes
- Logistic Regression
- SVM with linear kernel
- SVM with RBF kernel
- SVM with polynomial kernel
- SVM with sigmoid kernel

**Step 4: Predicting Prices**

- Determine whether to buy, sell, or keep
- Analyze accuracy of our prediction
- Perform error analysis and improve models

Results

**Bernoulli Naive Bayes for Bitcoin**

- %Hit Correct: 70%
- %Down Correct: 70%
- %Average Correct: 70%

**Logistic Regression for Bitcoin**

- %Hit Correct: 60%
- %Down Correct: 60%
- %Average Correct: 60%

**Multinomial Naive Bayes for Bitcoin**

- %Hit Correct: 65%
- %Down Correct: 65%
- %Average Correct: 65%

Data

**Examples of raw tweets collected**

2015-11-12 00:26:38:626000 bitcoin has no future, says ubs chairman axel weber at bank of england's open forum

2015-11-12 00:21:38:362000 uk chancellor sees a big future for digital currencies #bitcoin #bitcoinnews #btc #cryptocurrency

2015-11-12 00:26:45:489000 #bitcoin, sideways today, but then back down to just above $300 again.

**Examples of prices per hour collected**

2015-11-15 00:49:09:189000 332.65712695 { 'ticker': 'BTC', 'volume': '41988.81016457', 'price': '332.65712695', 'base': 'USD', 'target': 'USD', 'change': '0.96702316', 'success': True, 'error': 'u'

2015-11-15 01:49:21:136000 333.57270673 { 'ticker': 'BTC', 'volume': '43053.76322028', 'price': '333.57270673', 'base': 'USD', 'target': 'USD', 'change': '0.6026489000 #bitcoin, sideways today, but then back down to just above $300 again.

Let’s examine the frequency at which people posted tweets about Bitcoin.

<table>
<thead>
<tr>
<th>Tweet Metric</th>
<th>Median</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processed tweets (after cleaning)</td>
<td>16,000</td>
<td>16,000</td>
<td>2,000</td>
<td>19,000</td>
<td>13,000</td>
</tr>
<tr>
<td>Pre-processed tweets (before cleaning)</td>
<td>1</td>
<td>14</td>
<td>8.2</td>
<td>20,000</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note: After 20 day collection, rounded to 2 significant figures.