

Big Data

Bitcoin Kurse & Twitter

ALICIA HIRT

JAN-PHILIPP PRAETORIUS

TRISTAN KREUZIGER

Zielsetzung

Gibt es einen Zusammenhang zwischen der Stimmung der Twitter User und dem Bitcoin Kurs?

Daten

Twitter Daten, ausgelesen via Tweepy

- #Bitcoin, Bitcoin
- Cryptocurrency
- ...

Output als JSON mit vielen Attributen

 Reduktion auf Datum & Uhrzeit, Text

Zeitraum: 18.01.2018 – 01.02.2018

Etwa 120.000 Tweets

Daten

Bitcoin Kurs von www.coindesk.com

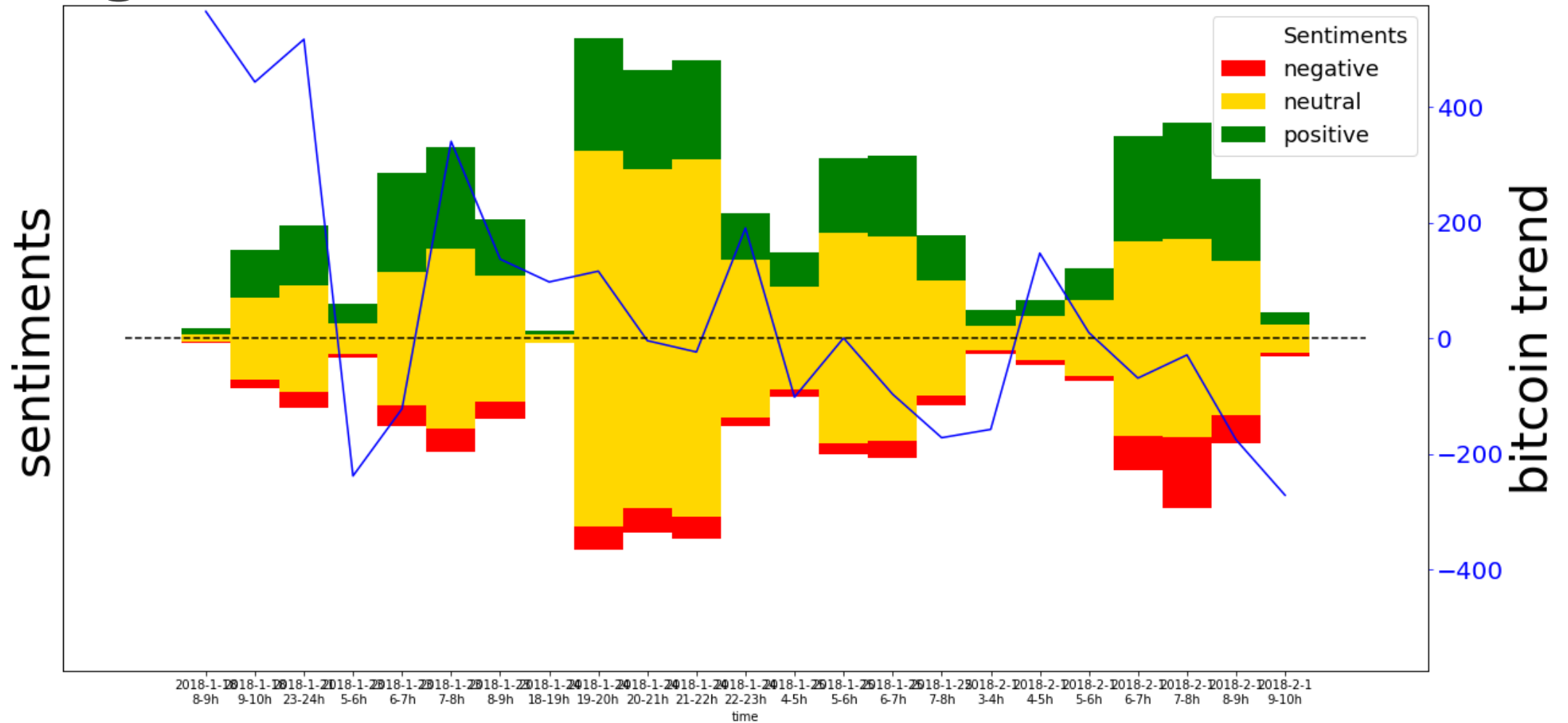
- Datum & Uhrzeit
- Open Preis
- Close Preis

| | A | B | C | |
|----|------------------|------------|-------------|--|
| 1 | Date | Open Price | Close Price | |
| 2 | 2018-01-10 00:00 | 14073,16 | 14073,16 | |
| 3 | 2018-01-10 01:00 | 14073,16 | 14193,43 | |
| 4 | 2018-01-10 02:00 | 14193,43 | 14290,4 | |
| 5 | 2018-01-10 03:00 | 14290,4 | 14139,18 | |
| 6 | 2018-01-10 04:00 | 14139,18 | 14374,32 | |
| 7 | 2018-01-10 05:00 | 14374,32 | 14182,03 | |
| 8 | 2018-01-10 06:00 | 14182,03 | 14240,25 | |
| 9 | 2018-01-10 07:00 | 14240,25 | 14115,24 | |
| 10 | 2018-01-10 08:00 | 14115,24 | 13500,39 | |
| 11 | 2018-01-10 09:00 | 13500,39 | 13790,67 | |
| 12 | 2018-01-10 10:00 | 13790,67 | 13862,86 | |
| 13 | 2018-01-10 11:00 | 13862,86 | 14073,29 | |
| 14 | 2018-01-10 12:00 | 14073,29 | 13898,03 | |
| 15 | 2018-01-10 13:00 | 13898,03 | 13907,03 | |
| 16 | 2018-01-10 14:00 | 13907,03 | 13778,42 | |
| 17 | 2018-01-10 15:00 | 13778,42 | 14420,99 | |
| 18 | 2018-01-10 16:00 | 14420,99 | 14498,19 | |

Spark Workflow

| | |
|--------------------------------|--|
| Daten einlesen | <pre>data_lines = sc.textFile(txt)</pre> |
| Vorverarbeitung | <pre>prepared = data_lines.map(lambda x: split_data_line(x)) converted = prepared.map(lambda x: convert_date(x))</pre> |
| Sentiment - Analyse | <pre>sentiments = converted.map(lambda x: check_sentiment(x)) sentiments = sentiments.map(lambda x: cluster_sentiment(x, levels)) sentiments = sentiments.map(lambda x: cluster_date(x, levels))</pre> |
| Sentiments zählen | <pre>for i in range(len(sent_levels)): temp = s.filter(lambda x: x[1] == i) .groupByKey() .map(lambda x: (x[0], (i, len(list(x[1]))))) final = final.union(temp)</pre> |

Ergebnis



Ausblick

Mehr Daten aus unterschiedlichen Quellen (facebook, RSS)

Konsistenz der Daten

Bessere Sentiment-Analyse

- Bisher: textblob
- eigener Classifier