Boğaziçi University

EC 48T - Term Project

TRADING SIGNAL MODEL USING GDELT

Oğuzhan Pınar, Mehmet Sami Boz, Ömer Aydemir

12.02.2021

WHAT IS OUR GOAL

To obtain day-trading signals by associating the GDELT’s news in its database and the tone scores created according to those news with the volatility in the stock market

WHAT IS GDELT

- Global Database of Events, Language and Tone
- Monitors the world's news media
- Every moment of every day
- Database of over a quarter-billion event records
- More than three quarters of a trillion emotional assessments and over 1.5 billion location references
- A catalog of human societal-scale behavior and beliefs

WHAT ARE WE TRYING TO MAKE MONEY FROM? WHY?

We want to examine all possible Indexes with different volatilities. Because we are not sure how these indexes to react our news shock. We will examine 3 index:

- ***BIST30***
- This Index contains 30 stocks with highest market value and average daily volume. So this Index is most conservative index in the BIST
- ***BIST100***
- This Index is the benchmark of the BIST. It contains top 100 stock with highest market value and average daily volume
- ***BISTTUM***
- This Index contains every stock in the BIST with some minor exceptions. So this index is more volatile than the BIST100

In short; from perspective of volatility:

BISTTUM > BIST100 > BIST30

PRELIMINARY RESULTS

We calculate manipulated news vector’s mean values in every off-session value and if it exceeds the determined threshold we buy at open next day and sell at close price same day.

By using this strategy we gain:

- ➔ 3% compound return in bist30
- ➔ 13.2% compound return in bist100
- ➔ 14.8% compound return in bisttum

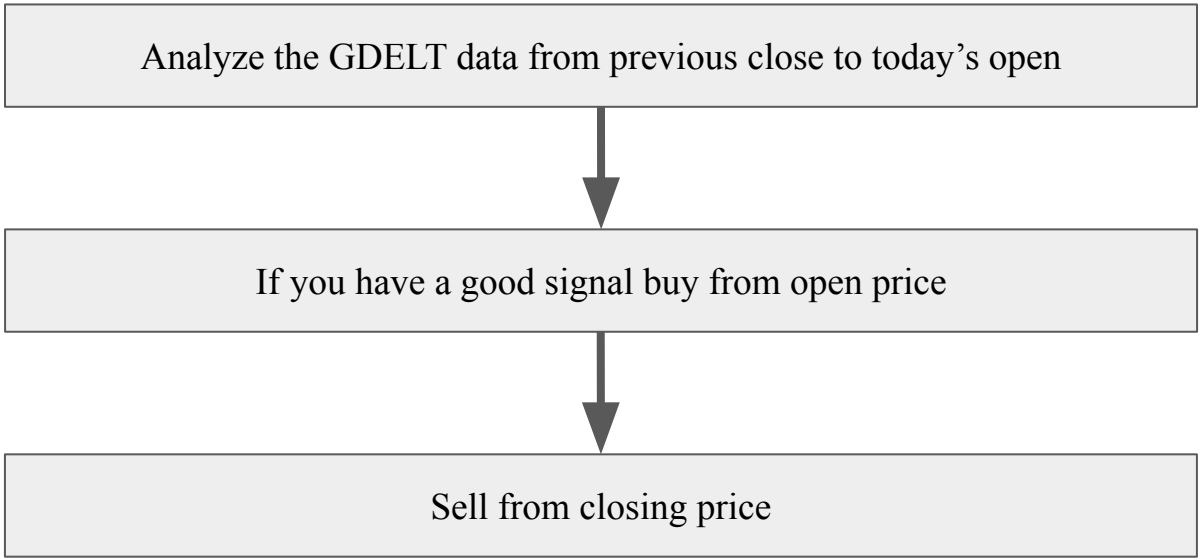
- 1.13% compound interest return for the days that alarm rings

EXTRACTING and MANIPULATING DATA

1. Determine the target themes (we used 193 different themes)
2. Download the global data for each 15 minute from 2015 April.
3. Get the required tone scores of target themes in Turkey.
4. Merge the data and we have 190285 rows × 386 columns.
5. Conduct the PCA and pick enough component to explain 80% of the total variance and than calculated weighted sum of this components and to obtain a vector that mimics the variance of whole news matrix
6. Aggregate the news observations from 15 minute segments to 2 part for whole day: in trading session and off-session.

HOW TO COMBINE GDELT AND STOCKS DATA

1. Decide on the profit making scenario

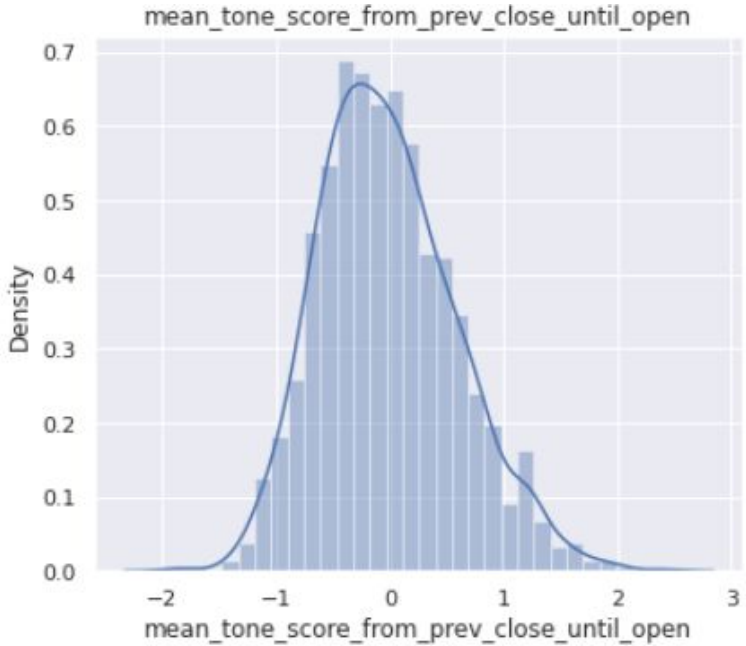


- This is only one scenario, too many other profitable scenarios can be produced.

HOW TO COMBINE GDELT AND STOCKS DATA

2. How to Determine what is a Good signal?

By deciding a Threshold level that gives the best compound return

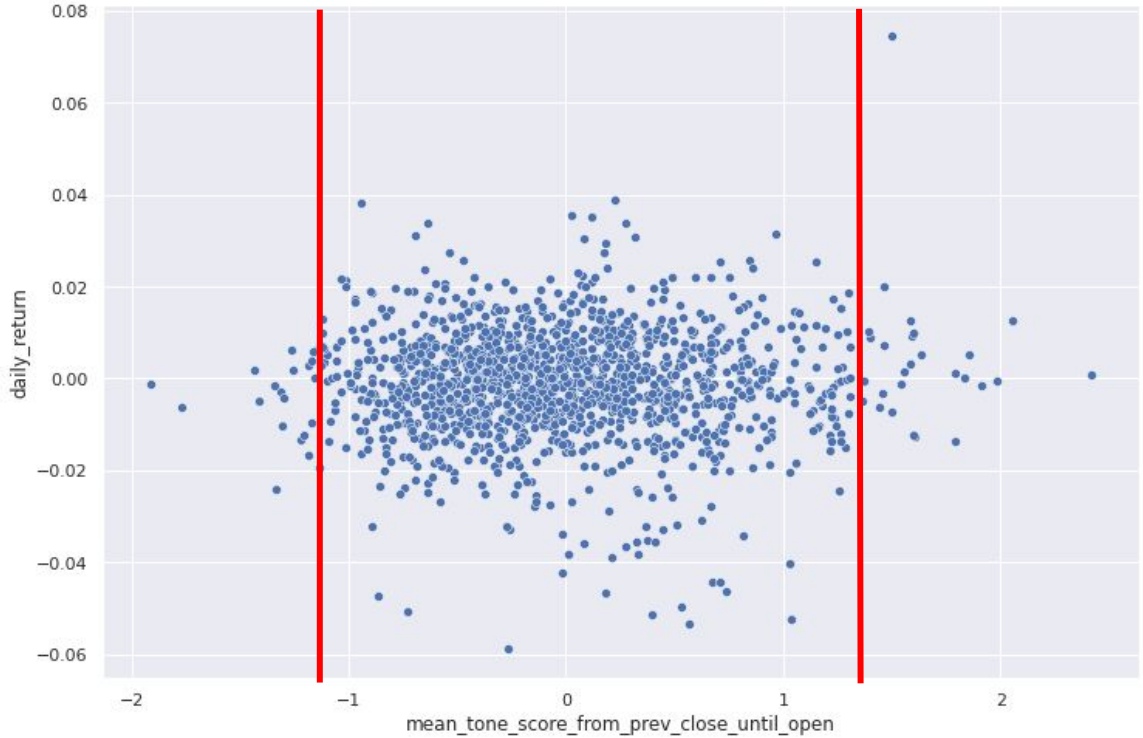


	alarm_count	cumulative_return	mean_return
1.29	34.0	1.148156	1.004176
1.30	32.0	1.136579	1.004119
1.27	36.0	1.133527	1.003591
1.28	35.0	1.130748	1.003624
1.39	25.0	1.126899	1.004920
...	...	...	...
1.59	13.0	1.002864	1.000254
1.60	12.0	0.993632	0.999501
1.02	87.0	0.955017	0.999576
1.01	88.0	0.952534	0.999551
1.00	88.0	0.952534	0.999551

100 rows × 3 columns

HOW TO COMBINE GDELT AND STOCKS DATA

2. How to Determine what is a Good signal?



PERFORMANCE ANALYSIS

	BIST 30	BIST 100	BIST TUM	Daily Interest
best treshold	1.29	1.29	1.29	1.29
total alarm count	34.00	34.00	34.00	34.00
compound_return %	3.03	13.21	14.82	1.13
mean_of_return %	0.05	0.38	0.42	0.03

POSSIBLE IMPROVEMENTS

- Trading session news
- Strategy to exit
- Trade strategy from close to open instead of open to close
- If we could obtain 15 min historical data of BIST, more frequent trades could be done.