

NEWS & STOCK

Using News to Predict Stock Volatility

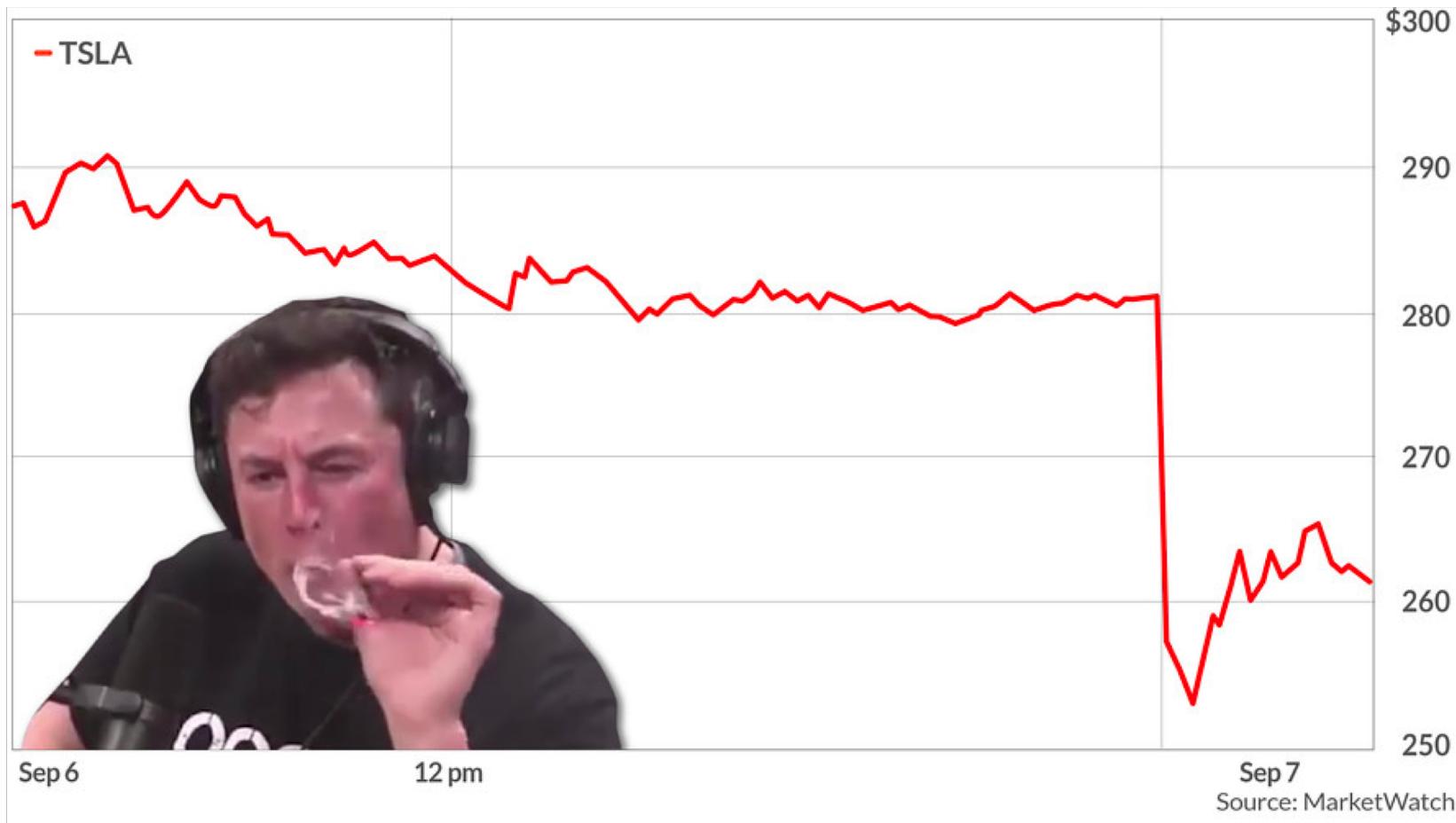
Sibyl (Yimeng) He

M.S., Communication Data Science 2019

University of Southern California

Nov 14, 2018

Case: Tesla Inc

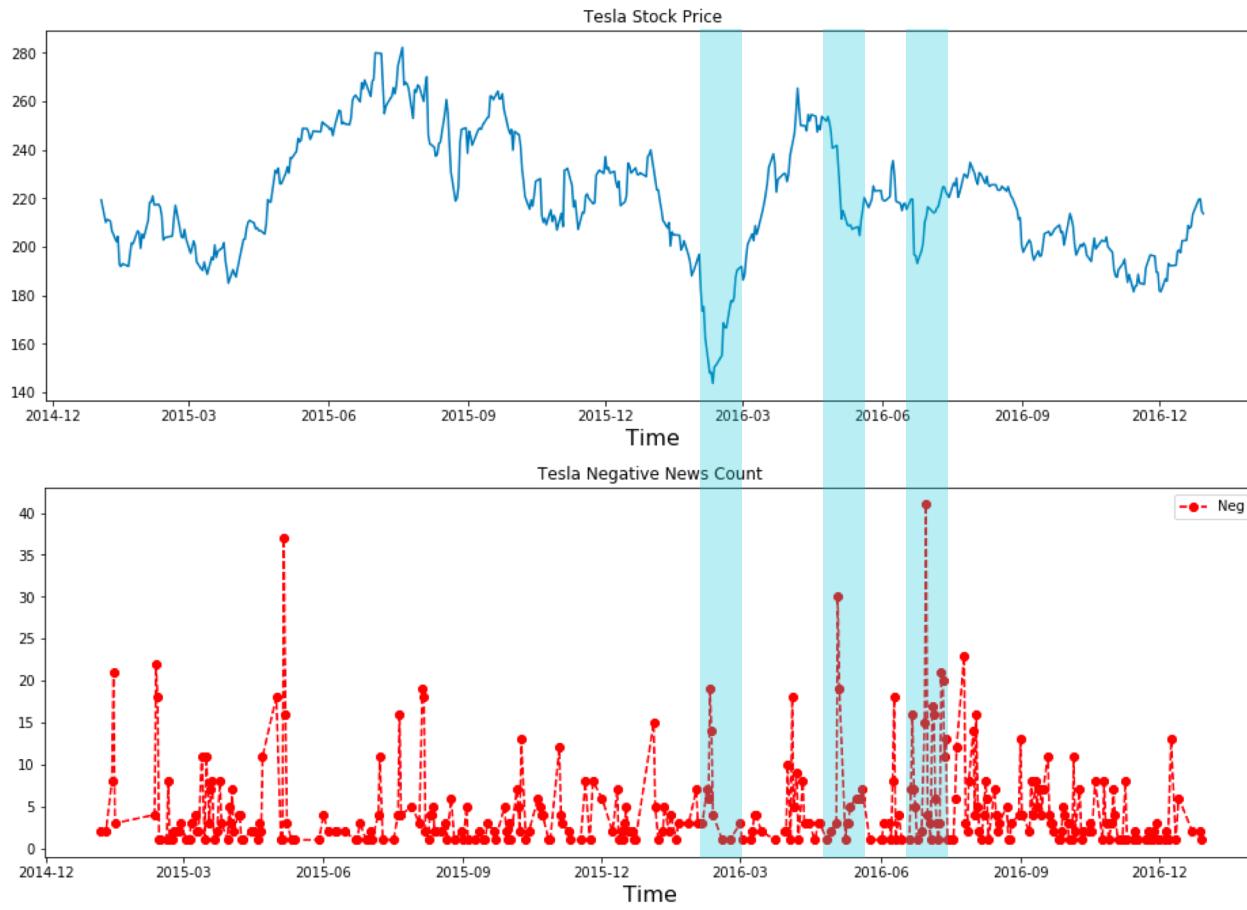


Elon Musk's high and Tesla's low: Stock and bonds tumble on CEO's latest shenanigans

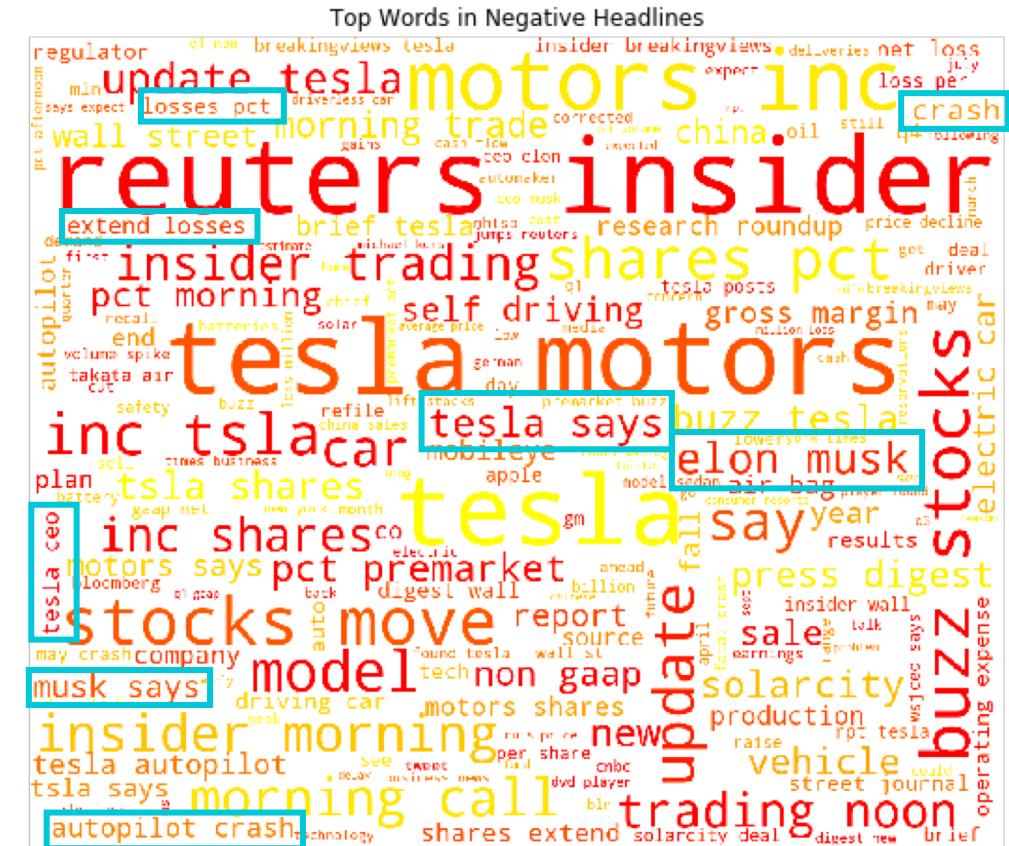
Source: <https://www.marketwatch.com/story/elon-musks-high-and-teslas-low-stock-and-bonds-tumble-on-ceos-latest-shenanigans-2018-09-07>

Case: Tesla Inc

Stock Price vs Negative News Count



Data source: www.kaggle.com/c/two-sigma-financial-news/data



Dataset

Two Sigma: <https://www.kaggle.com/c/two-sigma-financial-news>

3000+ assets, 4M+ market data, 9M+ news data, from 2007-01-01 to 2016-12-31

Market

Time, assetCode, assetName, universe, volume, close, open,
returnsClosePrevRaw1, returnsOpenPrevRaw1,
returnsClosePrevMktres1, returnsOpenPrevMktres1,
returnsClosePrevRaw10, returnsOpenPrevRaw10,
returnsClosePrevMktres10, returnsOpenPrevMktres10,
returnsOpenNextMktres10

News

Time, sourceId, headline, urgency, provider, subjects,
audiences, bodySize, companyCount, headlineTag,
marketCommentary, sentenceCount, wordCount,
assetCode, assetName, relevance, sentimentClass,
sentimentNegative, sentimentNeutral, sentimentPositive,
sentimentWordCount, noveltyCount, volumeCounts...

Target

confidenceValue: $\hat{y}_{ti} \in [-1, 1]$

In what direction, to what extent will the stock change
Near 1.0 – large, positive return
Near -1.0 – large, negative return

Evaluation

For each day: $x_t = \sum_i \hat{y}_{ti} r_{ti} u_{ti}$

returnsOpenNextMktres10
confidenceValue

Two-sigma score:

$$\text{score} = \frac{\bar{x}_t}{\sigma(x_t)}$$

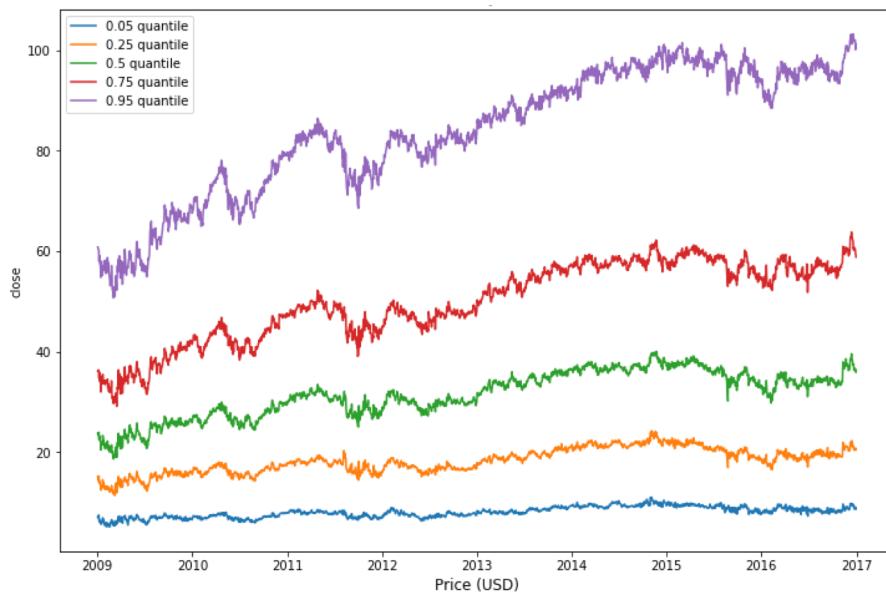
EDA – Market Data

returnsOpenNextMktres10

Market-residualized returns in the next 10 days

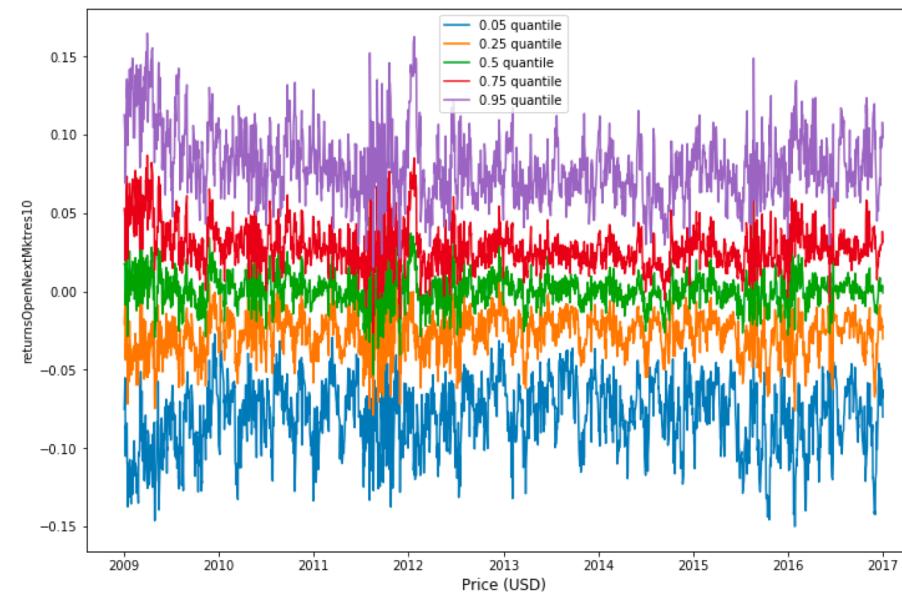
How a specific stock will change – in the next 10 days – compared to the board market

Close Price by Quantile

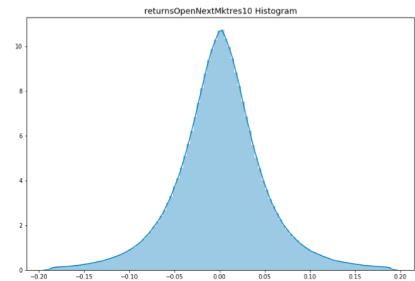


Similar trends

returnsOpenNextMktres10 by Quantile



Polarized



Normal Distribution
Mean: 0.000597
Standard deviation: 0.050352

EDA – News Data

Top source: Reuters (79%)

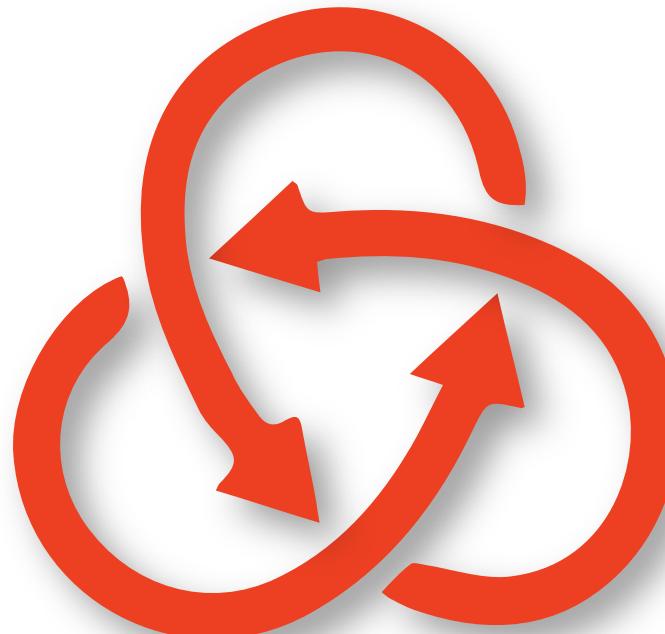
Sentiment class: 40% positive, 30% neutral, 30% negative



Challenges

News-Stock Alignment

Which news really trigger the trends and should be assigned to a related stock?



Multiple Assets

Diversified patterns. Hard to please all in terms of overall accuracy.

Model Tuning

Model structure and hyperparameter tuning

Modeling

Train: 80%

Valid: 20%

Test: 100,000 rows in Valid

LSTM Network (Long Short Term Memory Network)

Model	Data	train_acc	valid_acc	test_acc	twosigma_score
lstm_1	market (2009-2016)	0.5557	0.5478	0.5475	0.5051
lstm_2	market (2012-2016)	0.5627	0.5402	0.5378	0.4664
lstm_3	market (2014-2016)	0.558	0.5465	0.5465	0.5753
lstm_4	market & news (2014-2016)	0.5424	0.5202	0.5065	-0.1171

LightGBM

Model	Data	train_auc	valid_auc	test_acc	twosigma_score
lgb_1	market (2014-2016)		0.5408	0.5387	0.5679
lgb_2	market & news (2014-2016)		0.5276	0.4915	-0.1019

Joining irrelevant news with assets weakens model

Findings

LSTM

- Time span of training data the longer the better.
- Increasing cells in a single LSTM layer contributes little.

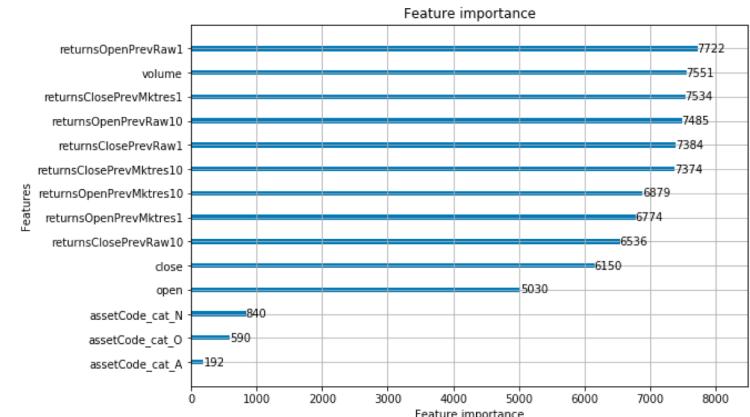
Important market features (by LGB):

- returnsOpenPrevRaw1
- volume
- returnsClosePrevMktres1
- returnsOpenPrevRaw10
- returnsClosePrevRaw1
- returnsClosePrevMktres10

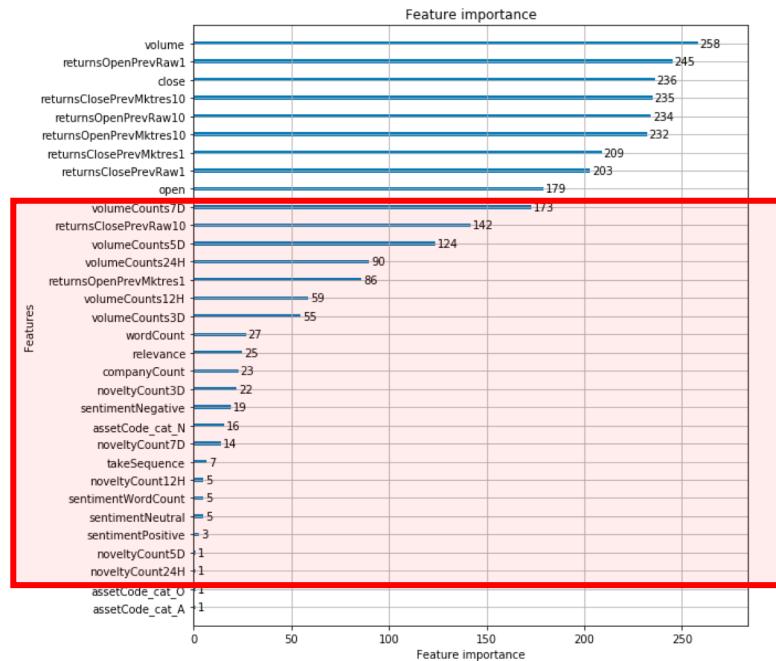
Irrelevant news hurts model performance.

- Negative news may overlap with positive trends in the same period, but they don't trigger the trends.
- A naïve merge of market and news data on time and asset may confuse the model.

lgb_1 trained on market data



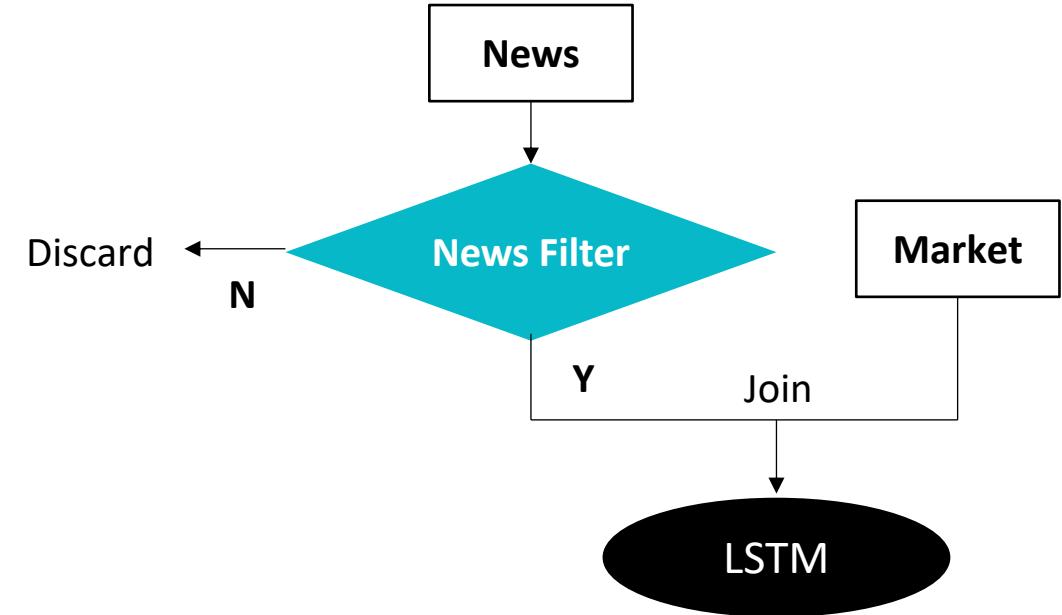
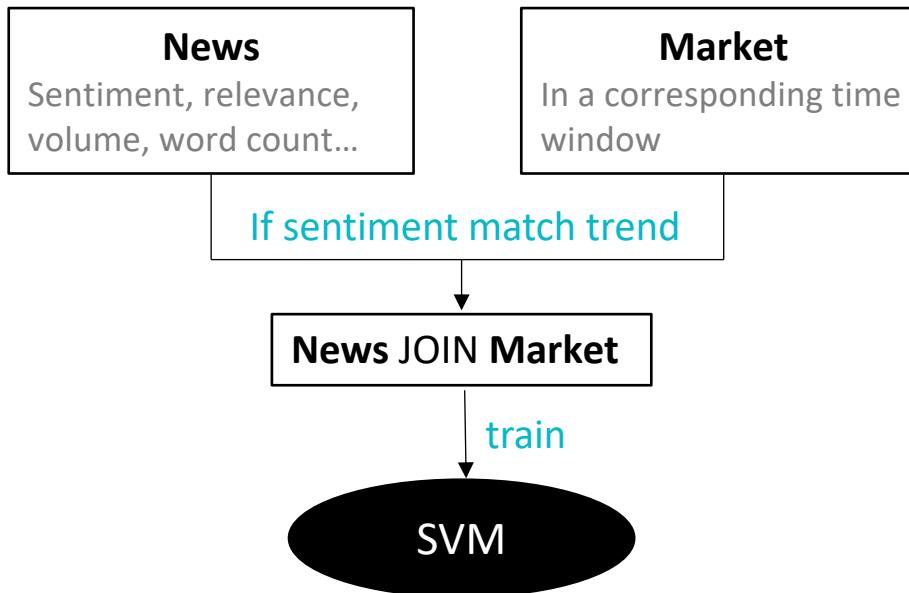
lgb_2 trained on market & news data



Tentative Approach

1

News Filter – classifier to Identify real trend-triggering news



2

LSTM – test more network structures

References

Stock Price Prediction | AI in Finance

<https://www.youtube.com/watch?v=7vunJlqLZok>

US-Stock-Market-Prediction-by-LSTM

<https://github.com/christsaizyt/US-Stock-Market-Prediction-by-LSTM>

LSTM Model on Market Data

<https://www.kaggle.com/sergeykalutsky/lstm-model-on-market-data>

Sentiment Analysis for Event-Driven Stock Prediction

<https://github.com/WayneDW/Sentiment-Analysis-in-Event-Driven-Stock-Price-Movement-Prediction>

StockPrediction

<https://github.com/jerry81333/StockProdiction/>

News Sensitive Stock Trend Prediction

https://www.researchgate.net/publication/220895244_News_Sensitive_Stock_Trend_Prediction

Thank you!

