



**CANTAB**  
PART OF GAM SYSTEMATIC

24<sup>TH</sup> MAY 2018

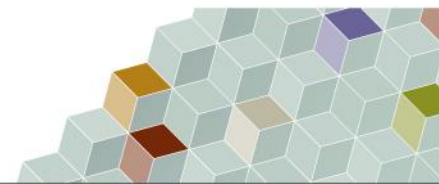
**PRIVATE AND CONFIDENTIAL. NOT TO BE DISTRIBUTED**  
AUTHORISED AND REGULATED BY THE FINANCIAL CONDUCT AUTHORITY

# TRADING ON TEXTUAL DATA USING AI

MEELIS LOOTUS, JANICE LISTER, DEREK DRIGGS

# Presentation Roadmap

---



- Background (Meelis)
  - Motivation, media, events, trading
- Demo on 1000 stocks (Janice)
  - Article analysis, event study, trading strategy
- Improve article analysis (Derek)
  - LSTM on IMDB and news headlines
- Conclusion and future work (Meelis)

# The Dream of Textual Data



## Create artificial super-analyst:



WIKIPEDIA  
The Free Encyclopedia



1 Model of the world  
(context)



2 Detect novelties  
in the world

3 Predict novelties'  
impact on stocks



Trade stocks  
for profit



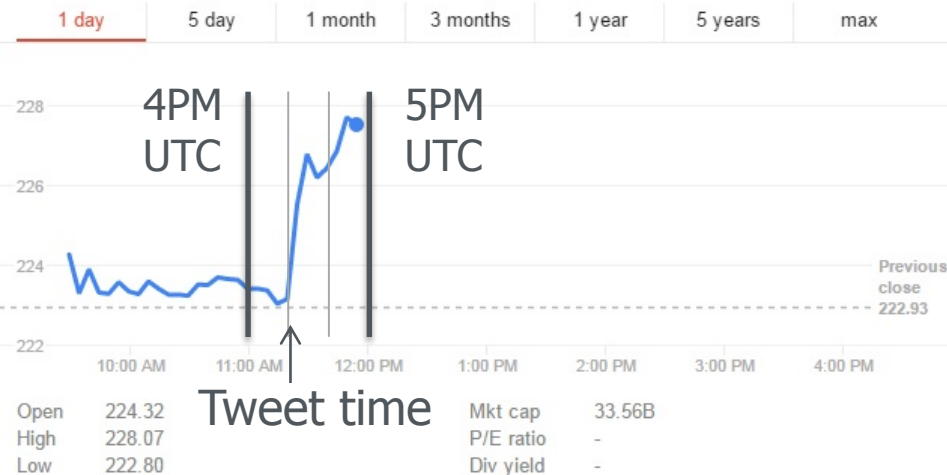
# Motivation

Does text move prices?



Tesla Motors Inc  
NASDAQ: TSLA - Aug 23, 11:59 AM EDT

227.55 USD ↑4.62 (2.07%)



# Map of Media

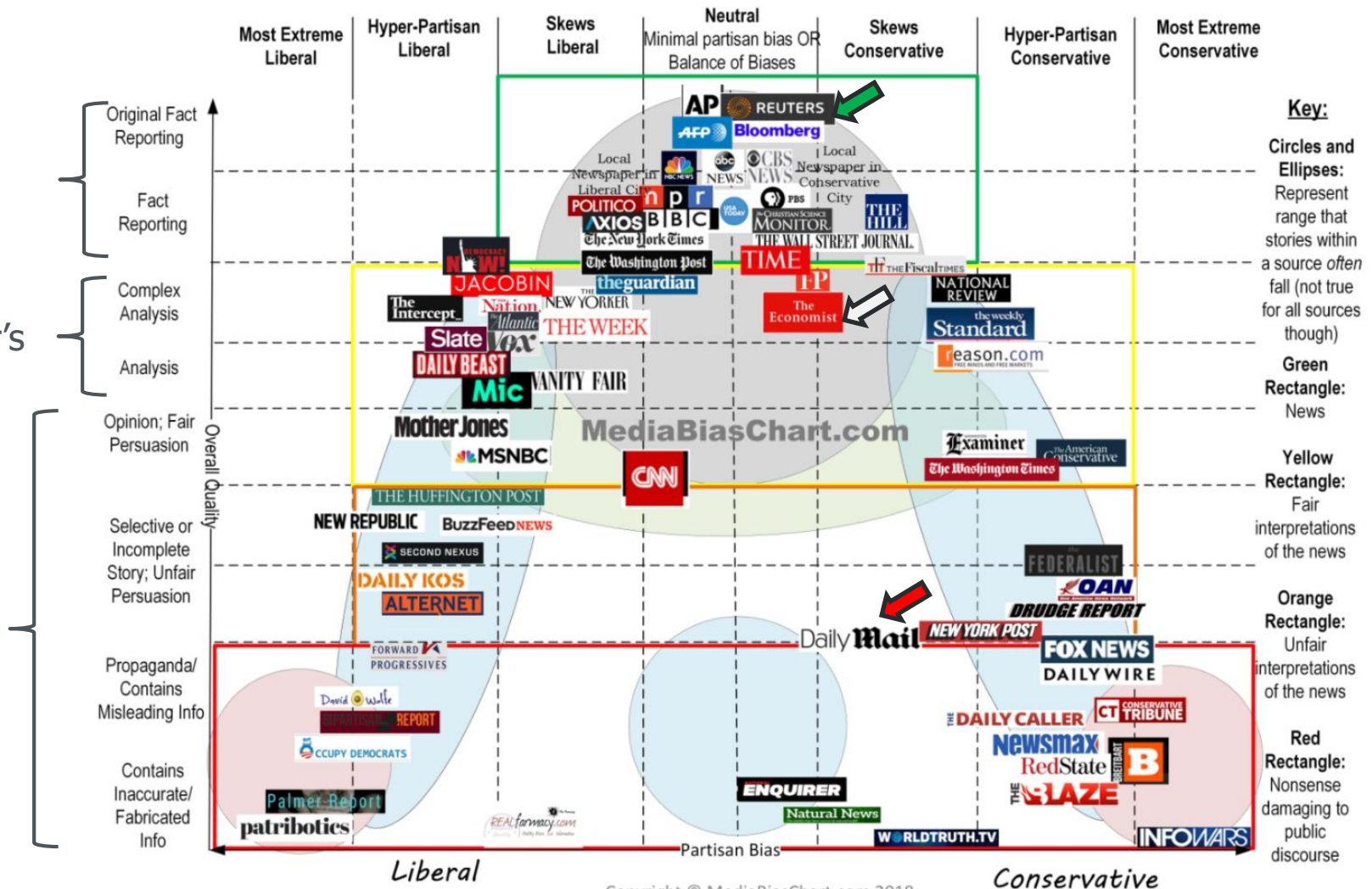


## Why difficult?

Need context to interpret novelties

Able to extract & comprehend author's argumentation

Recognise bias or irrelevance of info

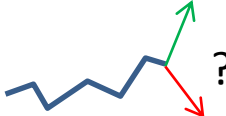

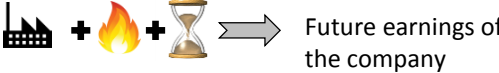

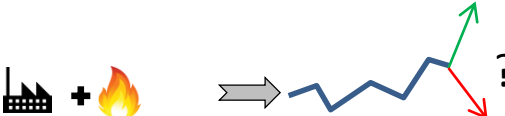
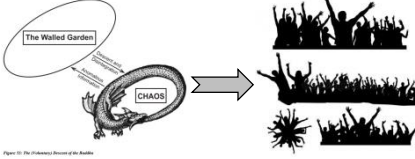
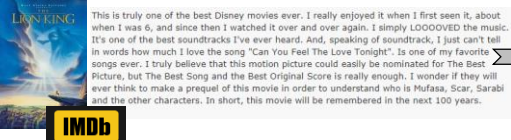



Copyright © MediaBiasChart.com 2018

For licensing and requests for royalty-free usage, please contact  
mediabiaschart@gmail.com

# Interpreting Events for Trading Stocks

Some events are much easier to interpret than others:

	Novel event	Prediction task	World to be modelled	Trading
Easy	Quarterly earnings surprise	Earnings surprise → 		HFT on specific event classes
	Difficult	Company factory fire	 Future earnings of the company	What was in the warehouse How was it going to be used What % of revenue impacted 
				Ignore events, analyse word stats
<hr/>				
Movie reviews:				
 				

We focus on (relatively) “easy to interpret” events in our experiments



# How does article analysis work in practice?



- Use commercial analyser from RavenPack
  - Detect entities and produce sentiment scores for articles
  - Processes millions of articles per year from over 10,000 sources
  - History back to 2000

## Tesla's Risky SolarCity Acquisition Gets Shareholder Approval

2016-11-17 21:26:02 UTC

**Tesla Motors ( TSLA )** shareholders overwhelmingly approved that acquisition of **SolarCity ( SCTY )** on Thursday, clearing the way **Tesla CEO** Elon Musk to create a vertically-integrated energy company.

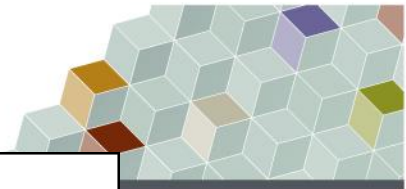
"Excluding the votes of **Elon** and other affiliated shareholders, more than 85% of shares voted were cast in favor of the acquisition," **Tesla** said in a statement. **SolarCity's shareholders** approved the acquisition Thursday morning. The transaction will be completed in the coming days, **Tesla** said.

**SolarCity shareholders** will receive 0.11 share of **Tesla**, worth \$20.75 based on Thursday's closing price, for each **SolarCity** share, making the deal worth about \$2.45 billion.

Musk initially announced the acquisition offer on July 21, which at the time valued **SolarCity** at \$2.8 billion. The offer price was renegotiated to about \$2.6 billion in August after widespread criticism of the deal and further review by **Tesla**.

# Who is it about?

## Detect entities



### Tesla's Risky SolarCity Acquisition Gets Shareholder Approval

2016-11-17 21:26:02 UTC

Tesla Motors ( TSLA ) shareholder, overwhelmingly approved that acquisition of SolarCity ( SCTY ) on Thursday, clearing the way Tesla CEO Elon Musk to create a vertically-integrated energy company.

"Excluding the votes of Elon and other affiliated shareholders, more than 85% of shares voted were cast in favor of the acquisition, Tesla said in a statement. SolarCity's shareholders approved the acquisition Thursday morning. The transaction will be completed in the coming days Tesla said.

SolarCity shareholders will receive 0.11 share of Tesla worth \$20.75 based on Thursday's closing price, for each SolarCity share, making the deal worth about \$2.45 billion.

Musk initially announced the acquisition offer on July 21, which at the time valued SolarCity at \$2.8 billion. The offer price was renegotiated to about \$2.6 billion in August after widespread criticism of the deal and further review by Tesla

Musk had called the combination of both companies "a no-brainer," believing that the acquisition of SolarCity would transform it into a vertically-integrated energy company' that would be able to sell electric cars, make and sell energy storage for buildings and the grid, and make and install solar panels, all under one roof.

Tesla stock closed Wednesday at 188.86, up 2.6% in the stock market today SolarCity stock rose 2.4% and closed at 20.27.

After the closing bell, when the news broke, with Tesla climbing to about 192 and SolarCity to 21.12.

The deal had been widely criticized by equity analysts, with Goldman Sachs saying it would create a " higher risk entity ." But early this month Tesla received the endorsement of a major shareholder advisory firm in its bid to acquire SolarCity International. Shareholder Services said shareholders of both companies should vote in favor of the controversial deal. It came two days after Tesla put forth its strongest case yet about why shareholders should approve the acquisition. In a lengthy blog post Tesla said it can achieve more than \$150 million in cost savings in the first full year after completing the acquisition. The transaction is also expected to add \$500 million to Tesla's cash balance sheet over the next three years and add \$1 billion in revenue to the combined company in 2017.

But another proxy advisory firm, Glass Lewis, blasted the deal as a " thinly veiled bailout ." Glass Lewis cited the "uncomfortable relationship between SolarCity's near-crippled financial position and Mr. Musk's personal interest - both direct and indirect - in preventing the rather pronounced, public collapse of an affiliate enterprise." SolarCity has debt of about \$3.25 billion, which has been essential for its growth. The two companies combined would have debt of more than \$6 billion. SolarCity had revenue in 2015 of \$399.6 million, up 57% year over year. Tesla showed revenue of \$4 billion in 2015, up 27%.

Musk made a Tesla SolarCity merger a big part of his recent " Secret Master Plan Part Deux " that outlined his long-term vision for the electric-car company in combination with SolarCity

The acquisition plan comes at a time when Tesla is juggling several clubs. As it works to complete the deal and integration of both companies, Tesla is working on ramping production as it works on completing a massive battery plant in Nevada, while it also gears up to produce the Model 3, which is slated to start deliveries in late 2017.

#### RELATED

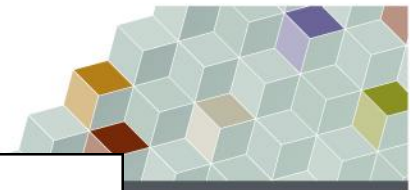
Tesla SolarCity Acquisition 'Creates A Higher-Risk Entity' Goldman

Tesla Q3 Earnings Go Full Throttle, Topping Estimates



# Who is it about?

Detect entities... & measure relevance to the article



99

## Tesla's Risky SolarCity Acquisition Gets Shareholder Approval

2016-11-17 21:26:02 UTC

99

Tesla Motors ( TSLA ) shareholders overwhelmingly approved that acquisition of SolarCity ( SCTY ) on Thursday, clearing the way Tesla CEO Elon Musk to create a vertically-integrated energy company.

"Excluding the votes of Elon and other affiliated shareholders, more than 85% of shares voted were cast in favor of the acquisition," Tesla said in a statement. SolarCity's shareholders approved the acquisition Thursday morning. The transaction will be completed in the coming days, Tesla said.

SolarCity shareholders will receive 0.11 share of Tesla, worth \$20.75 based on Thursday's closing price, for each SolarCity share, making the deal worth about \$2.45 billion.

Musk initially announced the acquisition offer on July 21, which at the time valued SolarCity at \$2.8 billion. The offer price was renegotiated to about \$2.6 billion in August after widespread criticism of the deal and further review by Tesla.

Musk had called the combination of both companies "a no-brainer," believing that the acquisition of SolarCity would transform it into a vertically-integrated energy company that would be able to sell electric cars, make and sell energy storage for buildings and the grid, and make and install solar panels, all under one roof.

Tesla stock closed Wednesday at 188.66, up 2.6% in the stock market today SolarCity stock rose 2.4% and closed at 20.27.

6

After the closing bell, when the news broke, with Tesla climbing to about 192 and SolarCity to 21.12.

The deal had been widely criticized by equity analysts with Goldman Sachs saying it would create a "higher risk entity ." But early this month Tesla received the endorsement of a major shareholder advisory firm in its bid to acquire SolarCity International Shareholder Services said shareholders of both companies should vote in favor of the controversial deal. It came two days after Tesla put forth its strongest case yet about why shareholders should approve the acquisition. In a lengthy blog post , Tesla said it can achieve more than \$150 million in cost savings in the first full year after completing the acquisition. The transaction is also expected to add \$500 million to Tesla's cash balance sheet over the next three years and add \$1 billion in revenue to the combined company in 2017.

But another proxy advisory firm, Glass Lewis, blasted the deal as a "thinly veiled bailout ." Glass Lewis cited the "uncomfortable relationship between SolarCity's near-crippled financial position and Mr. Musk's personal interest - both direct and indirect - in preventing the rather pronounced, public collapse of an affiliate enterprise." SolarCity has debt of about \$3.25 billion, which has been essential for its growth. The two companies combined would have debt of more than \$6 billion. SolarCity had revenue in 2015 of \$399.6 million, up 57% year over year. Tesla showed revenue of \$4 billion in 2015, up 27%.

Musk made a Tesla-SolarCity merger a big part of his recent "Secret Master Plan Part Deux " that outlined his long-term vision for the electric-car company in combination with SolarCity

The acquisition plan comes at a time when Tesla is juggling several clubs. As it works to complete the deal and integration of both companies Tesla is working on ramping production as it works on completing a massive battery plant in Nevada, while it also gears up to produce the Model 3, which is slated to start deliveries in late 2017.

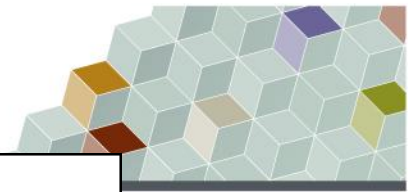
### RELATED

Tesla SolarCity Acquisition 'Creates A Higher-Risk Entity': Goldman

Tesla Q3 Earnings Go Full Throttle, Topping Estimates

# What is it about?

*Event detection, event relevance and sentiment*



## Tesla's Risky SolarCity Acquisition Gets Shareholder Approval

2016-11-17 21:26:02 UTC

**Tesla Motors ( TSLA ) shareholders** overwhelmingly approved that acquisition of **SolarCity ( SCTY )** on Thursday, clearing the way **Tesla CEO Elon Musk** to create a vertically-integrated energy company.

"Excluding the votes of Elon and other affiliated shareholders, more than 85% of shares voted were cast in favor of the acquisition," **Tesla** said in a statement. **SolarCity's shareholders approved the acquisition** Thursday morning. The transaction will be completed in the coming days, **Tesla** said.

**SolarCity shareholders** will receive 0.11 share of **Tesla**, worth \$20.75 based on Thursday's closing price, for each **SolarCity** share, making the deal worth about \$2.45 billion.

Musk initially announced the acquisition offer on July 21, which at the time valued **SolarCity** at \$2.8 billion. The offer price was renegotiated to about \$2.6 billion in August after widespread criticism of the deal and further review by **Tesla**.

Musk had called the combination of both companies "a no-brainer," believing that the acquisition of **SolarCity** would transform it into a vertically-integrated energy company' that would be able to sell electric cars, make and sell energy storage for buildings and the grid, and make and install solar panels, all under one roof.

**Tesla** stock closed Wednesday at 188.86, up 2.6% in the stock market today . **SolarCity stock rose 2.4%** and closed at 20.27.

After the closing bell, when the news broke, with **Tesla** climbing to about 192 and **SolarCity** to 21.12.

**The deal had been widely criticized by equity analysts , with Goldman Sachs saying it would create a " higher risk entity .** But early this month **Tesla** received the endorsement of a major shareholder advisory firm in its bid to acquire **SolarCity**. International **Shareholder Services** said shareholders of both companies should vote in favor of the controversial deal. It came two days after **Tesla** put forth its strongest case yet about why shareholders should approve the acquisition. In a lengthy blog post , **Tesla** said it can achieve more than \$150 million in cost savings in the first full year after completing the acquisition. The transaction is also expected to add \$500 million to **Tesla's** cash balance sheet over the next three years and add \$1 billion in revenue to the combined company in 2017.

But another proxy advisory firm, Glass Lewis, blasted the deal as a " thinly veiled bailout . " Glass Lewis cited the "uncomfortable relationship between **SolarCity's** near-crippled financial position and Mr. Musk's personal interest - both direct and indirect - in preventing the rather pronounced, public collapse of an affiliate enterprise." **SolarCity** has debt of about \$3.25 billion, which has been essential for its growth. The two companies combined would have debt of more than \$6 billion. **SolarCity** had revenue in 2015 of \$399.6 million, up 57% year over year. **Tesla showed revenue of \$4 billion in 2015, up 27%.**

Musk made a **Tesla-SolarCity merger** a big part of his recent " Secret Master Plan Part Deux " that outlined his long-term vision for the electric-car company in combination with **SolarCity**.

The acquisition plan comes at a time when **Tesla** is juggling several clubs. As it works to complete the deal and integration of both companies **Tesla** is working on ramping production as it works on completing a massive battery plant in Nevada, while it also gears up to produce the Model 3, which is slated to start deliveries in late 2017.

### RELATED

**Tesla SolarCity Acquisition 'Creates A Higher-Risk Entity': Goldman**

**Tesla Q3 Earnings** Go Full Throttle, Topping Estimates

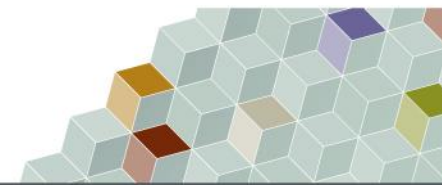
Group: acquisition  
Relevance: 80  
Sentiment: 0.58

Group: stock-price  
Relevance: 50  
Sentiment: 0.47

Group: revenue  
Relevance: 30  
Sentiment: 0

Group: merger  
Relevance: 25  
Sentiment: 0.63

Group: earnings  
Relevance: 0  
Sentiment: 0



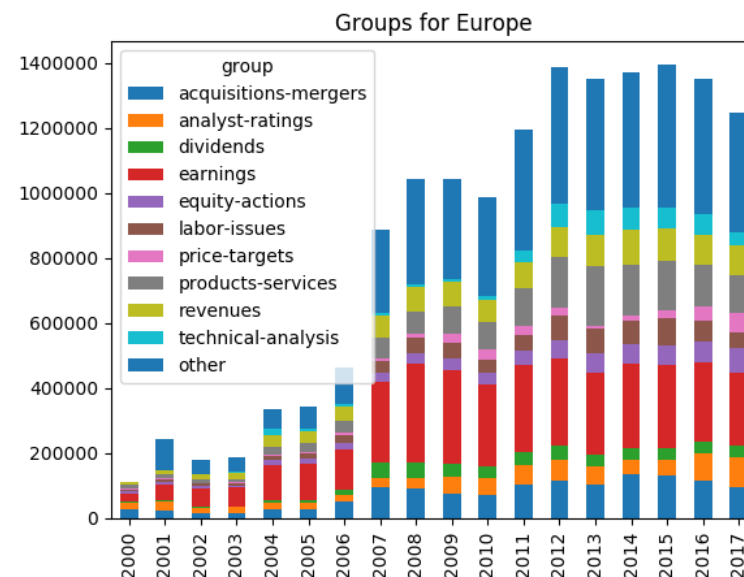
## Approach:

- Detect entities and relevance
- Detect events
- Score sentiment of events and article

## Events:

- Templates defined by financial experts
- Location determines the relevance of the event in a story. Higher up in the article has a higher relevance score.
- Novelty: How long since we detected an event of the same category.
- Tend to be easy to interpret types

*Sentiment scores for highly relevant and novel events may be better predictors.*

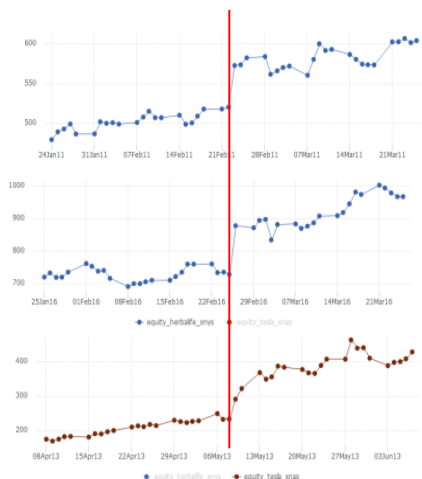


# News event study

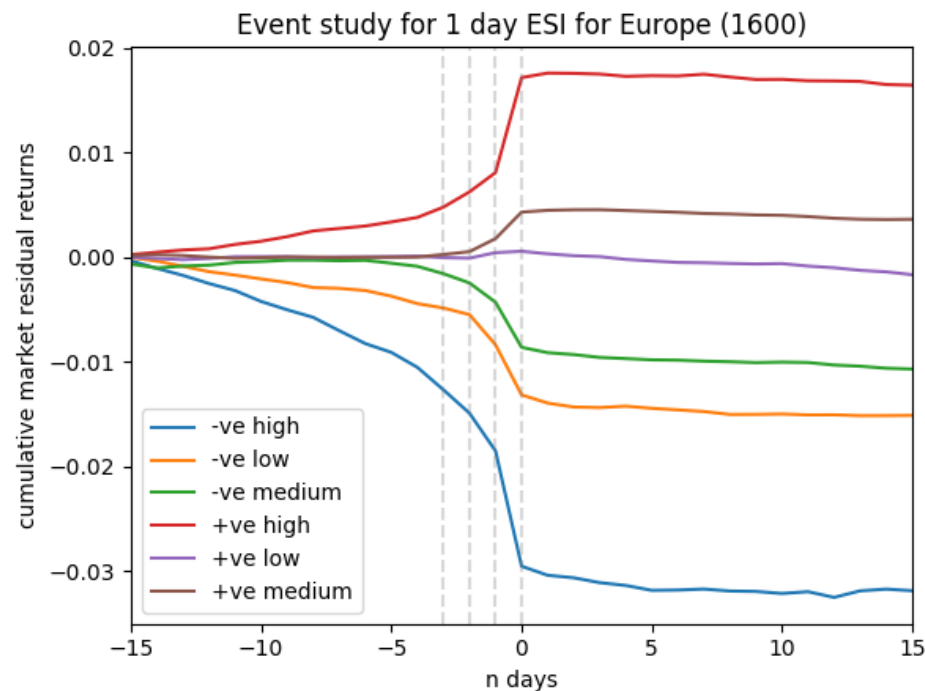
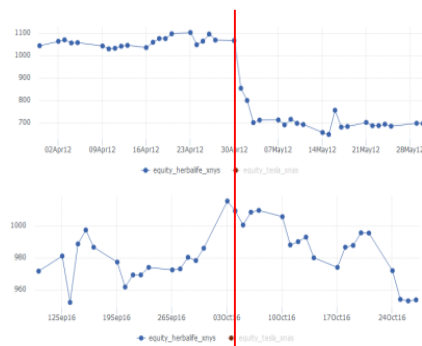
Lets look at stock returns around events:

- ~1000 European stocks
- Filter data by
  - Relevance > 90 (headlines)
  - Novelty > 90 days
- Build daily signal
- Group data by *Event Sentiment*

## Positive events



## Negative events

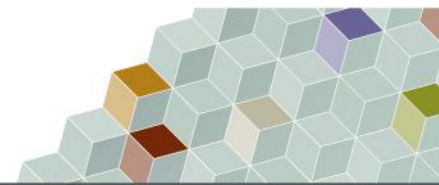


Returns of positive and negative sentiment stocks tend to be in the right direction

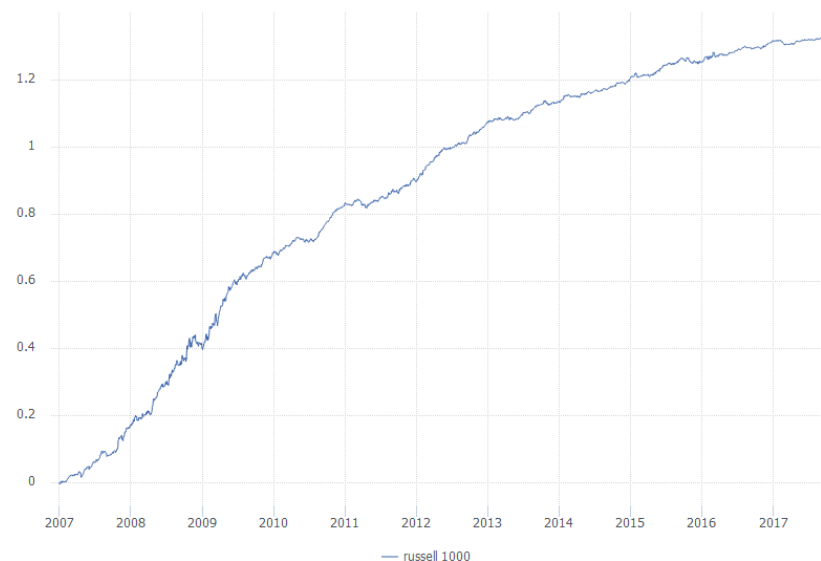
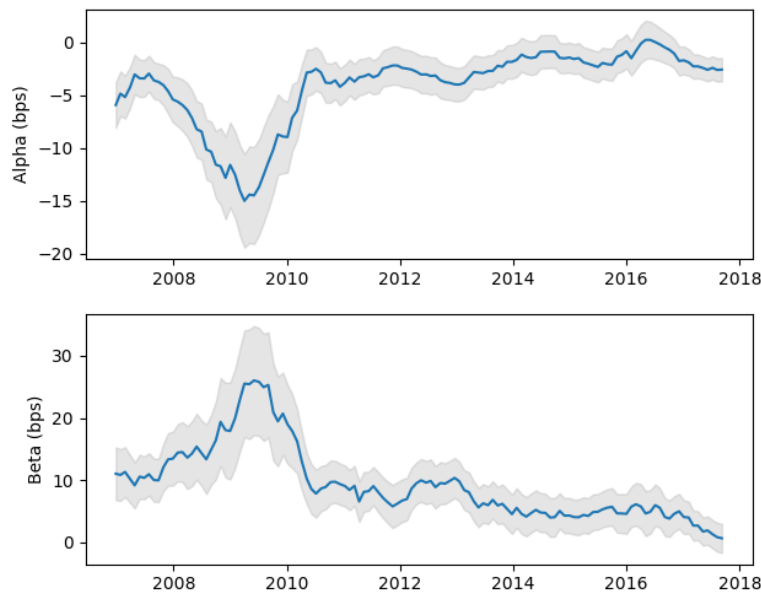
Some of the move is before the event

- Information leakage
- Data lag
- Some of the news is about price moves!

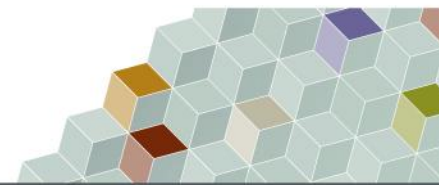
# From event study to a trading strategy



- Consider “residual returns” (stripping out other signals)  $R_{t+1} = B(t)*F(t) + u_{t+1}$
- Regress next day returns against Event Sentiment  $u_{i,t+1} = \alpha + \log(ESI_{i,t+1}) \beta + \varepsilon_{i,t+1}$



- Positive beta and gross performance, but some recent decay and high turnover (trade costs)
- Can we do better?**



Sainsbury Acquisition of ASDA Not to be Completed

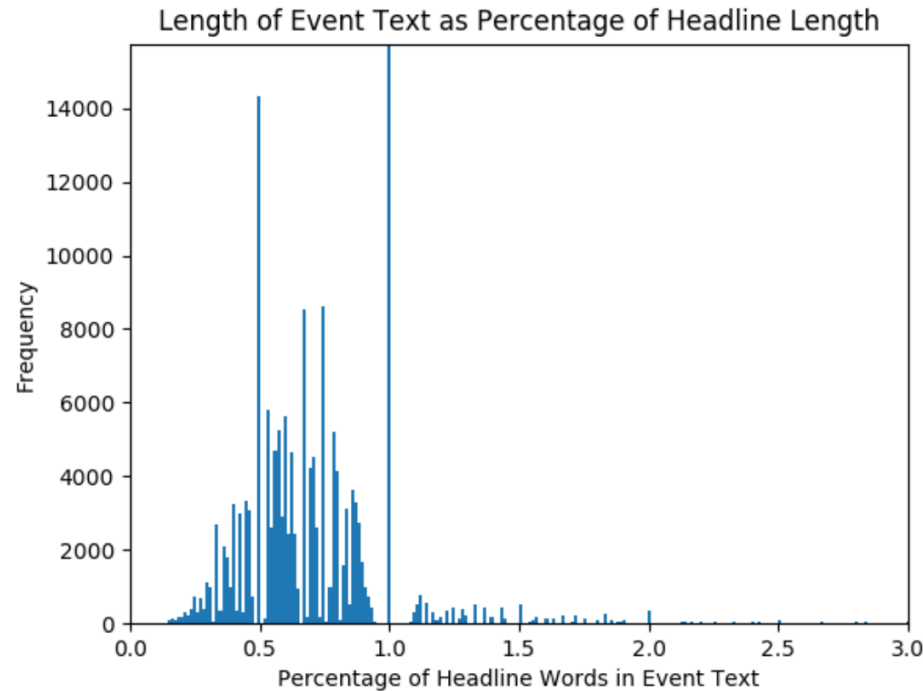
RavenPack's sentiment analyser is a **high precision, low recall** system: it performs well on a set of short text extracts but cannot be extended to longer sentences.

RavenPack extracts "Sainsbury Acquisition of ASDA" in the above headline, and incorrectly labels it "positive".

**RavenPack is prone to mislabeling "complicated" headlines**, especially long headlines or those containing negating clauses.



## How often can we trust the sentiment scores?



**Figure:** Event-text length as a percentage of the headline length

RavenPack often does not consider entire headlines when computing the ESS.

Event-texts containing negating clauses represent about 1.5% of total events.

# Bag-of-Words Embedding



	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	1	1	1							
Doc 2	1		1	1	1	1				
Doc 3					1	1	1	2	1	1

**Figure:** Bag-of-words embedding

Most sentiment classifiers use a bag-of-words embedding with a regression-based classifier, like logistic regression or naive Bayes.

# Linguistic Structure Analysis

## Part of speech:

NP NP RB VBD IN NN NP NP , CC PRP VBZ RB VBG PRP IN PRP .  
Mrs. Clinton previously worked for Mr. Obama, but she is now distancing herself from him .

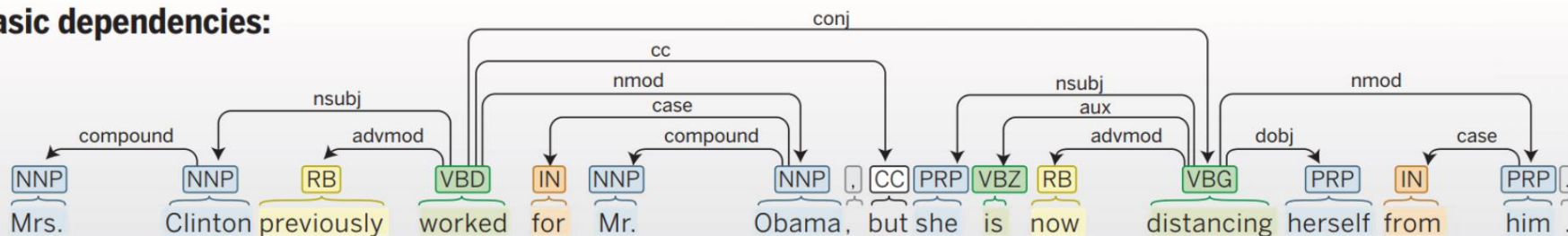
## Named entity recognition:

Person Date Person Date  
Mrs. Clinton previously worked for Mr. Obama, but she is now distancing herself from him.

## Co-reference:

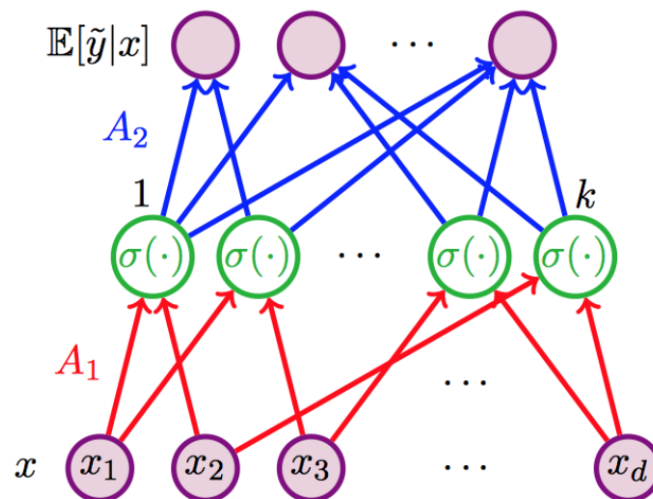
Coref Coref Coref  
Mention Ment M Mention M  
Mrs. Clinton previously worked for Mr. Obama, but she is now distancing herself from him.

## Basic dependencies:



**Fig. 1. Many language technology tools start by doing linguistic structure analysis.** Here we show output from Stanford CoreNLP. As shown from top to bottom, this tool determines the parts of speech of each word, tags various words or phrases as semantic named entities of various sorts, determines which entity mentions co-refer to the same person or organization, and then works out the syntactic structure of each sentence, using a dependency grammar analysis.

# A Primer on Neural Networks



**Figure:** The architecture of a simple (feed-forward) neural network (Anandkumar, 2014).

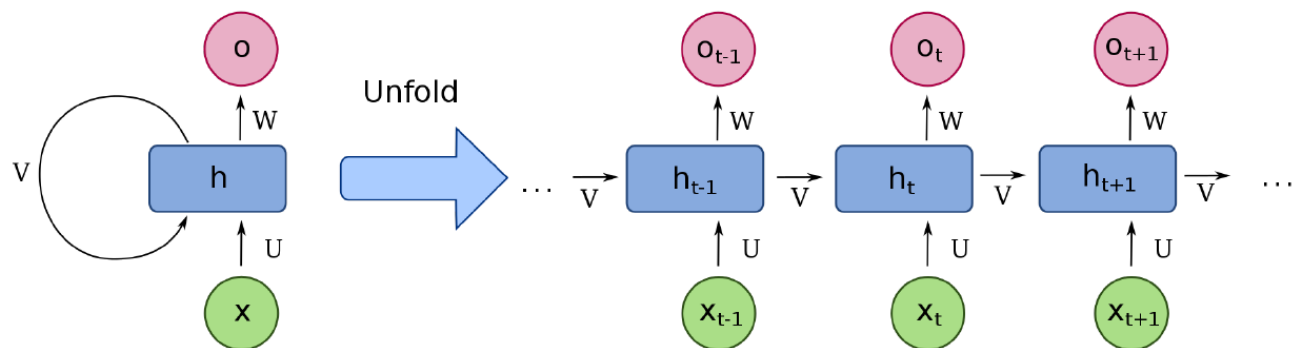
Neural networks attempt to “learn” the function  $\mathbb{E}[y|\mathbf{x}]$  for labels  $y \in \{0,1\}$  and feature-vector  $\mathbf{x} \in \mathbb{R}^n$ .

The function it learns has the form

$$A_n \sigma(\dots A_2 \sigma(A_1 \mathbf{x} + b_1) + b_2 \dots) + b_n$$

where  $\sigma$  is a nonlinear function (like tanh) applied element-wise.

# A Primer on RNN's



**Figure:** A basic RNN (Deloche, 2015).

RNN's incorporate self-recurrent connections to “remember” long-term dependences within data. This makes them particularly well-suited for sequence generation and prediction tasks (Tai et al., 2015).

Training RNN's is unstable because the gradients related to self-recurrent connections can grow/shrink exponentially.

**Long-short-term-memory (LSTM) networks ameliorate this problem by “forgetting” intelligently.**



We tested our LSTM on two datasets:

IMDB movie reviews (a standard dataset for testing sentiment analyzers)

“The Emoji Movie is a very powerful, moving film...”

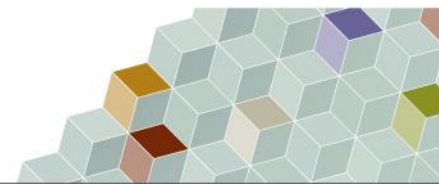
RavenPack extracted headlines with sentiment score

“Sainsbury Acquisition of ASDA Represents Opportunity but also Risk.”

We compare our model to a **logistic-regression baseline**.

Performance measured using **ROC curves, precision-recall curves, and accuracy (percentage of labels predicted correctly)**.





### IMDB Dataset:

12,500 positive, 12,500 negative reviews in total

After data processing, only 2,120 positive and negative reviews remained

(70/20/10) training/testing/validation split

Equal parts positive and negative reviews in each split



### RavenPack ESS:

20,000 positive, 20,000 negative news snippets

Considered only news snippets from headlines

(70/20/10) training/testing/validation split

Equal parts positive and negative sentiment in each split



Precision-Recall curve:

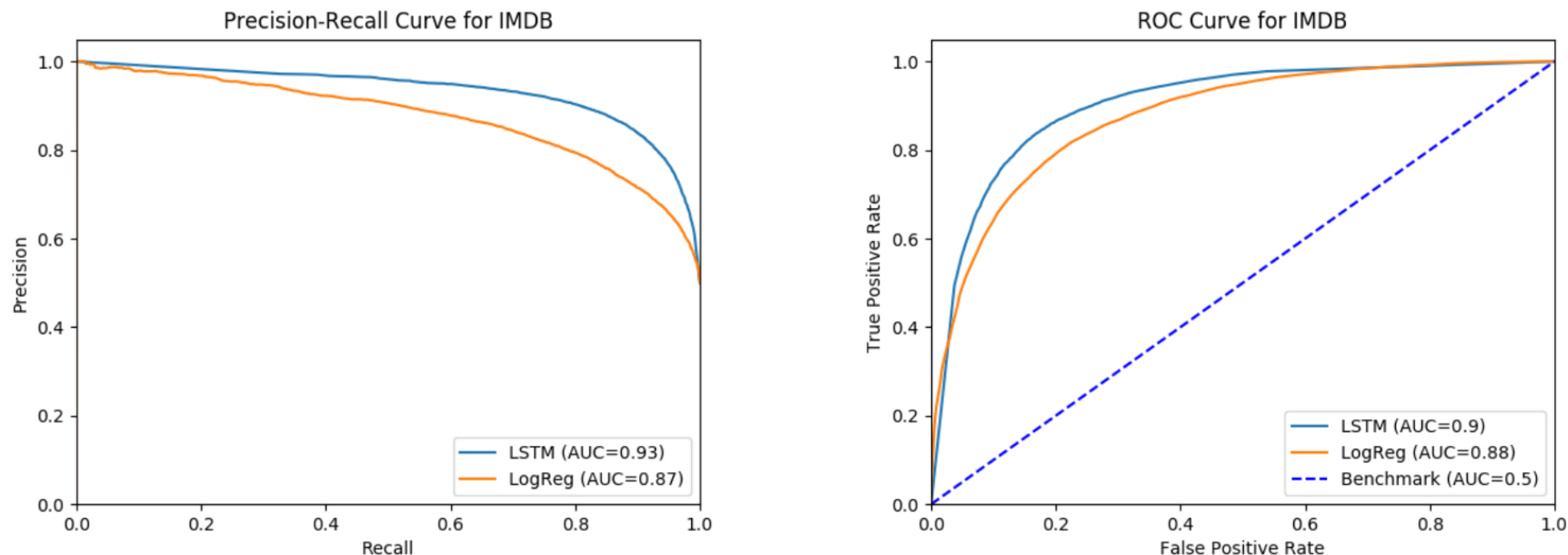
$$\text{Precision: } \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall: } \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

ROC curve: true positive rate by false positive rate

**For both plots, larger area under the curve implies better model performance.**

## Results on IMDB Dataset

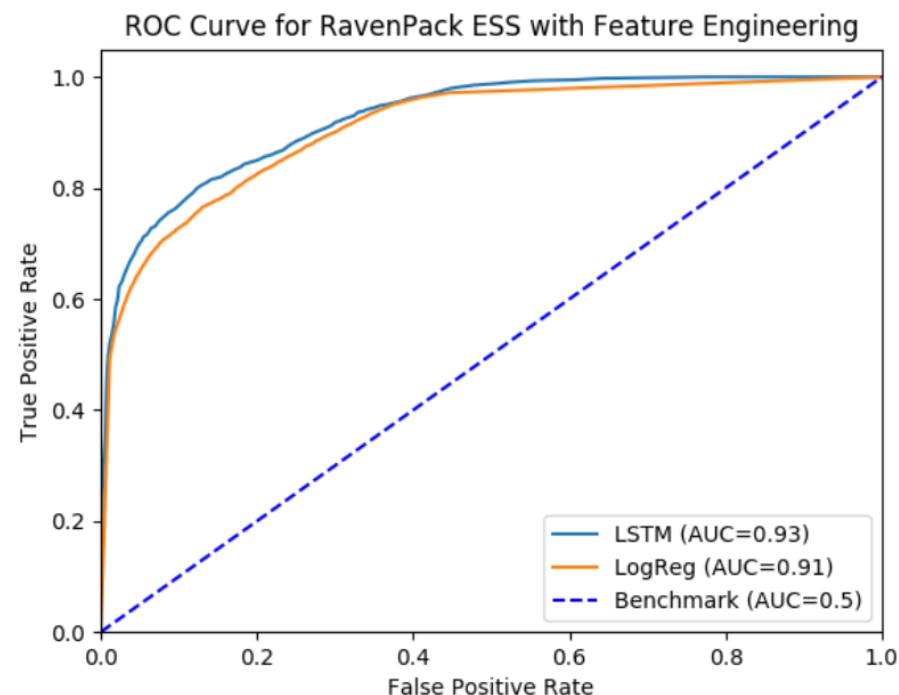
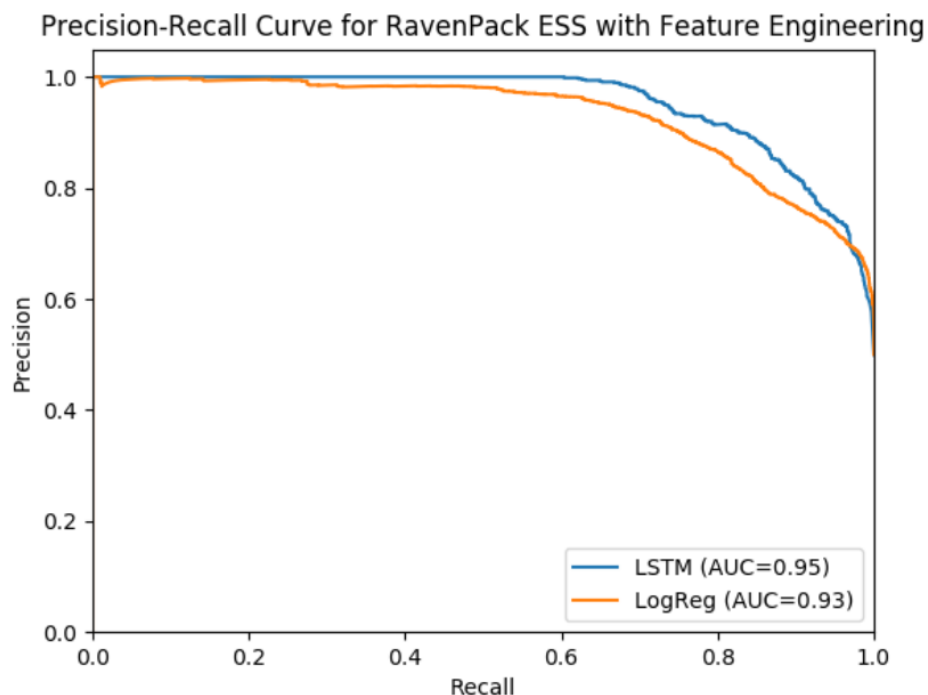
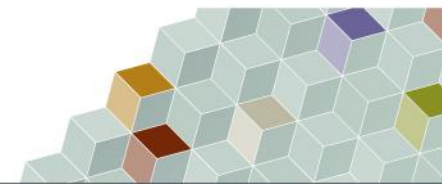


**Figure:** Precision-recall and ROC curves for our LSTM trained and tested on the IMDB dataset

Our model achieves an accuracy of 86%, which is comparable to the state-of-the-art, 89% (Maas et al., 2011).

The LSTM significantly outperforms the logistic-regression baseline, suggesting that the sentiment of movie reviews is dependent on semantic structure.

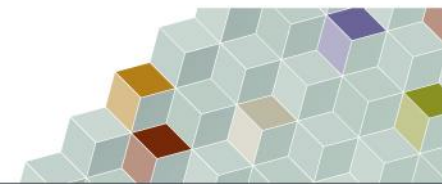
## Results on News Headlines



**Figure:** Models trained and tested on RavenPack ESS dataset

The LSTM offers a slight improvement over logistic regression, suggesting semantic structure affects sentiment in news extracts.

## How our LSTM “Reads”



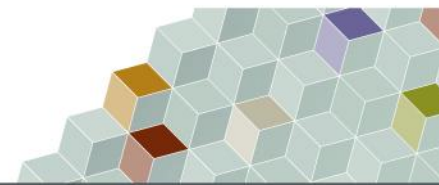
sainsbury acquisition of asda represents opportunity but also risk ✓  
sainsbury beats expectations but profits down 18% ✗  
electrolux sees rise in q2 profits but drops in sales ✓  
electrolux acquisition of ge appliances not to be completed ✗  
swatch sees sales rise but misses targets ✗  
polymetal executive says possible merger with polyus gold international not feasible ✗  
hargreaves lansdown cuts the cost of investing but it's not nearly enough ✓  
diageo: impressive dividend growth history, but not a strong buy ✓  
michelin profit softens, but tyre maker firm on 2017 goals ✓  
prysmian -3.4% cut to dividend not helping ✗  
cuba's communist economy in comparison to other latin american economies: corruption and failure ✓  
how general electric gambled on fossil fuels and lost ✓

sainsbury acquisition of asda represents opportunity but also risk ✗  
sainsbury beats expectations but profits down 18% ✗  
electrolux sees rise in q2 profits but drops in sales ✗  
electrolux acquisition of ge appliances not to be completed ✗  
swatch sees sales rise but misses targets ✗  
polymetal executive says possible merger with polyus gold international not feasible ✗  
hargreaves lansdown cuts the cost of investing but it's not nearly enough ✗  
diageo: impressive dividend growth history, but not a strong buy ✓  
michelin profit softens, but tyre maker firm on 2017 goals ✗  
prysmian -3.4% cut to dividend not helping ✗  
cuba's communist economy in comparison to other latin american economies: corruption and failure ✗  
how general electric gambled on fossil fuels and lost ✗

**Figure:** How the sentiment changes as each model “reads” a conjunctive or negating clause. (Left): LSTM. (Right): Logistic regression. Blue is positive, red is negative.

The LSTM performs better than the baseline on “complex” sentences, and can “change its mind” much more fluidly than the logistic regression model.





- Textual data moves prices
- Challenging problem
  - Need to focus the task (which sources, events, strategies) intelligently
- Evaluation of a stand-alone news signal
  - Demonstrated strategy a using commercial news analyser
  - Reasonable gross returns profile
- Improvements to sentiment analysis
  - Promising results from LSTM
  - Potential future intern projects,
    - deep learning on longer documents
    - Improved interpretation of harder-to-interpret event categories (e.g. products-services)