

An aerial night view of New York City, showing a dense grid of skyscrapers and city lights. A network of thin, glowing white lines is overlaid on the image, connecting various points across the cityscape, suggesting a complex system or data network. The text is positioned in the upper left quadrant, with the main title in white and orange, and the author's name in white script below it.

Predicting the Stock Market with the Daily News

Harry Smith

An aerial night view of a city skyline, likely New York City, with numerous skyscrapers illuminated. A network of thin, light-blue lines is overlaid on the image, connecting various points across the city. The word "Goal" is written in a large, orange, serif font in the upper left quadrant.

Goal

Use topic modelling and sentiment analysis
to make inferences about the stock market
from daily news headlines

Data

8th June 2008 – 1st July 2016, 1989 rows



/r/worldnews, 25 most popular headlines on each day



Dow Jones Industrial Average (DJIA), open & close

Topic modelling and sentiment analysis

Predictive modelling

Topics

Modelled with non-negative matrix factorization (NMF)

- ❖ Human rights
- ❖ Wikileaks/Julian Assange
- ❖ Asian countries
- ❖ Israel/Palestine
- ❖ Islam/Terrorism
- ❖ Climate change
- ❖ North Korea/South Korea
- ❖ Oil/Saudi Arabia
- ❖ World leaders
- ❖ Russia/Ukraine
- ❖ Nuclear weapons
- ❖ War/Afghanistan/Iraq
- ❖ NSA/Edward Snowden
- ❖ Egypt/Protests
- ❖ Child abuse

Sentiment analysis

Subjectivity

Subjective \longleftrightarrow Objective

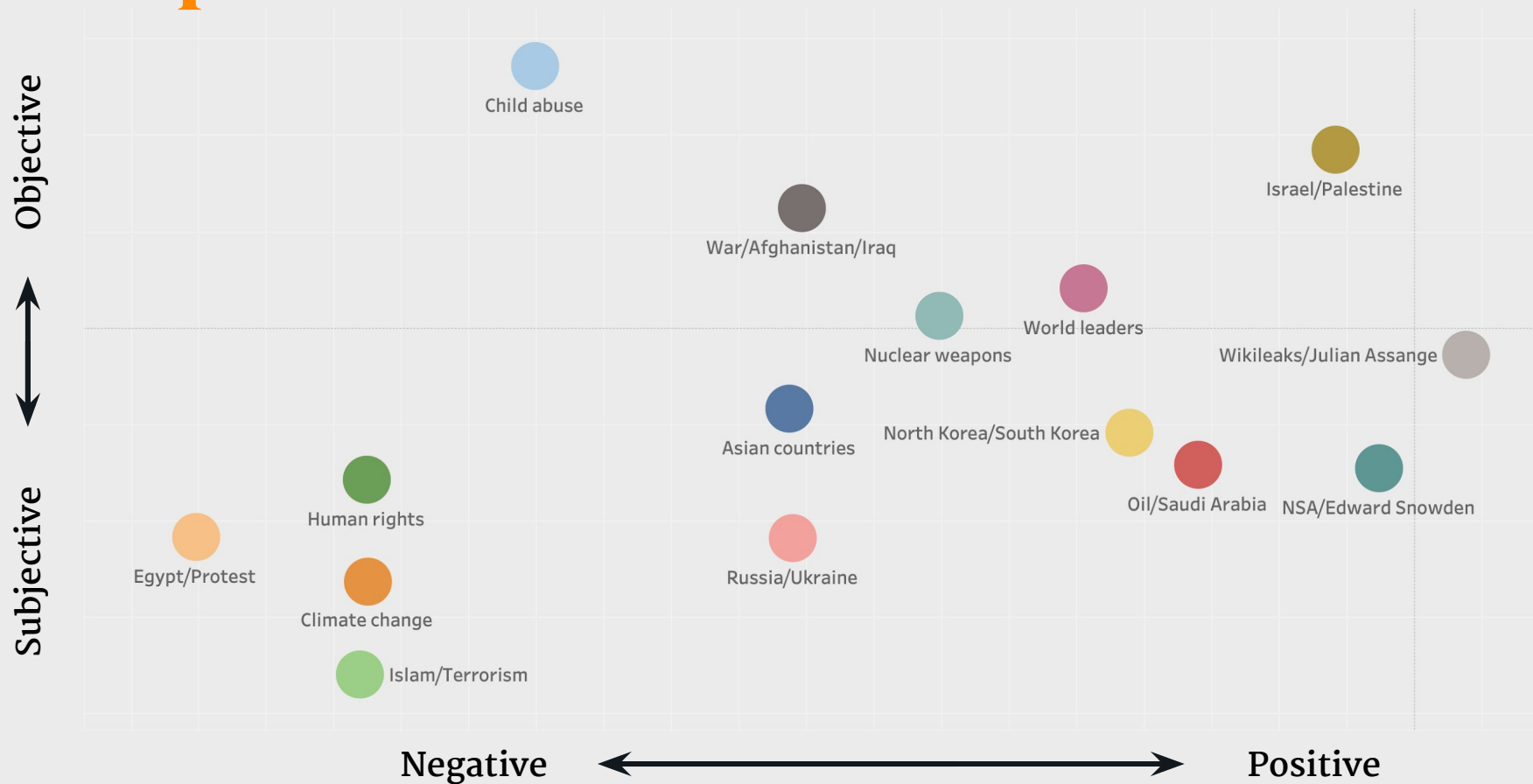
- ❖ Looks for a number of subjective “clues”
- ❖ Measures objectivity as the absence of subjectivity

Polarity

Negative \longleftrightarrow Positive

- ❖ Rule-based polarity classifier
- ❖ Works by counting positive, negative and neutral words

Topics vs Sentiment



Anomalies

Detected with an Isolation Forest algorithm

❖ November 2010:

Cablegate

Wikileaks leaks US diplomatic cables

❖ February, March 2011:

Arab Spring

Uprisings and protests in Bahrain, Egypt, Libya, Syria, Tunisia and Yemen

❖ June, July 2013:

Fugitivity of Edward Snowden

Modelling

Labels = $\begin{cases} 0: \text{Tomorrow's DJIA decreases} \\ 1: \text{Tomorrow's DJIA increases} \end{cases}$



An aerial photograph of the New York City skyline at sunset. The sky is filled with dark, dramatic clouds, and the sun is low on the horizon, casting a warm, golden glow over the city. The Empire State Building is prominent in the center. To the left, the MetLife building is visible. The word "Questions?" is overlaid in a large, orange, serif font. In the top right corner, there are faint, thin white lines that appear to be part of a network or data visualization. The overall mood is contemplative and urban.

Questions?