Machine Learning:
When Sentiment is News

Nadia Massoud

Dec 2019

Sydney Banking and Financial Stability Conference

Keynote Address

Artificial Intelligence (AI)

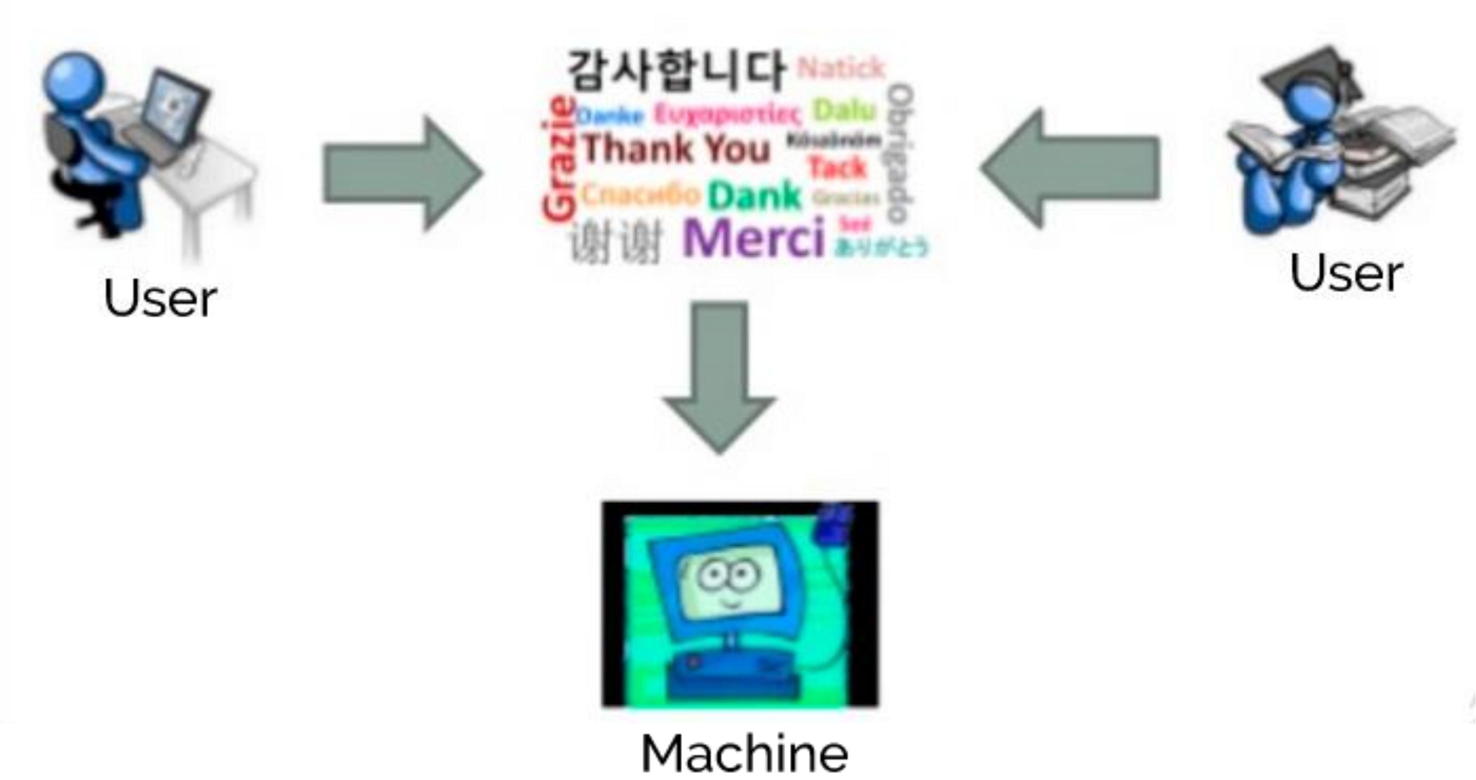
The study of how to make computers do things, which at the moment, people do better.

Goals of Artificial Intelligence

- Reasoning
- Automated Learning & Scheduling
- Machine Learning
- Natural Language Processing
- Computer Vision
- Robotics
- General Intelligence

Natural Language Processing (NLP)

- Natural Language Processing (NLP) is "**the ability of machines to understand and interpret human language as is written or spoken.**"



7 APPLICATIONS OF DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING

- **Text Classification**
  - **A popular classification example is sentiment analysis where class labels represent the emotional tone of the source text such as "positive" or "negative".**
- Language Modelling
- Speech Recognition
- Caption Generation
- Machine Translation
- Document Summarization
- Question Answering

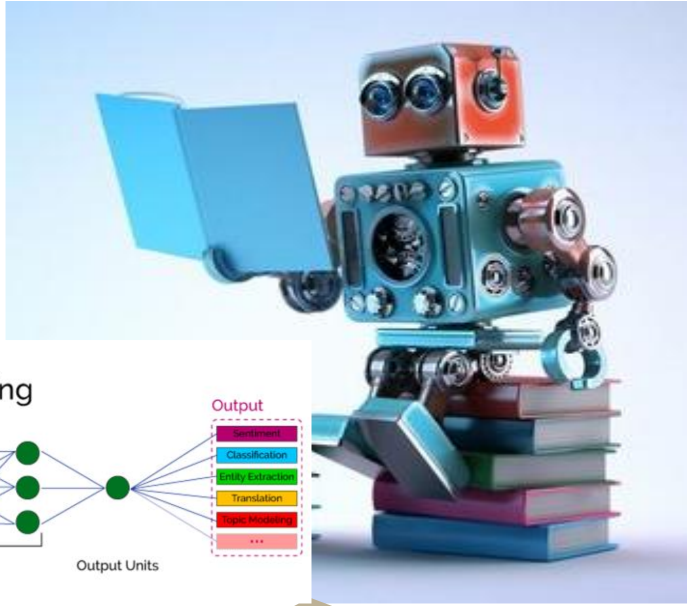
Sentiment Analysis (SA)

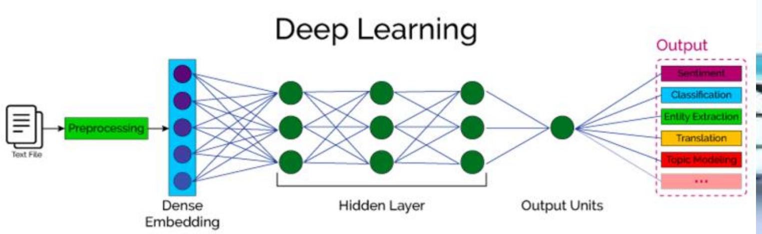
- ✓ “WiseTech shares plummet again after its short-seller launched another attack” → Negative
- ✓ “How Apple is Gearing Up for an Exciting 2020” → Positive
- ✓ “McDonald's And The Difficulty Of Repeating Its Past” → Negative

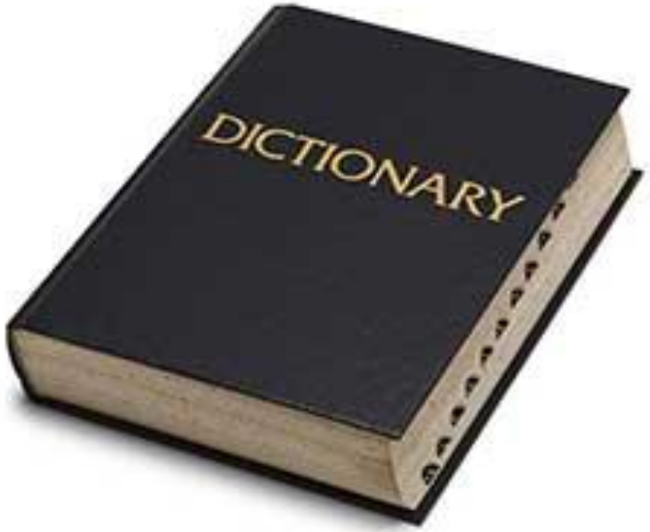
Why does Sentiment Analysis matter?

- ✓ It allows us to gain an overview of the public opinion behind certain topics.
- ✓ It is a tool to understand how people feel about a brand / person / system.
- ✓ It potentially affects stock prices.
- ✓ …

Sentiment Analysis Methods







Neural Networks

Classical Machine Learning

Domain-General “Bag-of-Words”

Domain-Specific

Bag of Words

Bag of Words (BoW)

In BoW approaches, text is represented as a **dictionary** in a multi-set (bag) of words, disregarding grammar and word order but recognizing multiplicity of meanings.

Bag of Words

Dictionary	
Word	**Sense**
Hate	-1
Best	+1
Risk	-1
Buy	+1

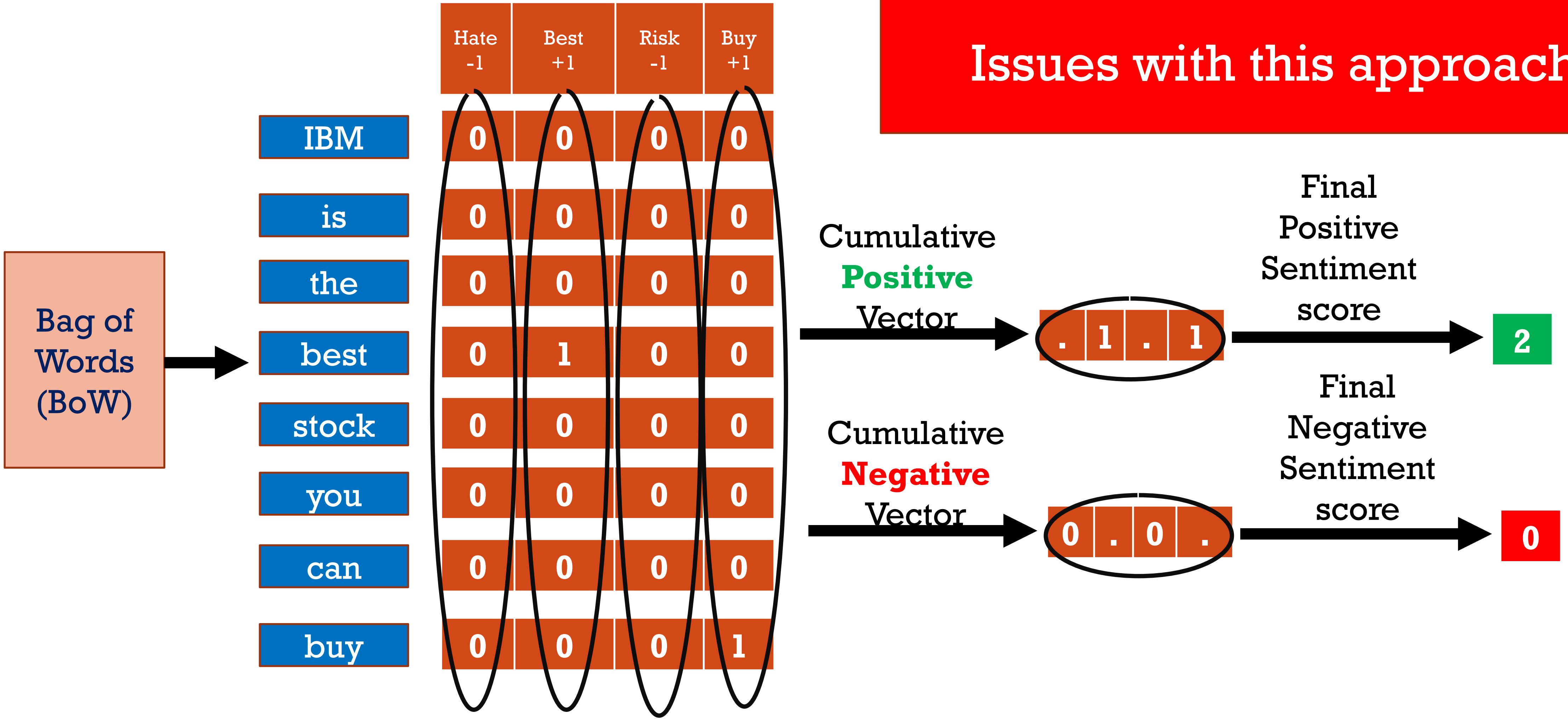
Bag of Words (BoW)

Text
IBM is the best stock you can buy

	Hate	Best	Risk	Buy
IBM	0	0	0	0
is	0	0	0	0
the	0	0	0	0
best	0	1	0	0
stock	0	0	0	0
you	0	0	0	0
can	0	0	0	0
buy	0	0	0	1

Bag of Words

Issues with this approach?



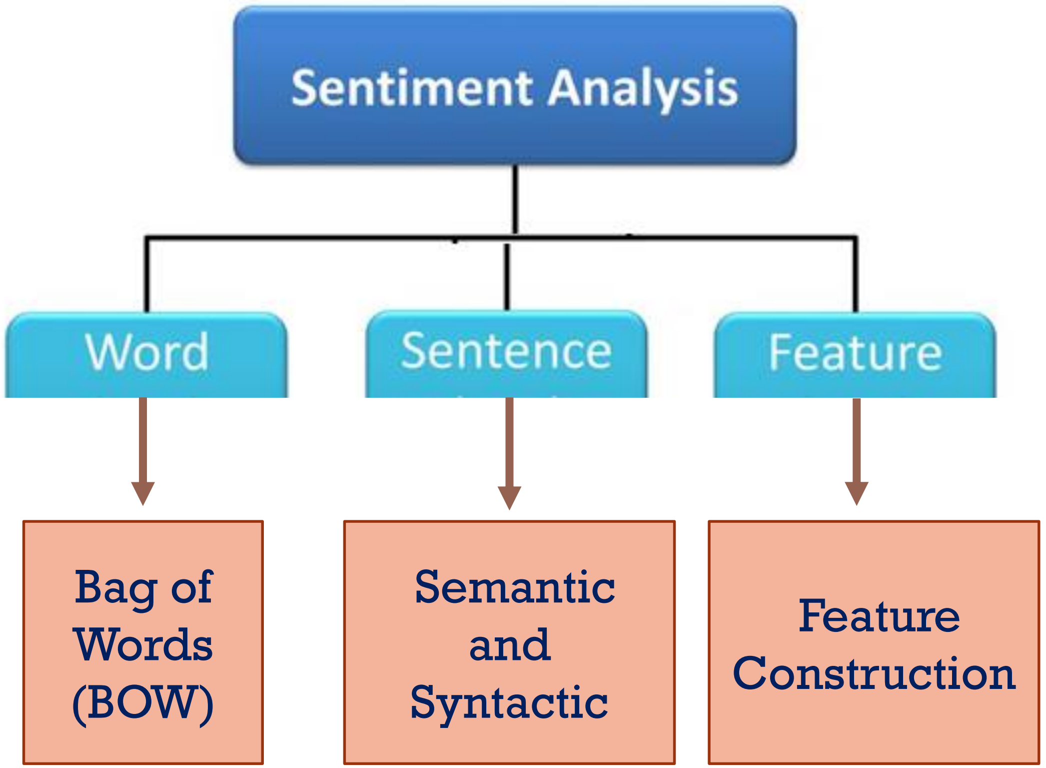
THE PROBLEMS WITH THE BOW APPROACH

- Human limitation with ontology— It is almost impossible to think of all the relevant keywords and their variants that represent a particular concept.

- Lack of domain expertise:
  - Many words that have a negative connotation in one context, may have a positive connotation in another context.

- The BoW Approach is unable to learn and improve itself.

Levels of Sentiment Analysis



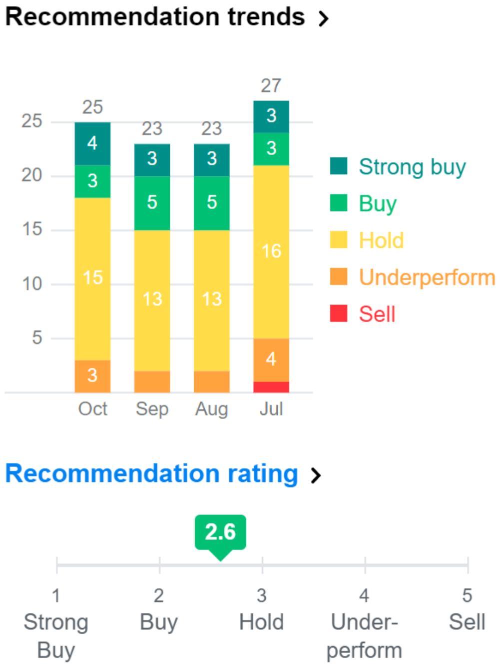
Semantic Patterns

Stock Market Area

Row	News Provider	Date	News Headline
1	Seeking Alpha	5/03/2014	Dr Pepper Snapple: 7 Different Insiders Have Sold Shares During The Last 30 Days
3	Investor's Business Daily	16/10/2014	Boeing May Sell Chinook Helicopters D11

Business Area

SOURCE OF DATA IN THE FINANCE LITERATURE



Earlier studies employed the dictionary-based approach in the textual sentiment literature.

- The General Inquirer (GI) built-in dictionary developed and used by Philip Stone (1966) a specialist in social psychology. GI includes **3,626** words.

- Perhaps the best-known polarity dictionary in finance built by Loughran and Mcdonald (2011) using 10-Ks and Bag of Words Method. That domain-specific dictionary includes **3,752** words.

General Polarity Dictionary			
General Inquirer (GI)	SentiWordNet	Diction	other
Tetlock (2007)	Malo et al. (2014)	Rogers et al. (2011)	Siganos et al. (2017)
Tetlock et al. (2008)	Nguyen et al. (2015)	Davis et al. (2012)	Ranco et al. (2016)
Engelberg (2008)	Mo et al. (2016)	Ferris et al. (2013)	Allen et al. (2017)
Feldman et al. (2008)	Salas Zarate et al. (2016)		
Kothari et al. (2009)	Chan & Chon (2017)		
Doran et al. (2012)	Yang et al. (2017)		
Ferris et al. (2013)			
Meyer et al. (2017)			
Jiang et al. (2017)			
Shapiro et al. (2018)			

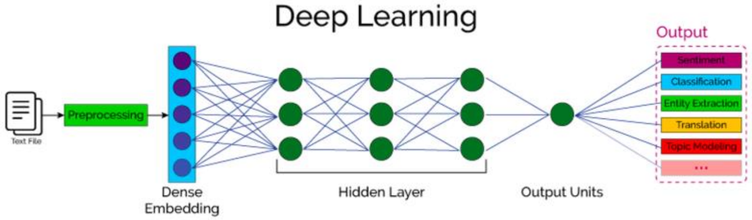
Researchers document that 73.8% of the negative words in the GI are not considered negative when used in the context of finance

General Polarity Dictionary				Domain-Specific Polarity Dictionary	
General Inquirer (GI)	SentiWordNet	Diction	other	Loughran & McDonald (L&M)	Manual Words List
Tetlock (2007)	Malo et al. (2014)	Rogers et al. (2011)	Siganos et al. (2017)	Loughran & McDonald (a2011)	Henry (2006)
Tetlock et al. (2008)	Nguyen et [illegible]	Davis et al. [illegible]	Ranco et al. (2016)	Rogers et al. (2011)	Li (2006)
Engelberg (2008)			[illegible] et al.	Loughran & McDonald (b2011)	Henry (2008)
Feldman et [illegible] (2008)				Jegadeesh and Wu (2012)	Rogers et al. (2011)
Kothari [illegible] (200[illegible]				Loughran & McDonald (2013)	Njolstad et al. (2014)
Doran et al. [illegible]				Liu & McConnell (2013)	Day & Lee (2016)
Ferris et al. (20[illegible]				Ferris et al. (2013)	Oliveira et al. (2017)
Meyer et al. (2017)				Garcia (2013)	Zhang et al. (2017)
Jiang et al. (2017)				Malo et al. (2014)	Daniel et al. (2017)
Shapiro et al. (2018)				Prollochs et al. (2016)	Cerchiello et al. (2017)
				Tsai & Wang (2017)	Krishnamoorthy (2018)
				Shapiro et al. (2018)	
				Krishnamoorthy (2018)	

These “Domain-Specific” dictionaries do not provide sufficient accuracy (Malo, et al. 2014, Meyer, Bikdash and Dai 2017, and Chan and Chong 2017).

L&M and GI

Headline	GI	L&M
Boeing May Sell Chinook Helicopters D11	Sell (Neutral)	Sell (Neutral)
Dr Pepper Snapple: 7 Different Insiders Have Sold Shares During The Last 30 Days	Sold (Neutral)	Sold (Neutral)



- Machine learning, Li (2010).
  - A proportion of the complete corpus of text to be analyzed is designated as the **‘training set’.** Each word in the training set is manually classified as ‘positive’, ‘negative’.
  - A selection of sentiment analysis algorithms is then trained on the training corpus.
  - The algorithms ‘learn’ the sentiment classification rules (or ‘grammar’) from the pre-classified data set.
  - Apply these rules out-of-sample to the whole corpus.

- Das and Chen (2007) use five algorithms to classify internet messages into bullish, bearish or neutral.
- Sinha (2010) uses the Reuters NewsScope Sentiment Engine to calculate the probabilities of news articles being positive, negative and neutral, respectively.

Machine Learning in literature

Method	Paper
Naïve Bayesian	Antweiler & Frank (2004)
Naïve Bayesian	Das & Chen (2007)
Naïve Bayesian	Feng Li(2010)
Support Vector Machine (SVM)	Hagenau et al (2013)
Support Vector Machine (SVM)	Smailovic et al (2014)
Support Vector Machine (SVM)	Malo et al (2014)
Othar	Prollochs et al (2016)
Othar	Oliveira et al (2016)
Support Vector Regression (SVR)	Tsai & Wang (2017)
Support Vector Machine (SVM)	Chan & Chon (2017)
Naïve Bayesian	Zhang et al (2017)
Support Vector Machine (SVM)	Meyer et al (2017)
Support Vector Regression (SVR)	Jiang et al (2017)
Neural Network	Lim et al (2018)

Sentiment Analysis Level in Literature

BOW		Syntactic and Semantic	Feature
Henry (2006) Li (2006)	Liu & McConnell (2013) Ferris et al. (2013)	Malo et al. (2014) Chan & Chon (2017)	Nguyen et al. (2015) Salas Zarate et al. (2016)
Tetlock (2007)	Garcia (2013)	Meyer et al. (2017)	
Das & Chen (2007)	Day & Lee (2016)		
Henry (2008)	Prollochs et al. (2016)		
Tetlock et al. (2008)	Oliveira et al. (2016)		
Engelberg (2008)	Ranco et al. (2016)		
Feldman et al. (2008)	Oliveira et al. (2017)		
Sinha (2010)	Tsai & Wang (2017)		
Loughran & McDonald (a2011)	Zhang et al. (2017)		
Rogers et al (2011)	Daniel et al. (2017)		
Loughran & McDonald (b2011)	Cerchiello et al. (2017)		
Jegadeesh and Wu (2012)	Yang et al. (2017)		
Doran et al. (2012)	Allen et al. (2017)		
Davis et al. (2012)	Siganos et al. (2017)		
Loughran & McDonald (2013)			

Sentiment Analysis Level in Literature

BOW		Syntactic and Semantic	Feature
Henry (2006)	Liu & McConnell (2013)	Malo et al. (2014)	Nguyen et al. (2015)
Li (20	t al. (2013)	Chan & Chon (2017)	
T	)	Meyer et al. (20	
Loug	)		
Jegad	7)		
Doran et	al. (2017)		
Davis et al. (2012)	Siganos et al. (2017)		
Loughran & McDonald (2013)			

Issues with the exiting ML literature:

1. The small size of ***<u>training data</u>*** for machine learning
2. Tagging financial concepts directional-dependence without considering them in financial text (tagging tokens in sentences).
3. The small size of financial concepts word lists.

The improvements were limited.
E.g. Malo, et al. 2014:

1. They consider 347 financial concepts from Investopedia.com with directional-dependence (252 Positive-If-Up and 95 Negative-If-Up).
2. Used 5,000 tagged news headlines
3. Reported improvement in the accuracy index is from 0.74 to 0.882

HOW IS ACADEMIC FINANCE RESEARCH DOING COMPARED TO LEADING INDUSTRY RESEARCH?

Tesla, the data company

Tesla is harnessing artificial intelligence and machine learning to build one of the most innovative neural networks in the world.

With 600,000 cars on the road, Tesla treats each vehicle, each sensor, each “event” (i.e. human interaction with the steering wheel, brake pedals, etc.) as data points.

It is then taking that data, analyzing it and utilizing it to improve its algorithms, create new algorithms and send those improvements over the air to the vehicles.

As of November 2018, Tesla has amassed 1 billion miles of Autopilot data. For comparison, Waymo has collected about 15 million miles.

The Startup That Could Help GM Beat Google to the Self-Driving Car

real-world examples to train the system. That's why Ford invested a billion dollars into artificial intelligence outfit Argo AI, why General Motors bought a startup called Cruise, why Waymo has driven 10 million autonomous miles on public roads (and billions more in simulation). Safe driving requires more than just knowing that a person is over there; you also have to know that said person is riding a bicycle, how they're likely to act, and how to respond. That's hard for a robot, but these budding Terminators are getting better, fast.

MAIN FINDINGS

- Sentiment in texts does convey incremental information over quantitative financial information.

- Sentiment in texts might have power in predicting market movements.

- The results are more robust for negative sentiments

- Sentiment is a pricing factor in addition to risk premia and other firm-level characteristics.

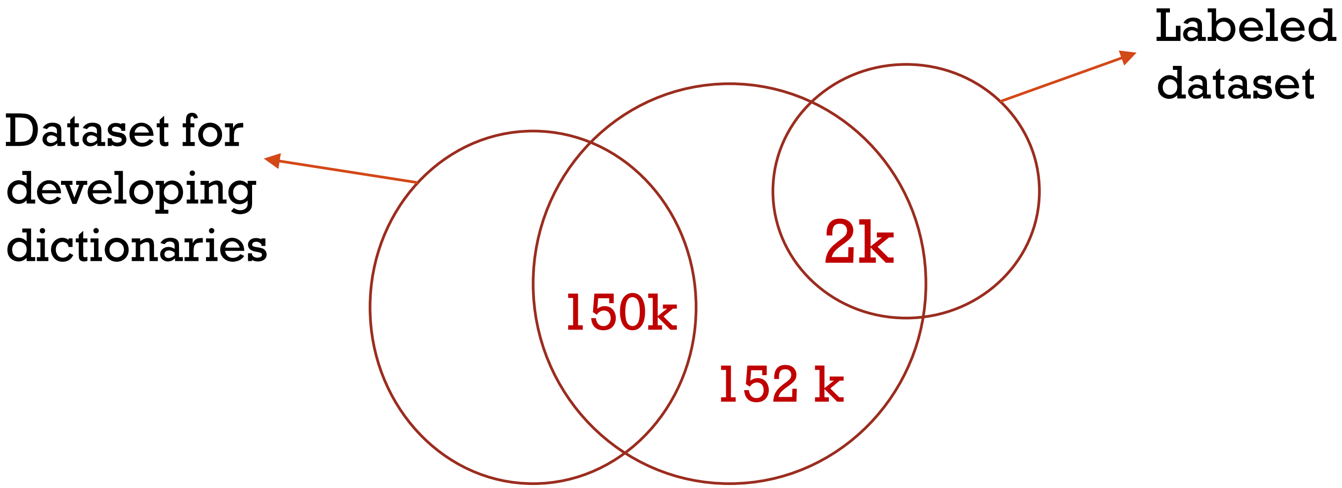
WHEN SENTIMENT IS NEWS

Nazanin Babolmorad

Nadia Massoud

Data Description

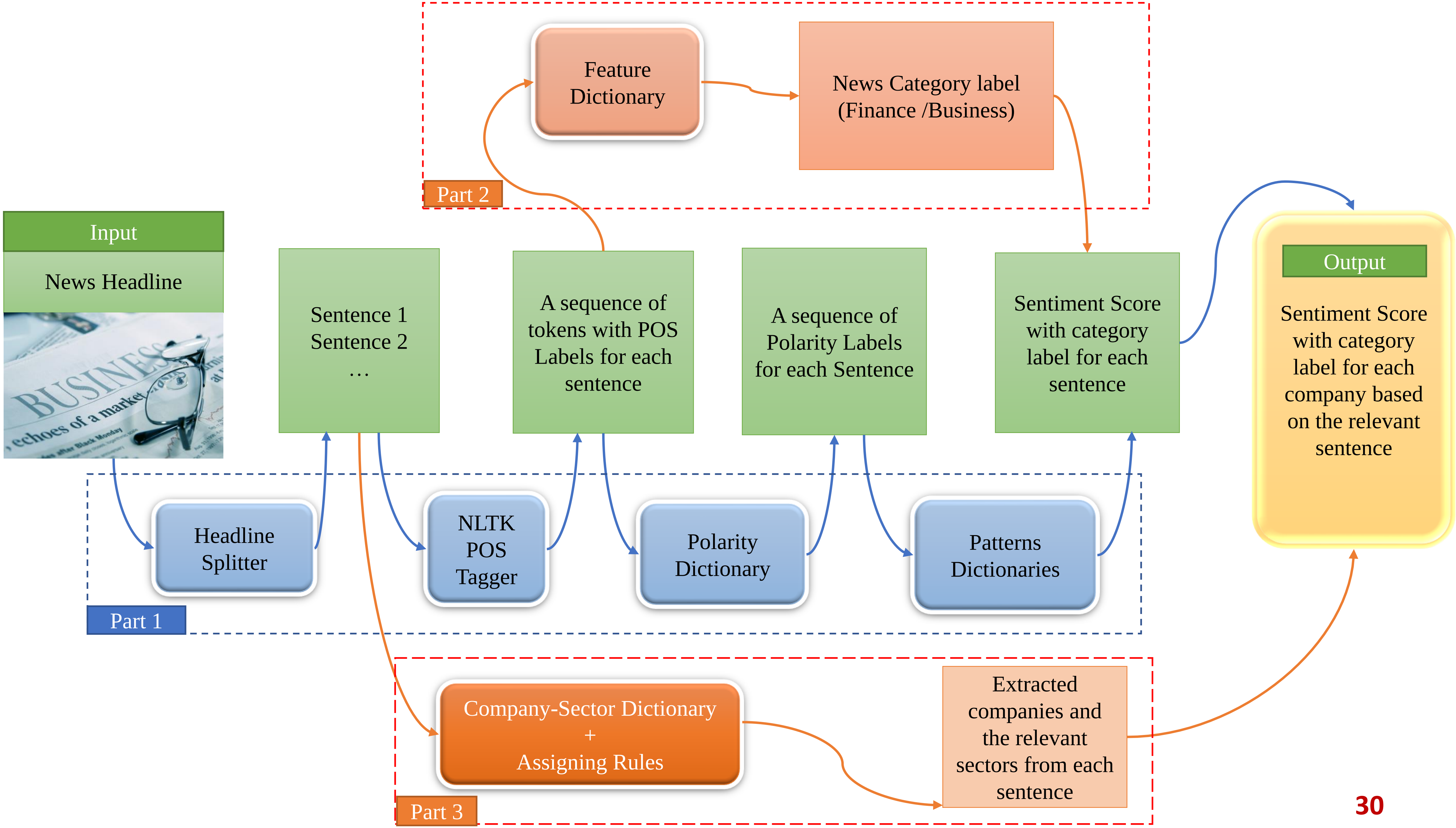
- ✓ We collect 304k news headlines From January 2014 to January 2019 to generate a meaningful daily sentiment index per company and choose randomly 150k of the collected headlines to create our dictionaries.
- ✓ 2k news headlines are randomly selected and have been tagged to cross validation phase. Those headlines are annotated by 21 Auditors and this sub-set is called Labeled Dataset.



Data Description

- ✓ Dataset is collected from 15 leading news providers: Seeking Alpha, Zacks, Wall Street Journal, Market Realist, Motley Fool, Yahoo Finance, Reuters, Bloomberg, InvestorPlace, Investor's Business Daily, GuruFocus, 247WallSt, Barron's, Fox Business, and Benzinga.

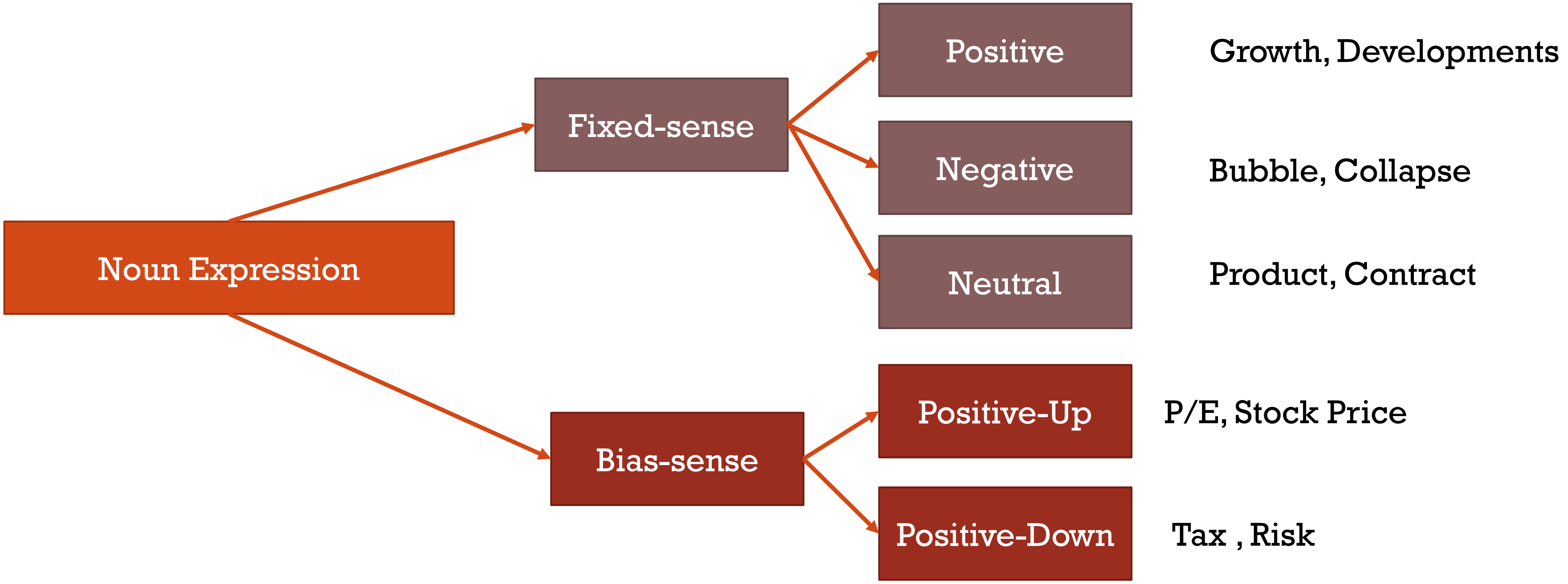
- ✓ We search the digital archives of each newspaper from January 2014 to January 2019 to obtain a daily count of news headlines that labeled by the news providers for 163 companies on the US Market.


Part 2
Feature Dictionary
News Category label (Finance /Business)
Input
News Headline
Sentence 1
Sentence 2
…
A sequence of tokens with POS Labels for each sentence
A sequence of Polarity Labels for each Sentence
Sentiment Score with category label for each sentence
Output
Sentiment Score with category label for each company based on the relevant sentence
Headline Splitter
NLTK POS Tagger
Polarity Dictionary
Patterns Dictionaries
Part 1
Company-Sector Dictionary
+
Assigning Rules
Extracted companies and the relevant sectors from each sentence
Part 3

Polarity Dictionary

Nouns and Noun Phrases

2,721 Nouns and nouns-phrases are extracted Based on frequency in headlines (75%).

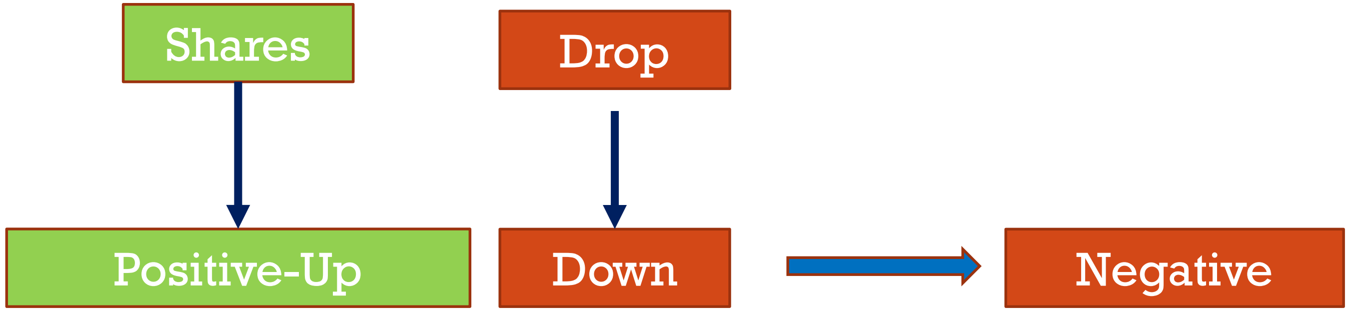


Polarity Dictionary

Noun:

Stock Market Area

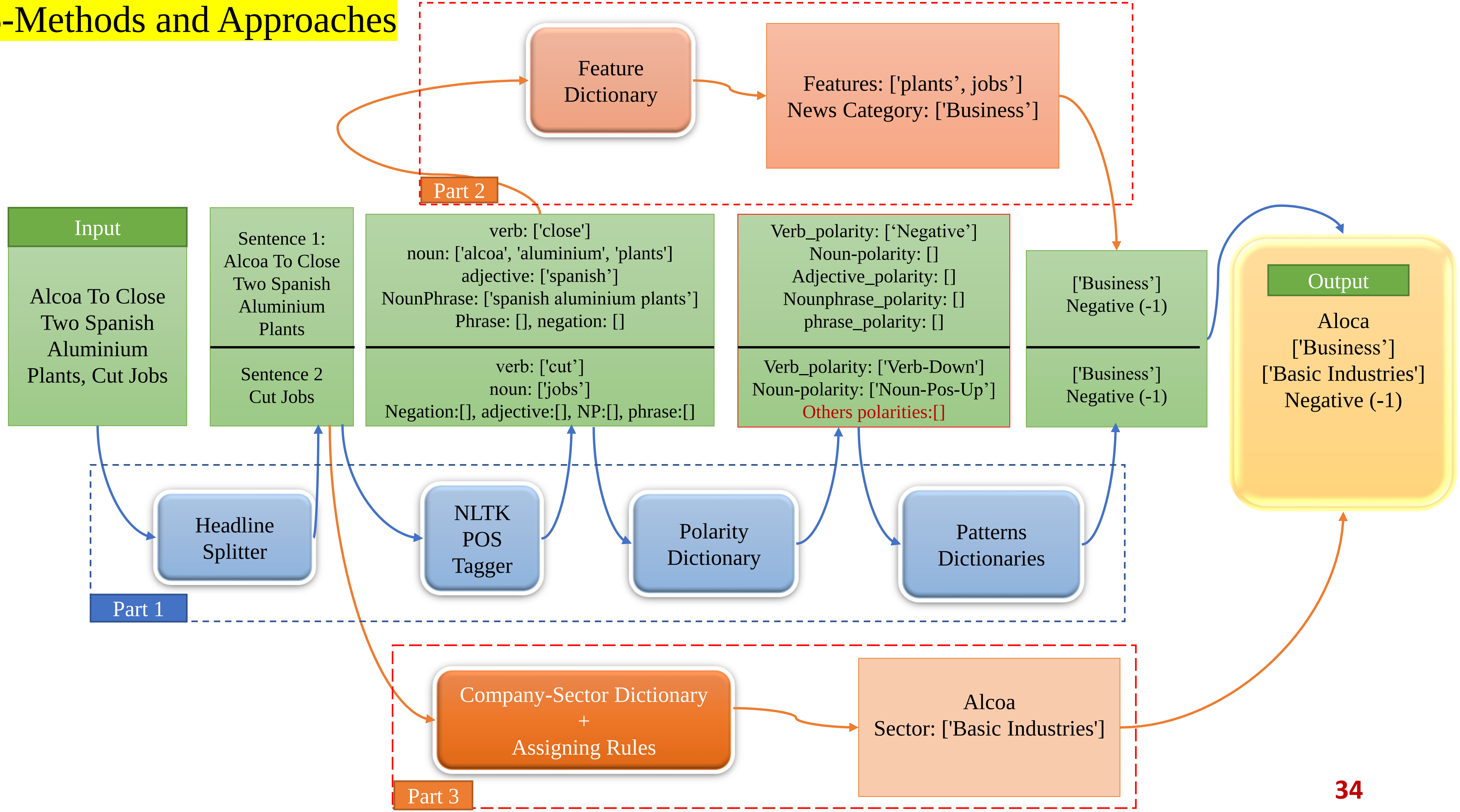
News Provider	Date	News Headline
247 WallSt	3/11/2015	Could Chipotle Shares Drop Another $200?



Our Approach VS. BOW Using L&M and GI

Headline	Our Polarity Dictionary	Our feature Dictionary	Final Label	L&M	GI
Boeing May Sell Chinook Helicopters D11	Sell (Positive)	Helicopters (Business)	Positive	Neutral	Neutral
Dr Pepper Snapple: 7 Different Insiders Have Sold Shares During The Last 30 Days	Sold (Negative)	Insiders (Finance)	Negative	Neutral	Neutral

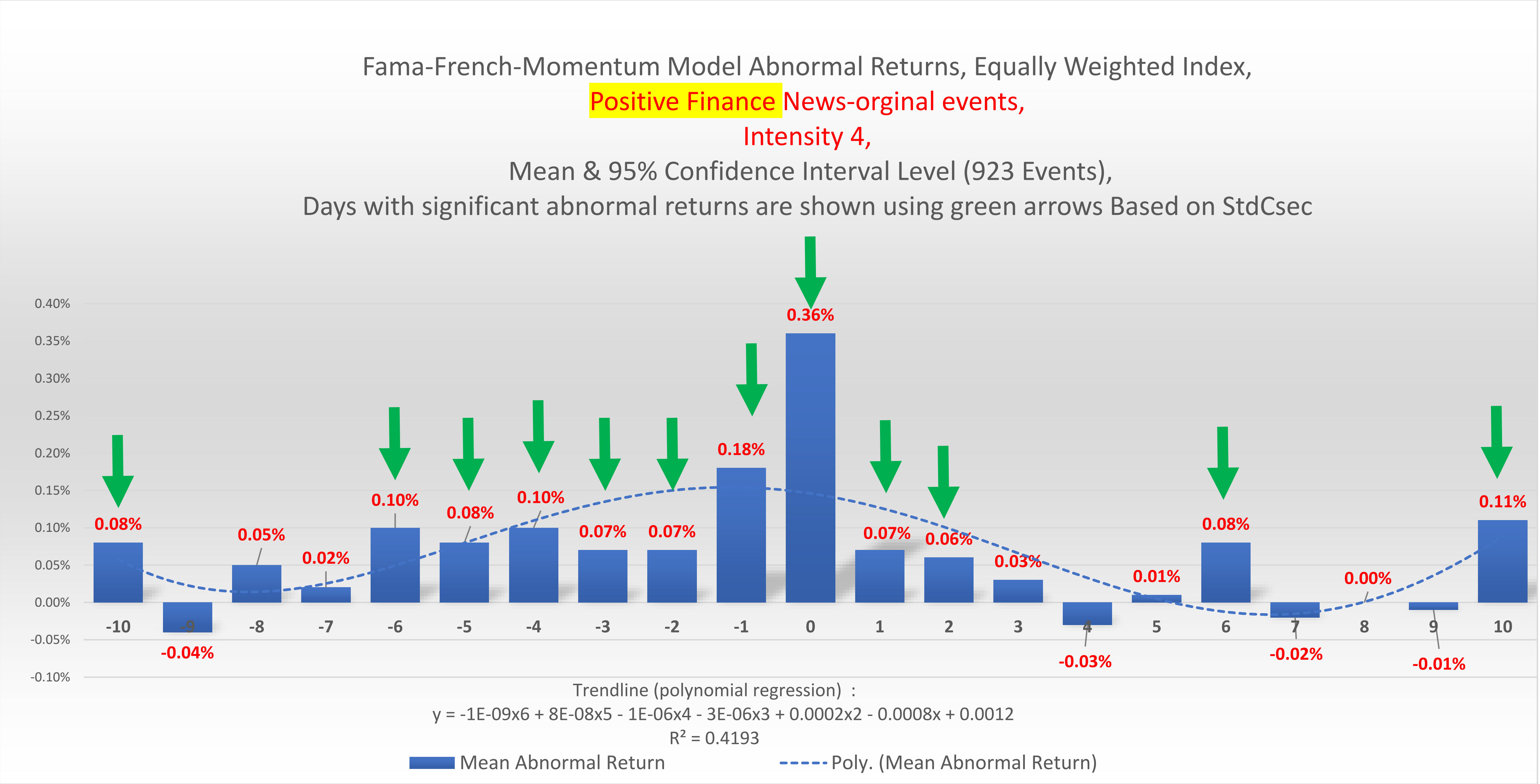
6-Methods and Approaches

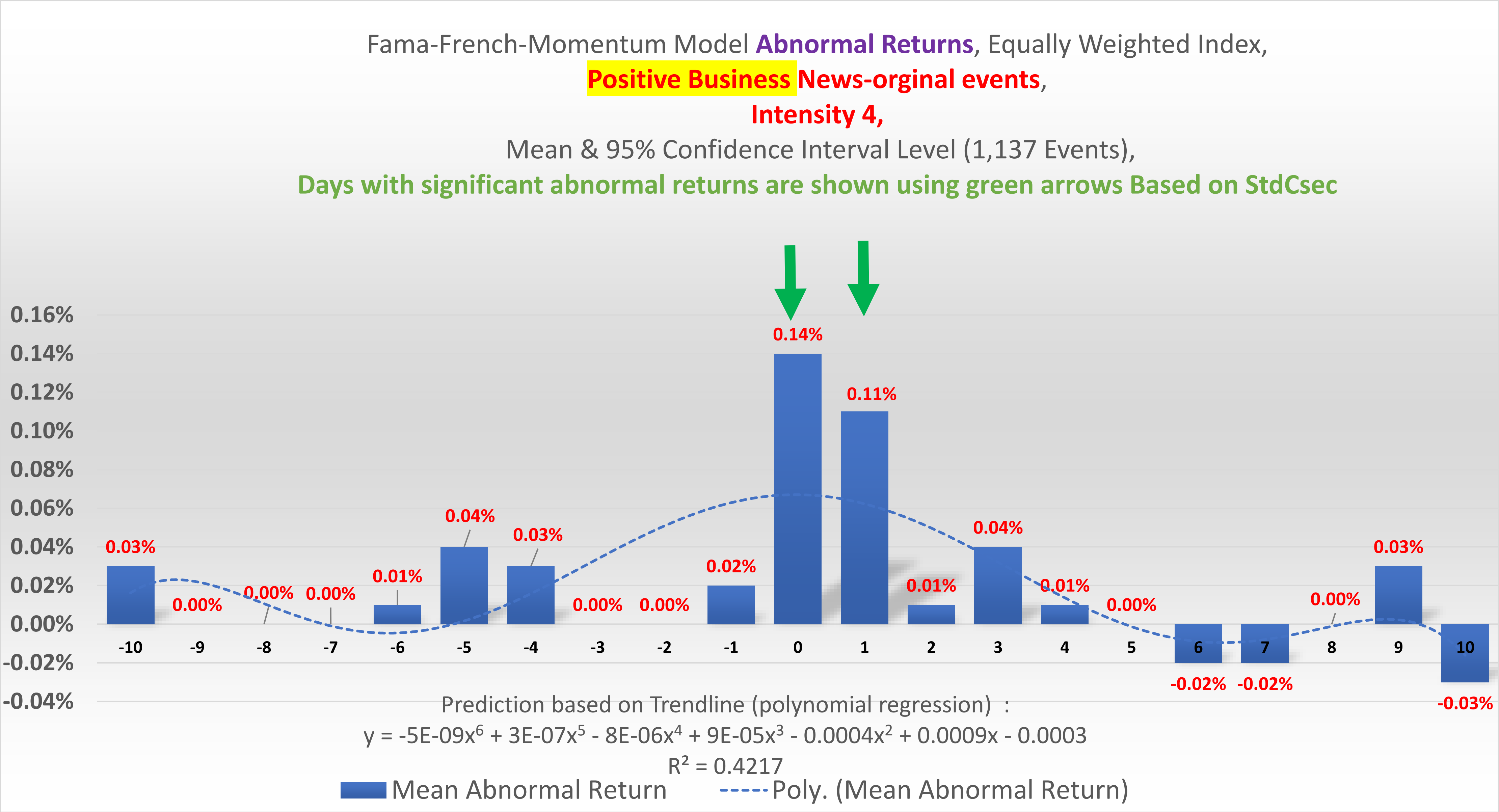


Accuracy of The Proposed Approach in Sentiment Analysis of News Headlines

Based on the 2K human tagged news headlines:

	Positive		Neutral		Negative	
	RECALL	PRECISION	RECALL	PRECISION	RECALL	PRECISION
Our Method	0.89	0.71	0.83	0.91	0.88	0.7
BOW (LM-Dictionary)	0.61	0.55	0.72	0.78	0.51	0.49


Fama-French-Momentum Model Abnormal Returns, Equally Weighted Index,
Positive Finance News-orginal events,
Intensity 4,
Mean & 95% Confidence Interval Level (923 Events),
Days with significant abnormal returns are shown using green arrows Based on StdCsec
0.40%
0.35%
0.30%
0.25%
0.20%
0.15%
0.10%
0.05%
0.00%
-0.05%
-0.10%
0.08%
-0.04%
0.05%
0.02%
0.10%
0.08%
0.10%
0.07%
0.07%
0.18%
0.36%
0.07%
0.06%
0.03%
-0.03%
0.01%
0.08%
-0.02%
0.00%
-0.01%
0.11%
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Trendline (polynomial regression) :
y = -1E-09x6 + 8E-08x5 - 1E-06x4 - 3E-06x3 + 0.0002x2 - 0.0008x + 0.0012
R² = 0.4193
Mean Abnormal Return
Poly. (Mean Abnormal Return)



Fama-French-Momentum Model Abnormal Returns, Equally Weighted Index,
Positive Business News-orginal events,
Intensity 4,
Mean & 95% Confidence Interval Level (1,137 Events),
Days with significant abnormal returns are shown using green arrows Based on StdCsec
0.16%
0.14%
0.12%
0.10%
0.08%
0.06%
0.04%
0.02%
0.00%
-0.02%
-0.04%
0.03%
0.00%
0.00%
0.00%
0.01%
0.04%
0.03%
0.00%
0.00%
0.02%
0.14%
0.11%
0.01%
0.04%
0.01%
0.00%
-0.02%
-0.02%
0.00%
0.03%
-0.03%
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
Prediction based on Trendline (polynomial regression) :
$y = -5E-09x^6 + 3E-07x^5 - 8E-06x^4 + 9E-05x^3 - 0.0004x^2 + 0.0009x - 0.0003$
$R^2 = 0.4217$
Mean Abnormal Return
Poly. (Mean Abnormal Return)

AR(0) Regressions with R(-1) to R(-3), AR(-1) to AR(-3), polarity, and news categories, across the quartiles of market capitalization in ascending order,

Report robust standard errors, clustered by industry, includes year fixed effect

t-statistics are in parentheses: * p<0.1, ** p<0.05, *** p<0.01

Model	market capitalization Q1	market capitalization Q2	market capitalization Q3	market capitalization Q3
DV	AR(0)	AR(0)	AR(0)	AR(0)
Business sentiment NEG	-0.4%***	-0.2%***	-0.1%***	-0.1%***
	(-3.21)	(-3.24)	(-3.33)	(-3.12)
Business sentiment POS	0.1%	0.0%	0.0%	0.0%
	(1.20)	(0.03)	(1.60)	(-0.19)
Finance Sentiment NEG	-0.5%***	-0.3%***	-0.2%***	-0.2%***
	(-2.92)	(-5.83)	(-3.26)	(-6.82)
Finance Sentiment POS	0.3%***	0.1%***	0.1%***	0.1%***
	(5.91)	(4.19)	(5.47)	(4.83)
Mixed sentiment NEG	-1.0%***	-0.4%***	-0.2%***	-0.2%***
	(-5.29)	(-3.71)	(-7.28)	(-3.51)
Mixed sentiment POS	0.4%***	0.1%***	0.1%***	0.1%***
	(5.00)	(4.58)	(3.17)	(4.46)
BUS. NEG x BUS. POS	0.0%	0.0%	0.0%	0.0%
	(0.50)	(-0.57)	(1.60)	(1.23)
BUS. NEG x FIN. POS	0.1%	0.0%	0.0%	0.0%
	(0.41)	(0.81)	(-1.57)	(-1.63)
BUS. POS x FIN. NEG	0.2%	0.0%	0.0%	0.000**
	(1.07)	(-0.39)	(0.32)	(2.20)
FIN. NEG x FIN. POS	-0.1%	0.0%	-0.000**	0.0%
	(-1.03)	(0.20)	(-2.58)	(0.87)
Constant	-0.1%	0.0%	0.0%	0.0%
	(-1.16)	(-0.04)	(-0.87)	(1.63)
Observations	14,408	14,183	14,269	14,243
Adjusted R-squared	0.014	0.011	0.019	0.015

log-AV(0) Regressions with V(-1) to V(-3), AV(-1) to AV(-3), polarity, and news categories,

Robust standard errors are clustered by industry,
t-statistics are in parentheses: * p<0.1, ** p<0.05, *** p<0.01

Model	(1)	(2)
DV	log-AV	log-AV
Business sentiment NEG	13.3%***	6.1%**
	(3.67)	(2.17)
Business sentiment POS	8.8%***	3.5%*
	(3.07)	(1.76)
Finance Sentiment NEG	8.1%***	6.3%***
	(4.53)	(4.52)
Finance Sentiment POS	4.3%***	2.3%*
	(2.84)	(1.88)
Mixed sentiment NEG	12.5%***	9.2%***
	(5.04)	(5.07)
Mixed sentiment POS	5.4%**	2.6%
	(2.70)	(1.52)
BUS. NEG x BUS. POS	-1.5%***	-0.6%
	(-3.37)	(-1.62)
BUS. NEG x FIN. POS	-0.9%***	-0.4%
	(-2.84)	(-1.51)
BUS. POS x FIN. NEG	-1.5%***	-0.8%*
	(-3.05)	(-1.80)
FIN. NEG . FIN. POS	-0.5%	-0.4%*
	(-1.40)	(-1.85)
Year fixed effect	yes	yes
market cap. Quartile fixed effect		yes
Constant	6.757***	6.372***
	(100.36)	(43.43)
Observations	58,127	57,184
Adjusted R-squared	0.060	0.229

Contributions

✓Using abnormal returns and abnormal volume, our approach provides new results:

	Business (+)	Business(-)	Stock Mkt (+)	Stock Mkt (-)
AR	(.) ✗	(-) ✓	(+) ✓	(-) ✓
AV	(+) weak	(+) ✓	(+) ✓	(+) ✓

	Mixed type (+)	Mixed type (-)	Mixed Polarity & type
AR	(+) ✓	(-) ✓	(.) ✗
AV	(+) & (.)	(+) ✓	(-) & (.)

New Results

Contributions

- ✓We develop a new polarity dictionary and a feature dictionary in finance , specially for news headline.
- ✓We develop a pattern dictionary which includes 1,544 semantic patterns on news headlines.
- ✓We propose a new approach based on polarity dictionary and semantic Patterns to Sentiment Analysis of news headlines.
- ✓We increase precision of Sentiment Analysis of news headlines compared with literature.

FUTURE RESEARCH

- Developing new hypothesis
- Improved Machine learning approach
- Use new sources of data, social media, IOT
- Improve training data set. Keep in mind it has to be customised to the research question.

THANKS…