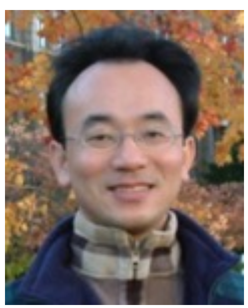
Tamkang University
淡江大學

Big Data Analytics for Financial Sentiment Analysis in FinTech
(金融科技財務大數據情感分析)

Time: 2016/11/15 (Tue) (10:25-12:10)
Place: 朝陽科技大學資訊工程研究所 <教室： E520 演講廳>
Host: 吳世弘 教授 (Professor Shih-Hung Wu)



Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail. tku.edu.tw/myday/

2016-11-15

戴敏育 博士
(Min-Yuh Day, Ph.D.)

淡江大學資管系專任助理教授

中央研究院資訊科學研究所訪問學人

國立台灣大學資訊管理博士

Publications Co-Chairs, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in TExt (IEEE EM-RITE 2012-)

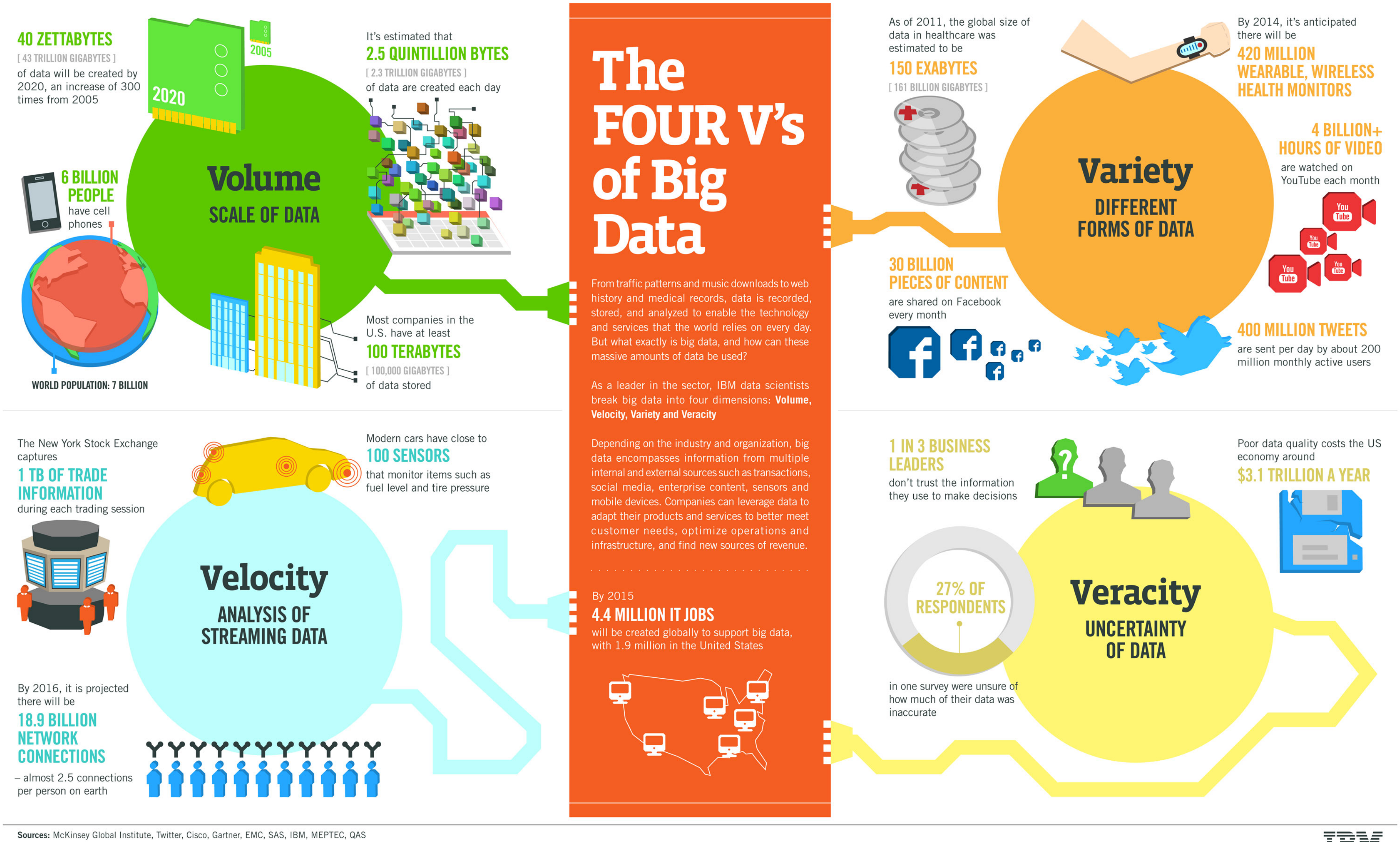
Workshop Chair, The IEEE International Conference on Information Reuse and Integration (IEEE IRI)

Outline

- Big Data Analytics
- FinTech
- Financial Sentiment Analysis

Big Data Analytics

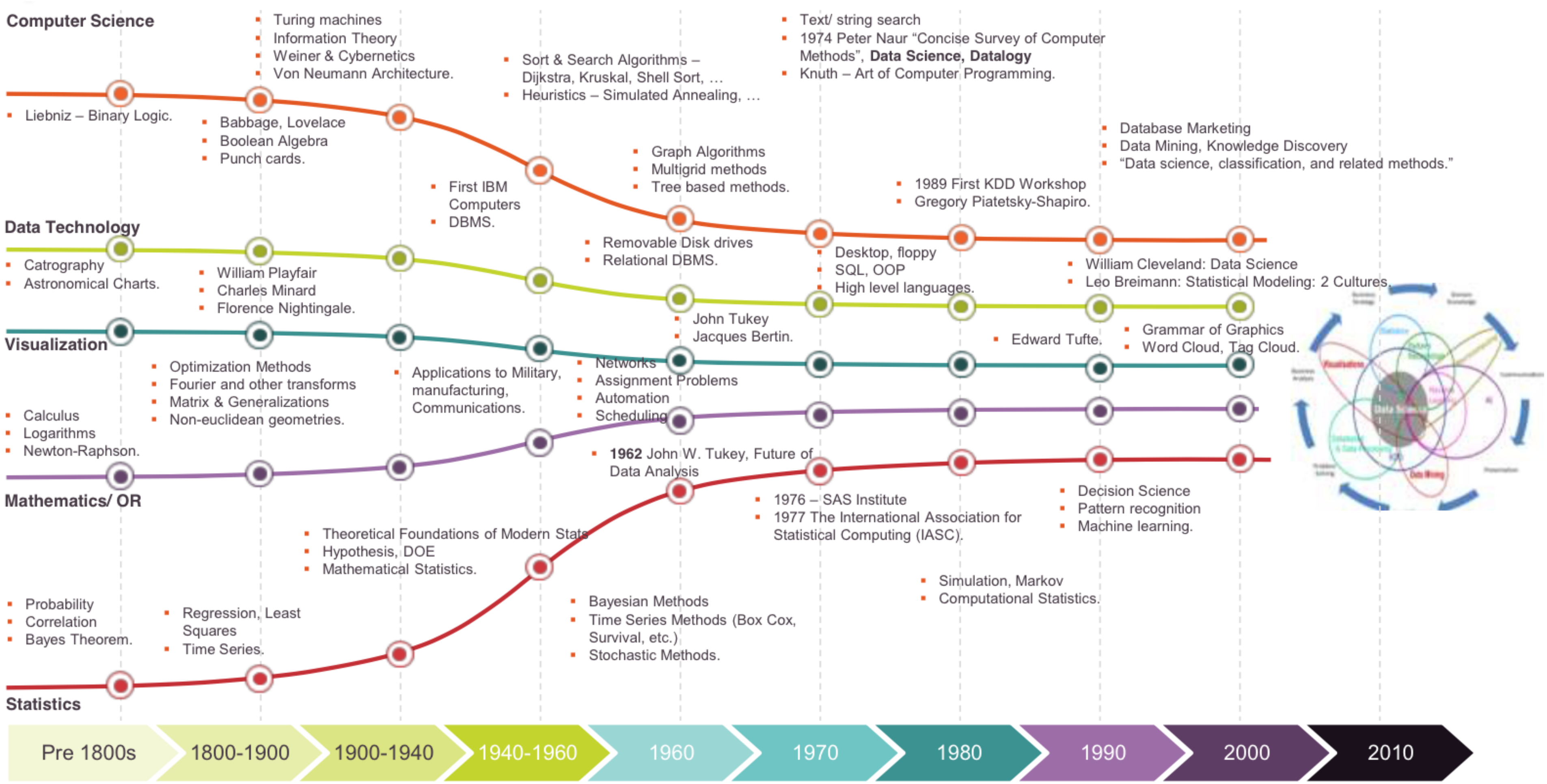
Big Data 4 V



Source: https://www-01.ibm.com/software/data/bigdata/

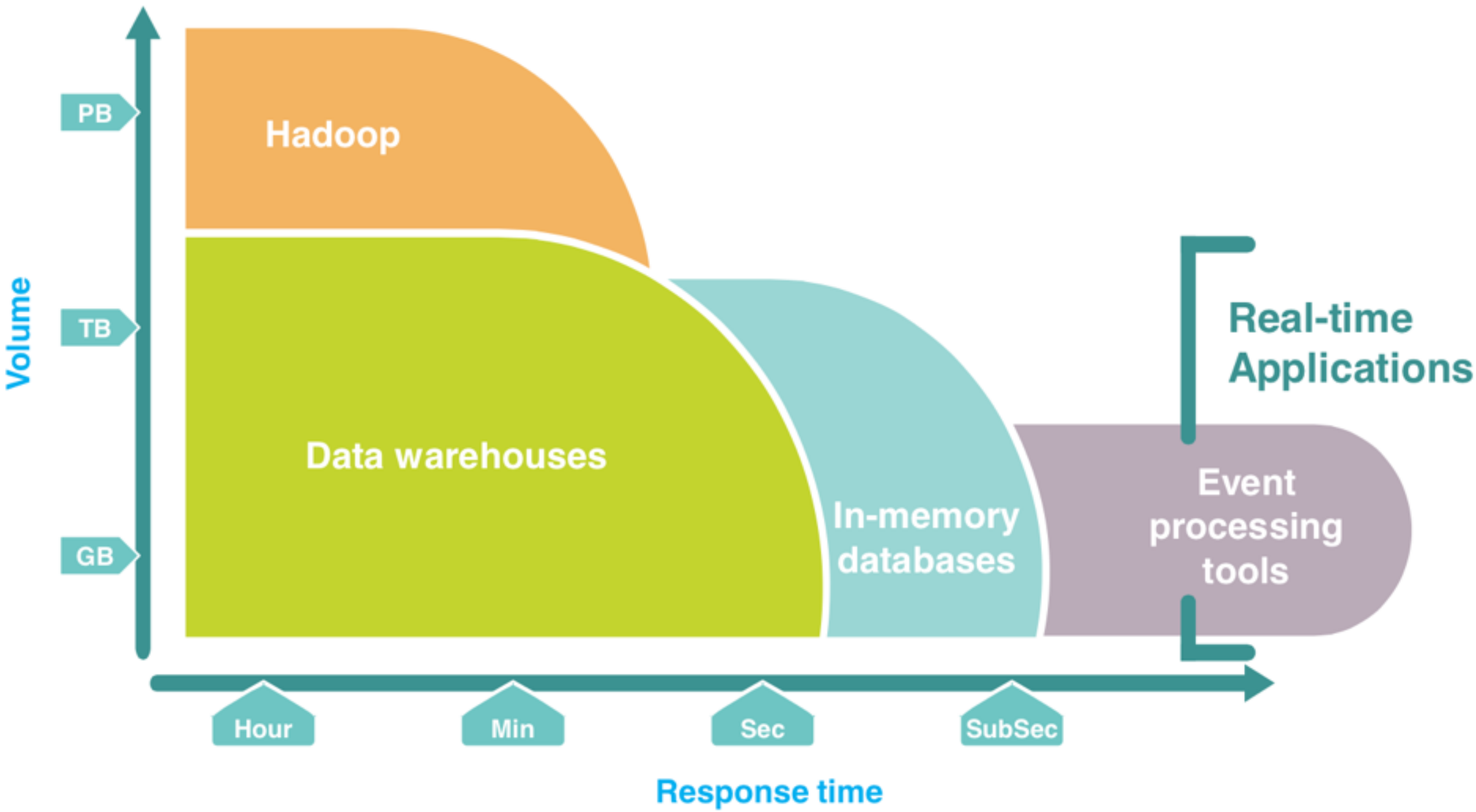
Value

History of Data Science



Source: http://www.kdnuggets.com/2015/02/history-data-science-infographic.html

Big Data Technologies are Enabling a New Approach



Source: http://www.doclens.com/119898/think-1-13-big-datas-impact-on-analytics/

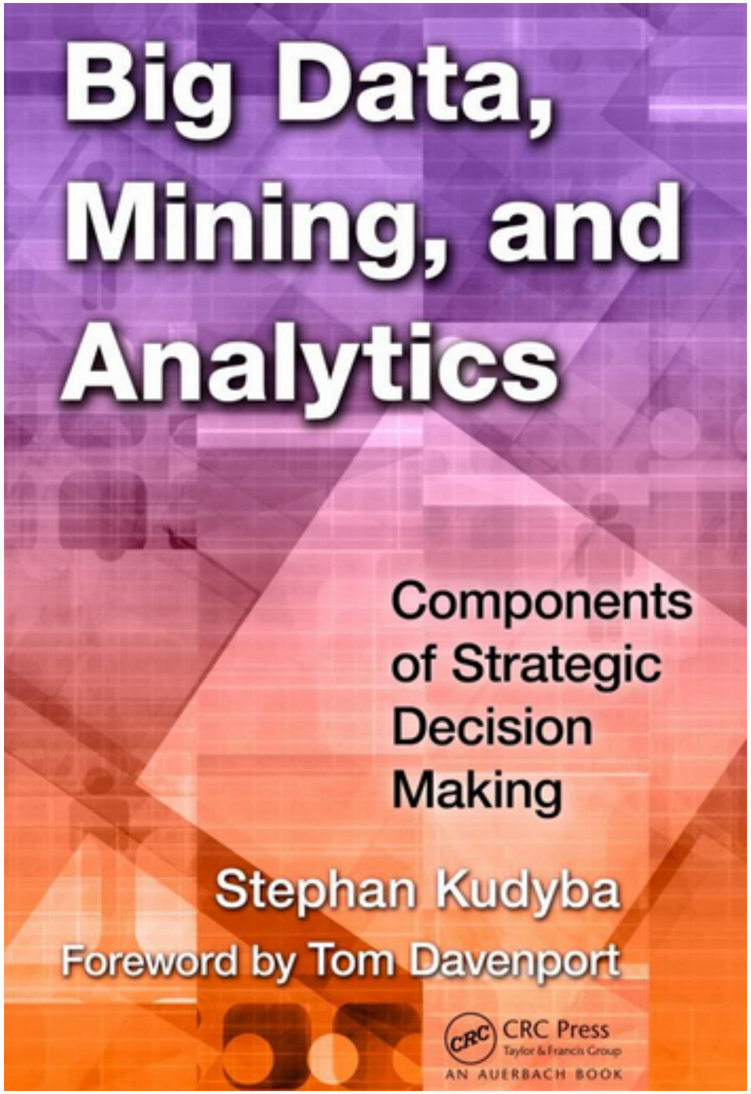
Big Data Analytics

and

Data Mining

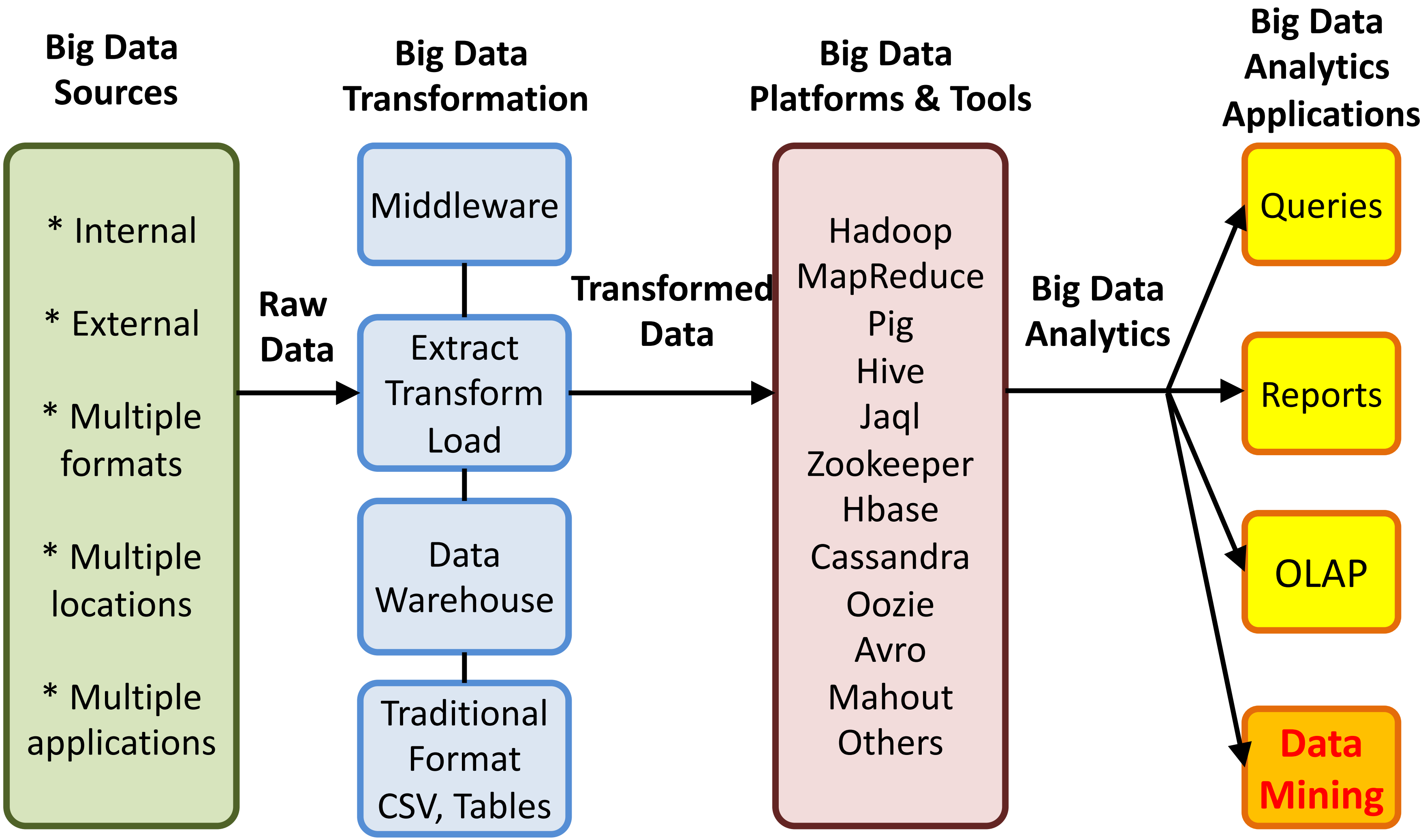
Stephan Kudyba (2014),

Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications



Source: http://www.amazon.com/gp/product/1466568704

Architecture of Big Data Analytics



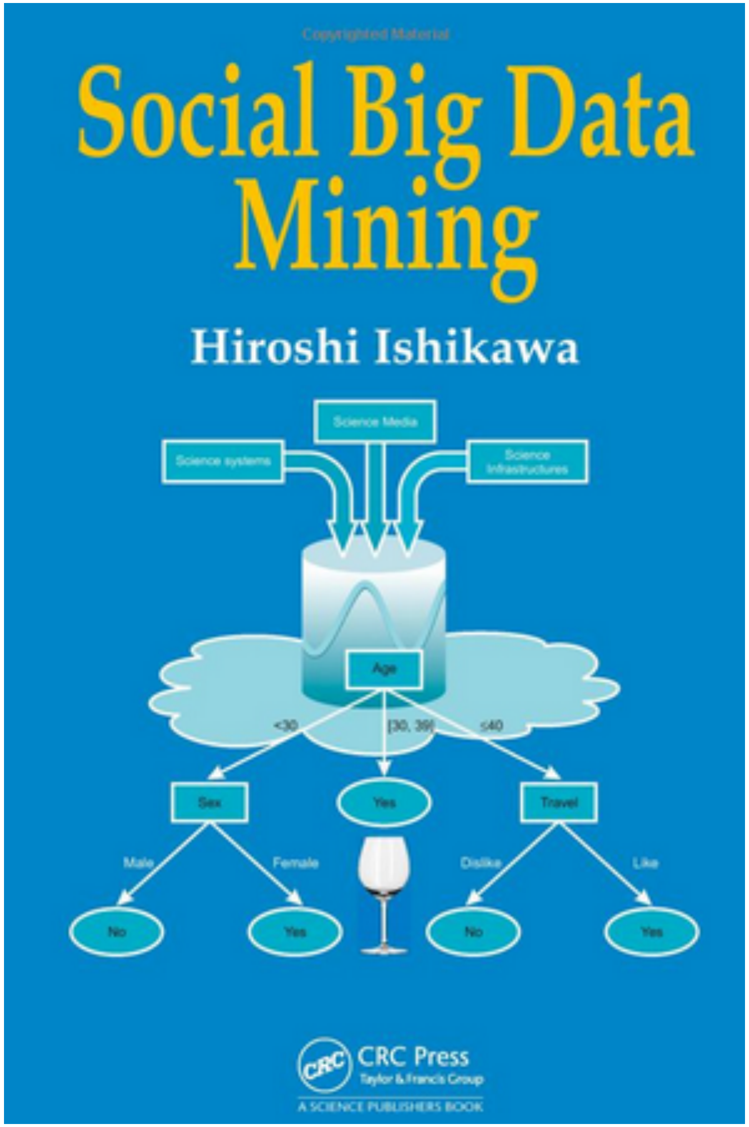
Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Architecture of Big Data Analytics

Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Social Big Data Mining

(Hiroshi Ishikawa, 2015)



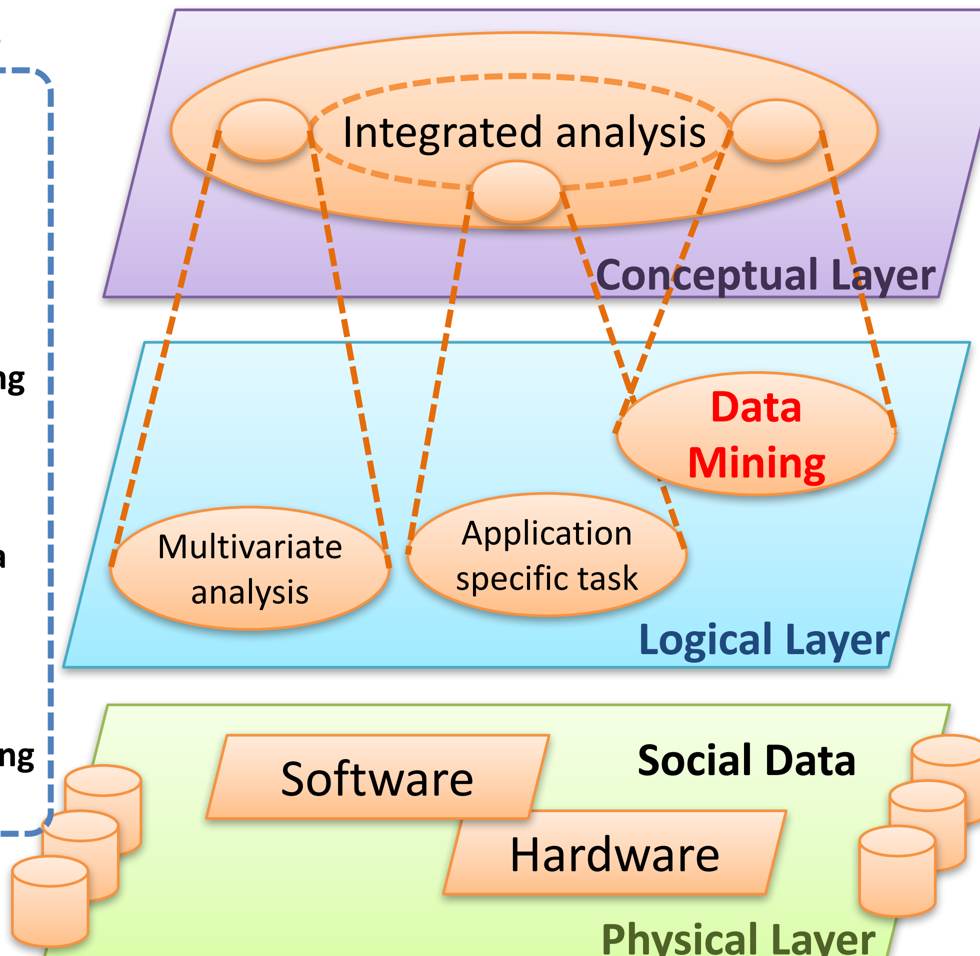
Source: http://www.amazon.com/Social-Data-Mining-Hiroshi-Ishikawa/dp/149871093X

Architecture for Social Big Data Mining

(Hiroshi Ishikawa, 2015)

Enabling Technologies

- Integrated analysis model
- Natural Language Processing
- Information Extraction
- Anomaly Detection
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distrusted processing

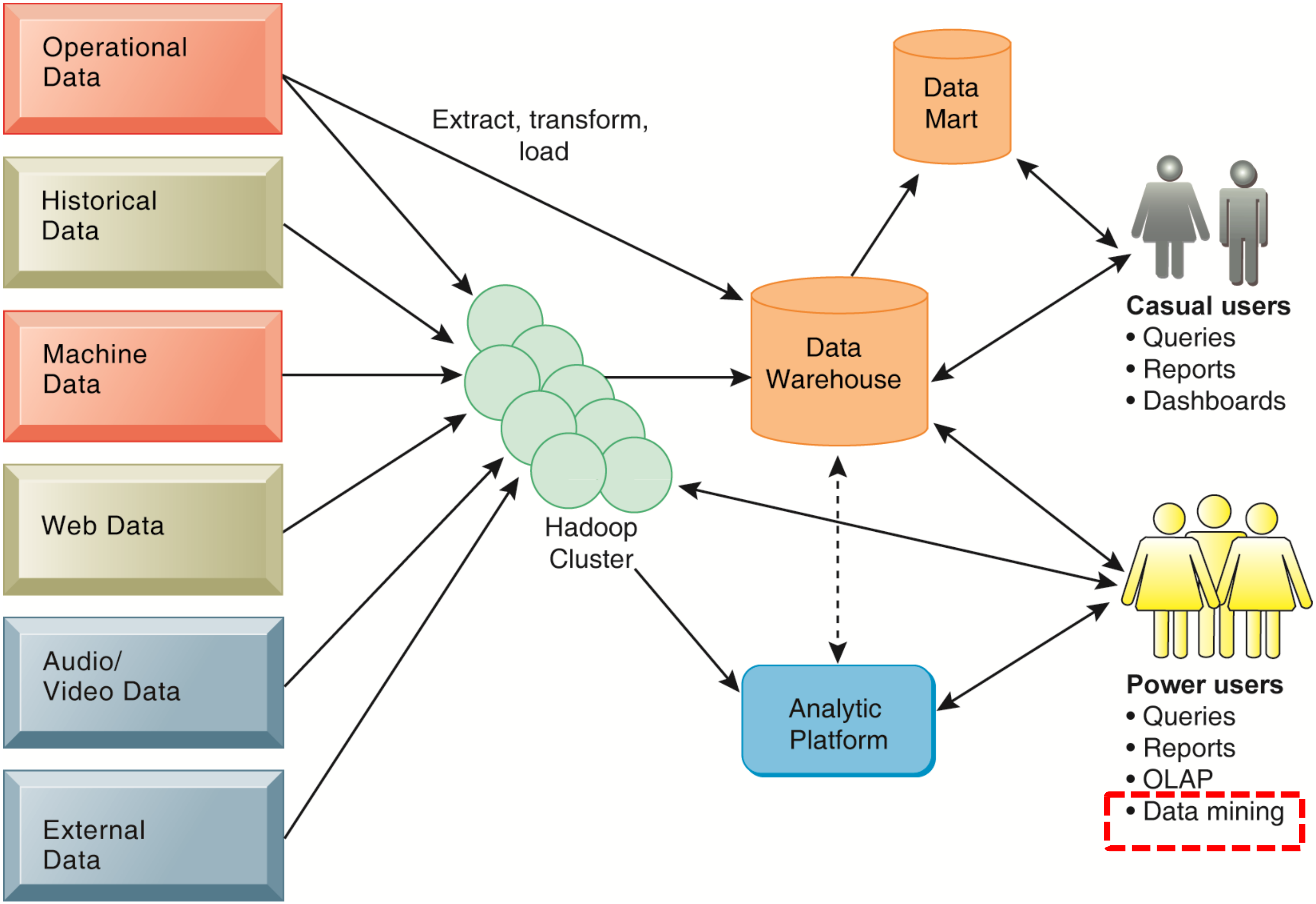


Analysts

- Model Construction
- Explanation by Model
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

Source: Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press

Business Intelligence (BI) Infrastructure



Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

References

- Jiawei Han and Micheline Kamber (2011),
  Data Mining: Concepts and Techniques, Third Edition, Elsevier
- Jennifer Golbeck (2013),
  Analyzing the Social Web, Morgan Kaufmann
- Stephan Kudyba (2014),
  Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications
- Hiroshi Ishikawa (2015),
  Social Big Data Mining, CRC Press

LeCun, Yann,
Yoshua Bengio,
and Geoffrey Hinton.

"**Deep learning**."

Nature 521, no. 7553 (2015): 436-444.

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

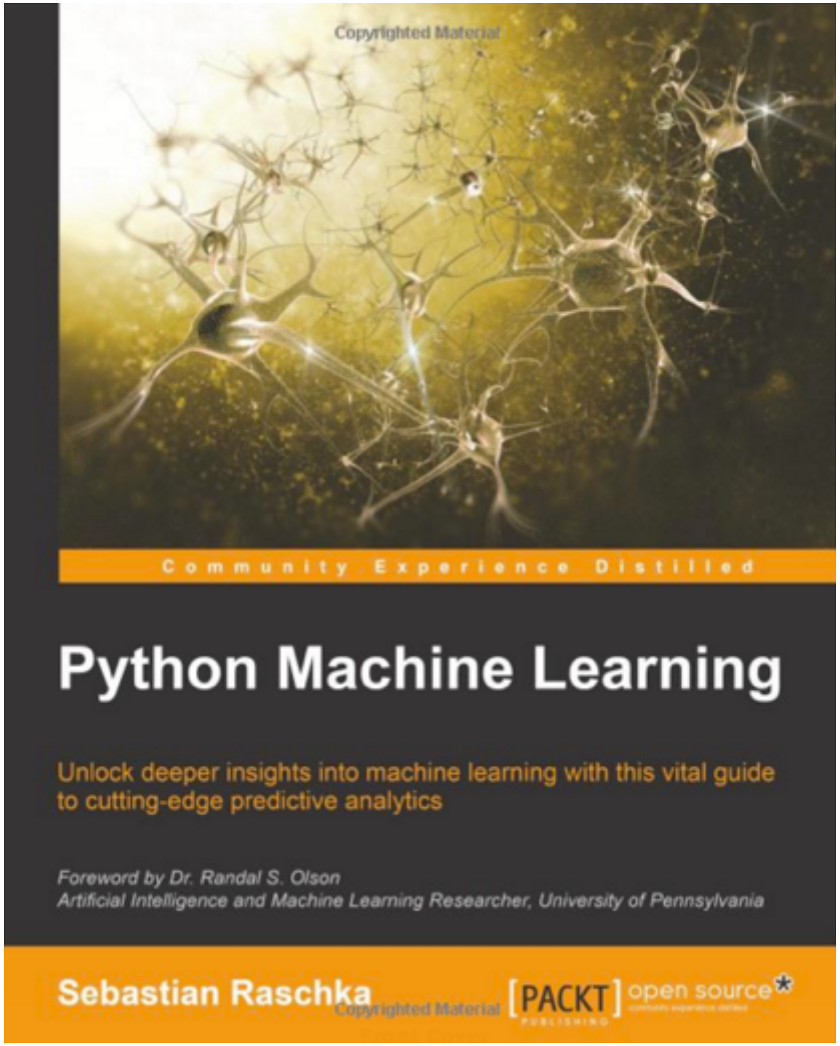
Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, con- intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

Source: LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

Sebastian Raschka (2015),

Python Machine Learning,

Packt Publishing

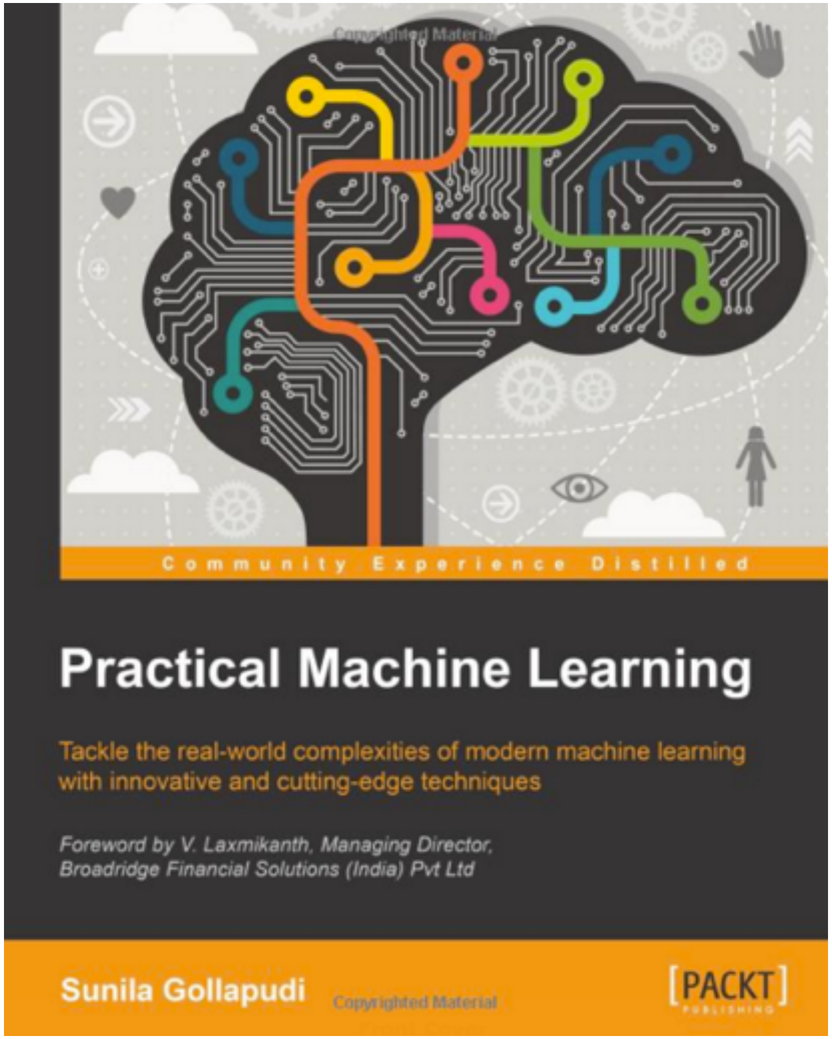


Source: http://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130

Sunila Gollapudi (2016),

Practical Machine Learning,

Packt Publishing



Source: http://www.amazon.com/Practical-Machine-Learning-Sunila-Gollapudi/dp/178439968X

Machine Learning Models

- Deep Learning
- Kernel
- Association rules
- Ensemble
- Decision tree
- Dimensionality reduction
- Clustering
- Regression Analysis
- Bayesian
- Instance based

Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing

Data Scientist
資料科學家

Source: http://www.ibmbigdatahub.com/infographic/what-makes-data-scientist

Deep Learning

Intelligence from Big Data

Source: https://www.vlab.org/events/deep-learning/

Big Data

Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Data Scientist:

The Sexiest Job of the 21st Century

(Davenport & Patil, 2012)(HBR)

Source: Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*

Data Scientist:

The Sexiest Job of the 21st Century

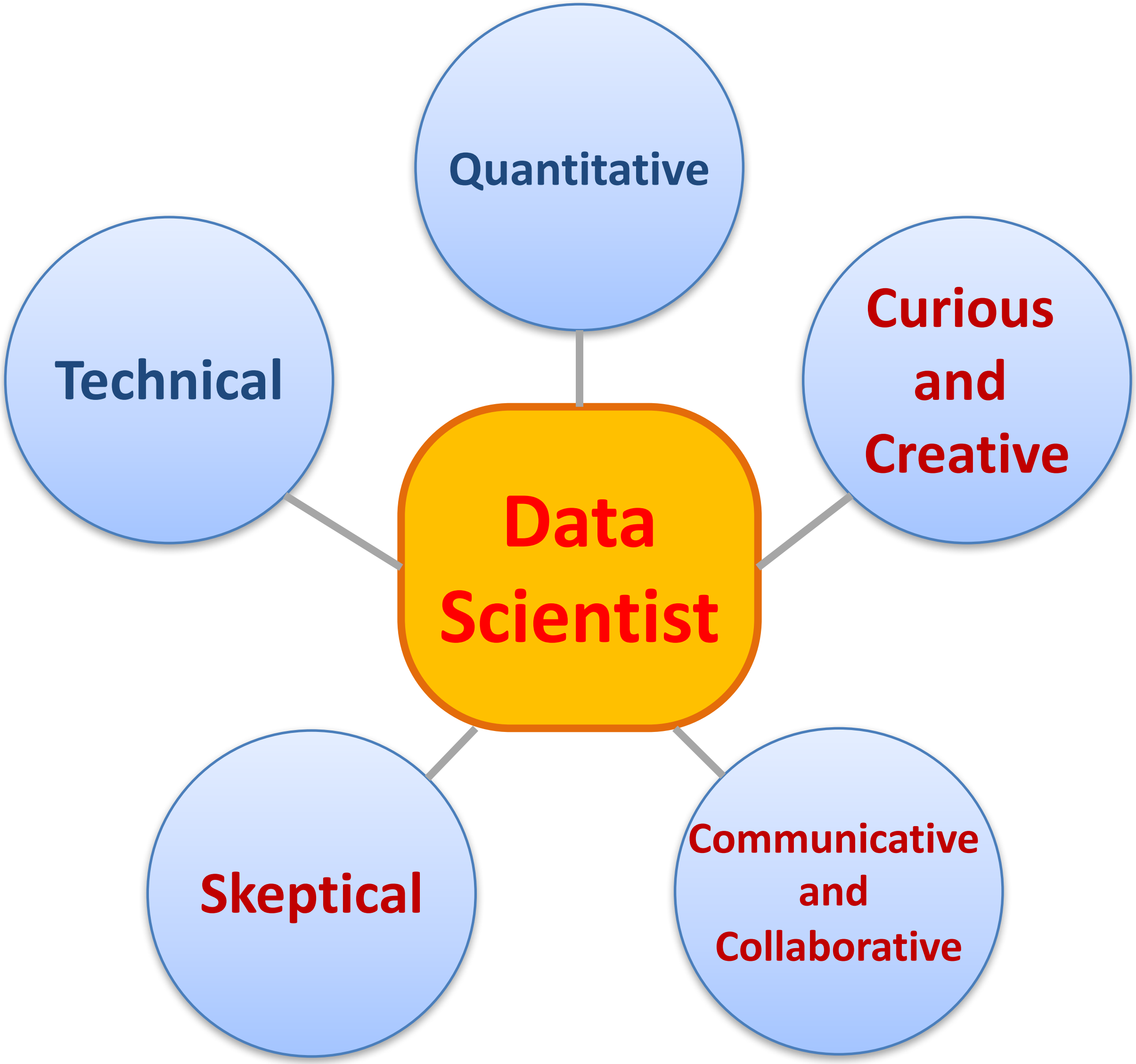
Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

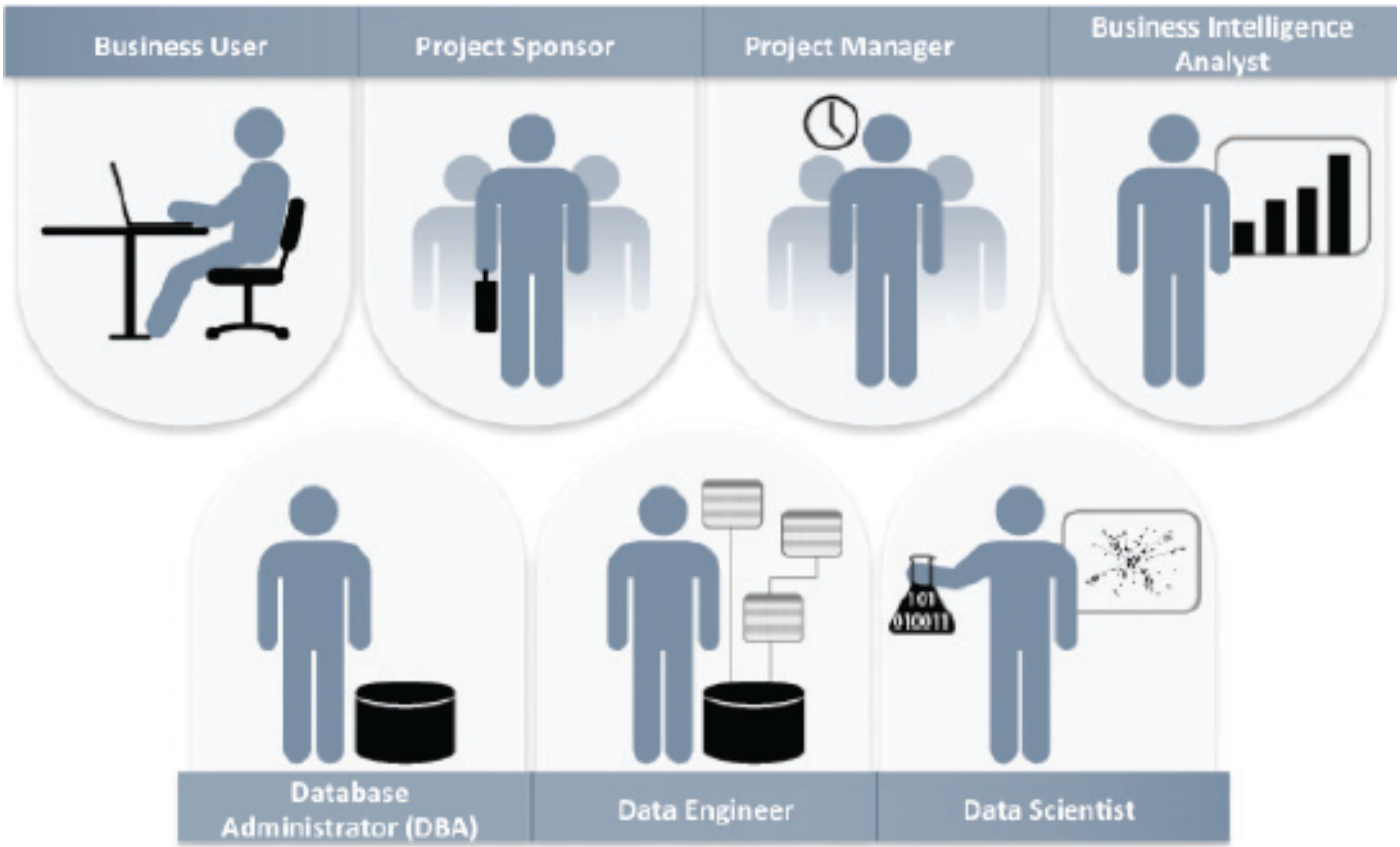
Source: Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*

Data Scientist Profile



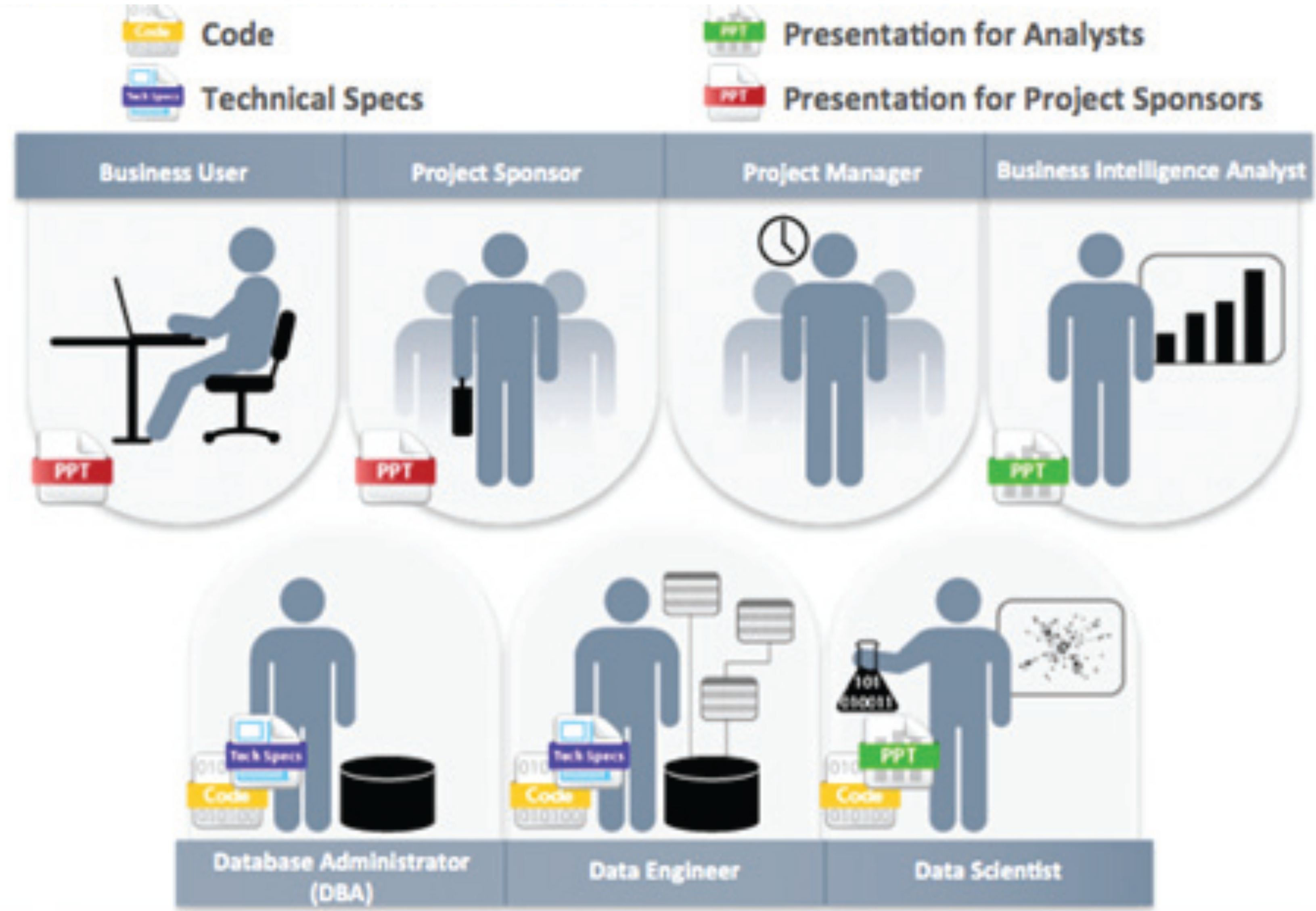
Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Key Roles for a Successful Analytics Project



Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Key Outputs from a Successful Analytics Project



Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Data Science vs.
Big Data vs. Data Analytics

Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Data Science vs. Big Data vs. Data Analytics

WHAT ARE THEY?

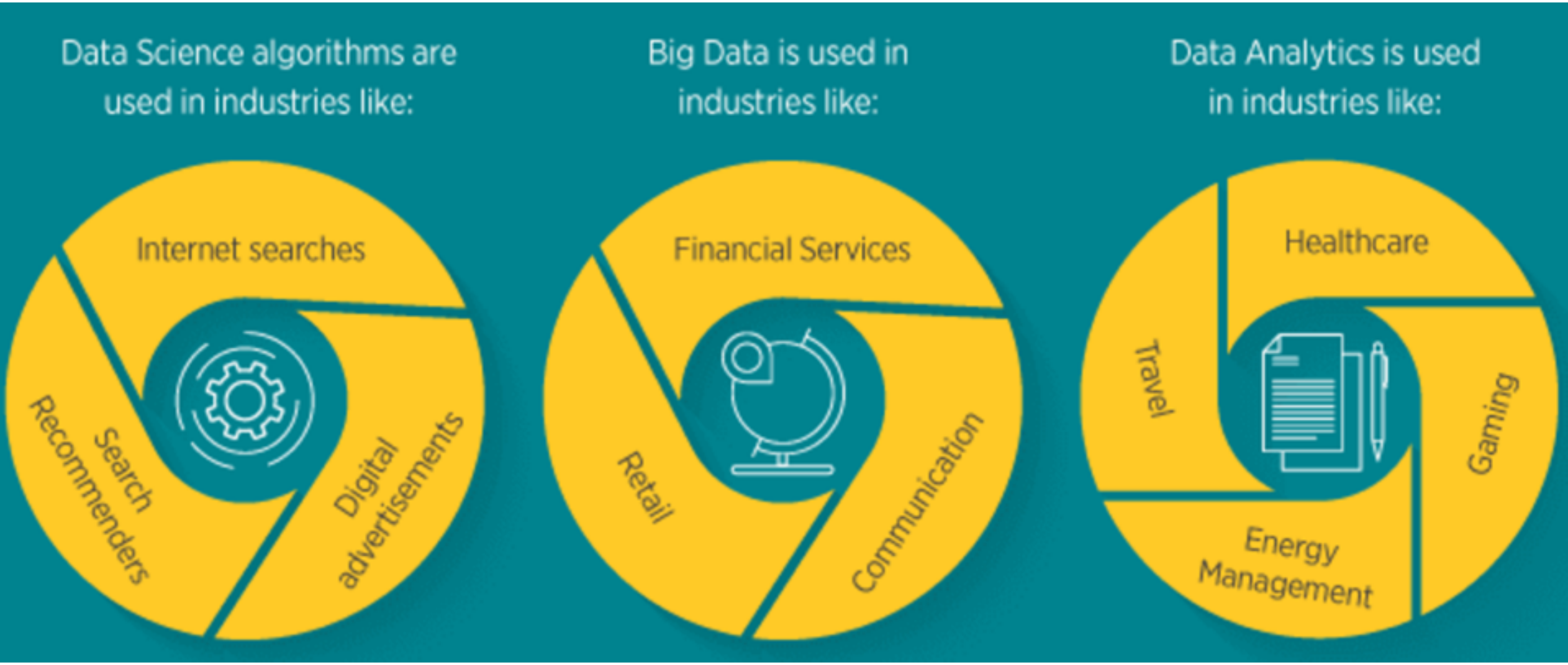
Data Science is a field that comprises of everything that related to data cleansing, preparation, and analysis.

Big Data is something that can be used to analyze insights which can lead to better decision and strategic business moves.

Data Analytics Involves automating insights into a certain dataset as well as supposes the usage of queries and data aggregation procedures.

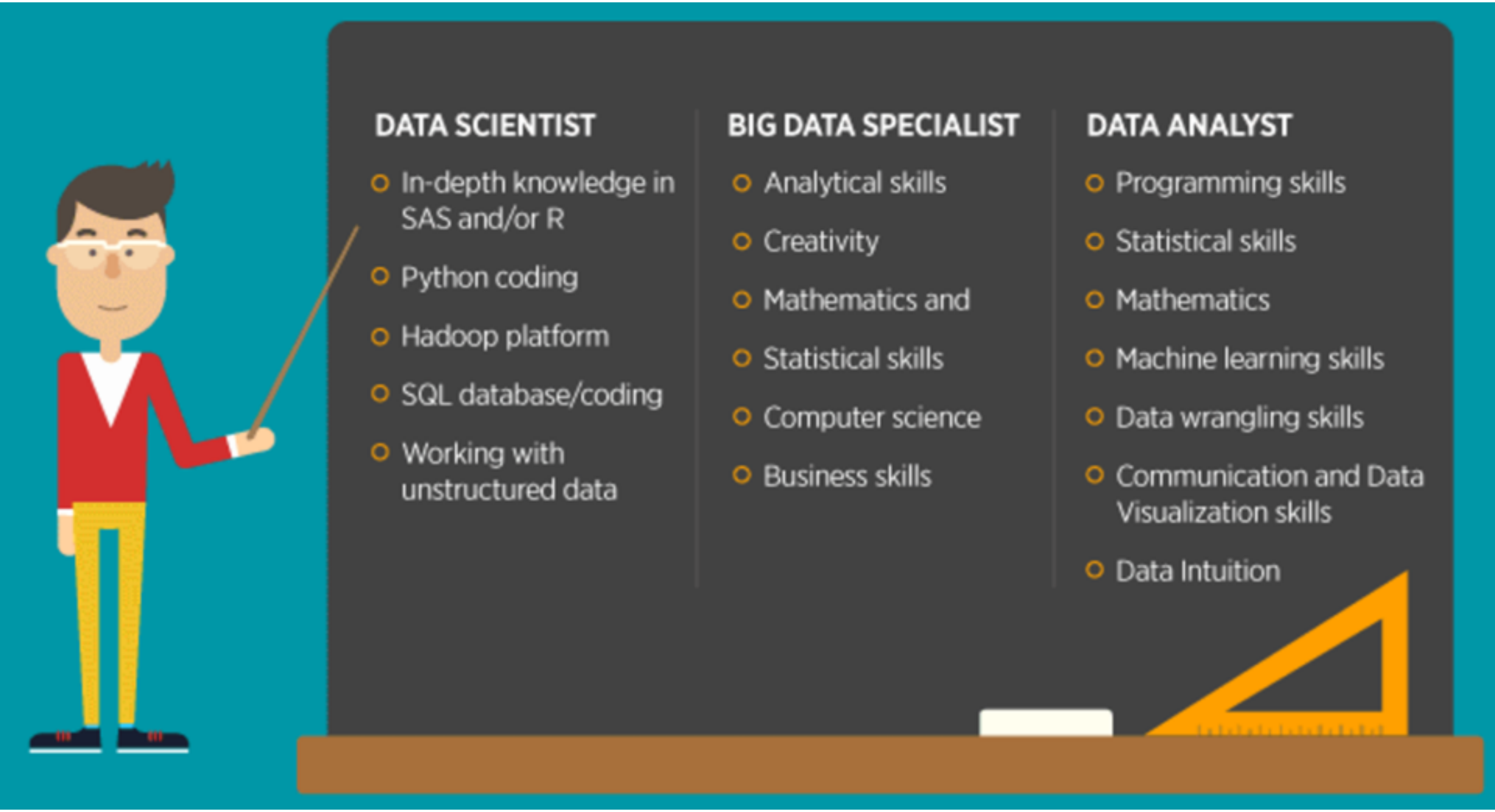
Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

What are they used?

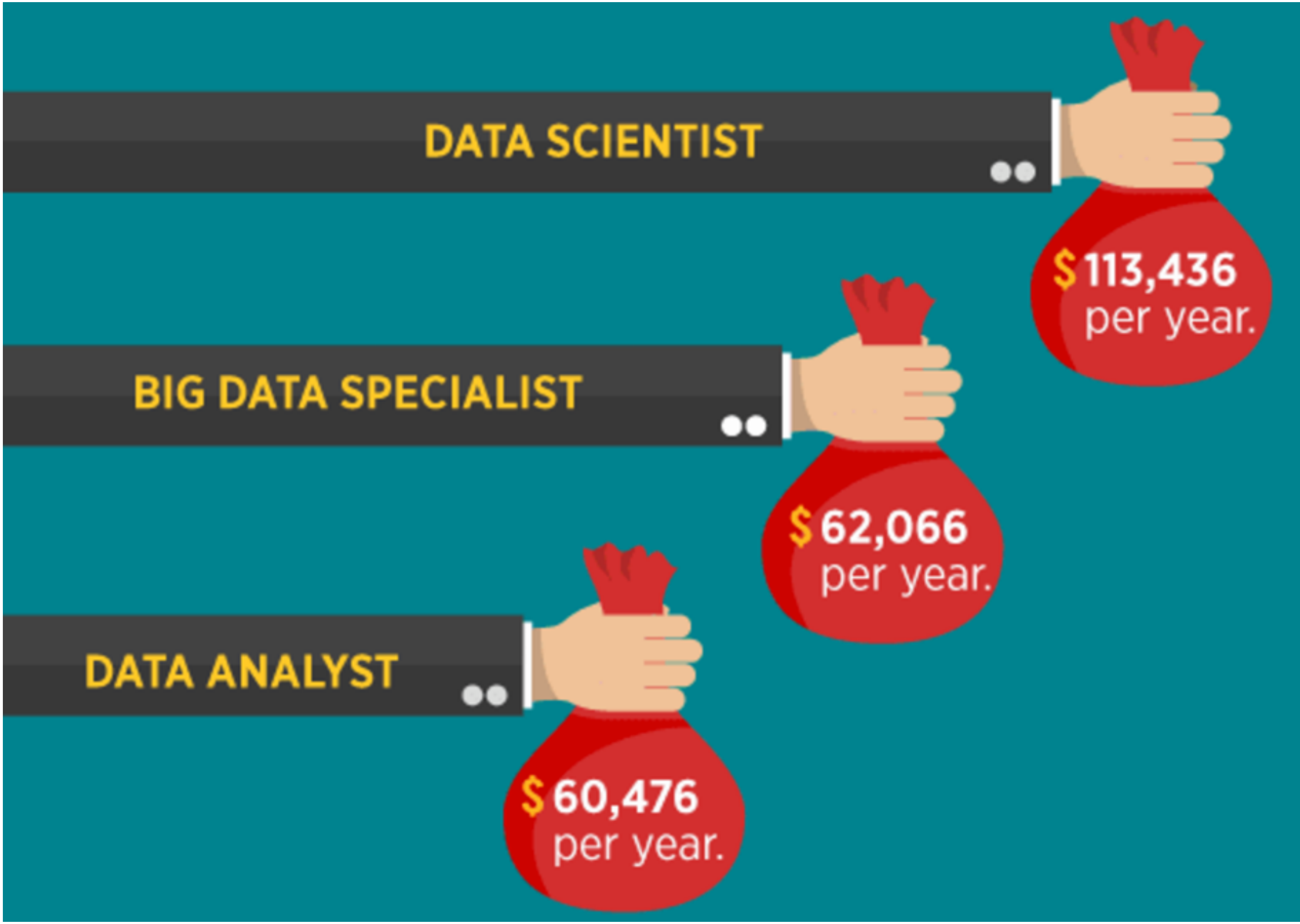


Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Data Science
What are the Skills Required?



Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article



Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

FinTech

FinTech

Financial Technology
FinTech

"providing
financial services
by making use of
software and
modern technology"

Source: https://www.fintechweekly.com/fintech-definition

Financial Revolution with Fintech

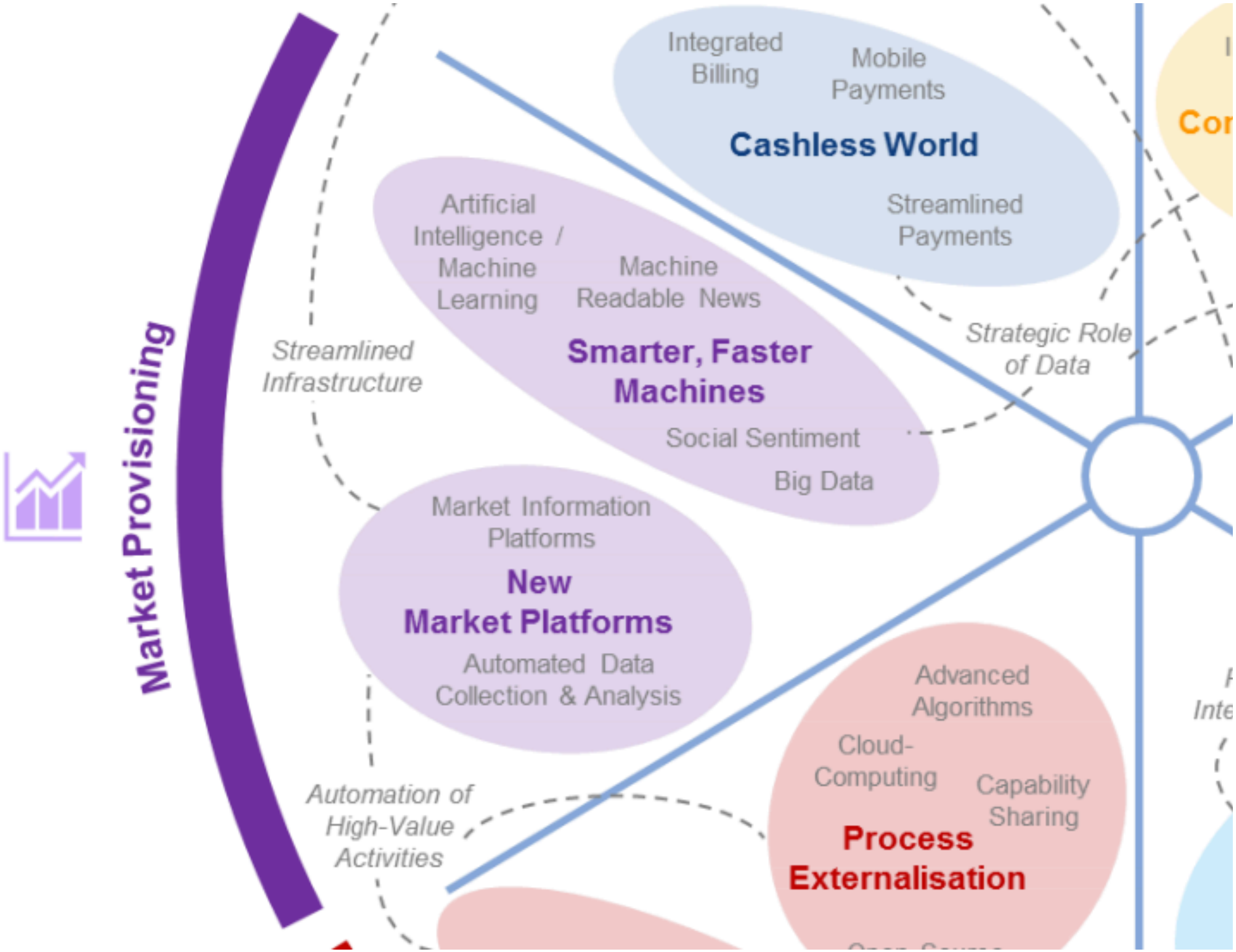
Source: http://www.hedgethink.com/fintech/european-fintech-top-100/

FinTech: Financial Services Innovation

Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

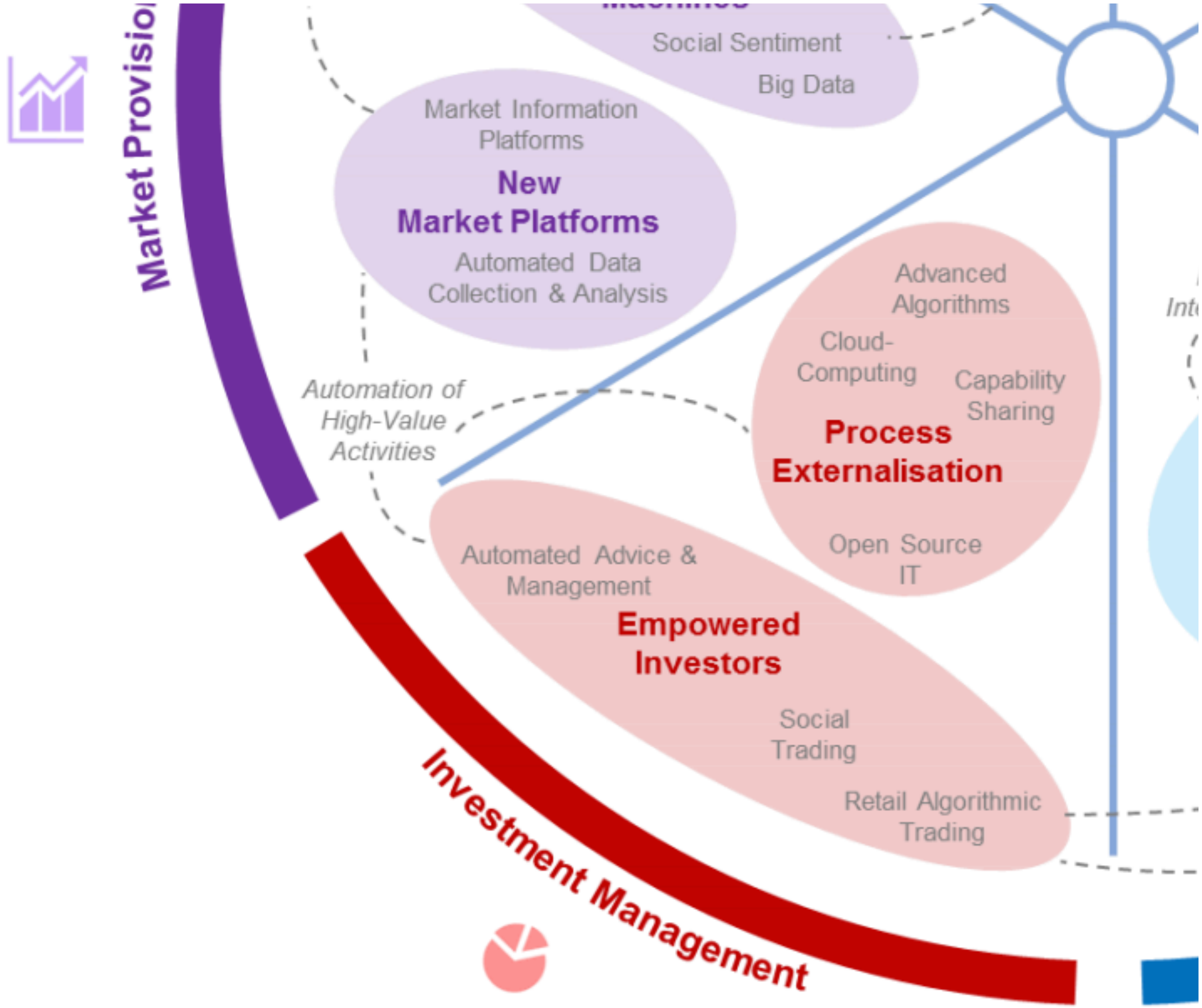
FinTech:
Investment Management Market Provisioning

FinTech: Market Provisioning

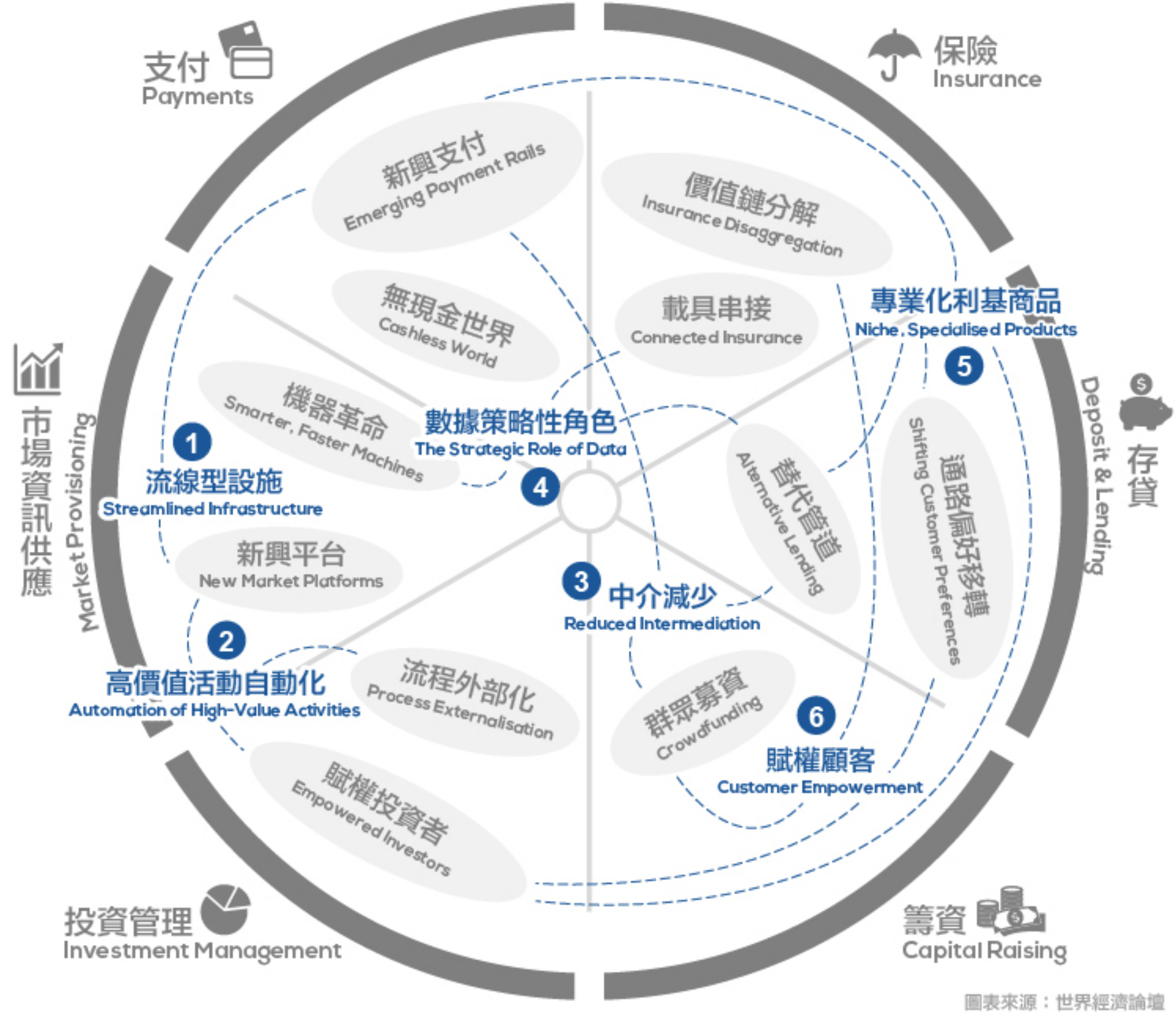


Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

FinTech: Investment Management



Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf



Source: https://www.stockfeel.com.tw/2015年世界經濟論壇－未來的金融服務/

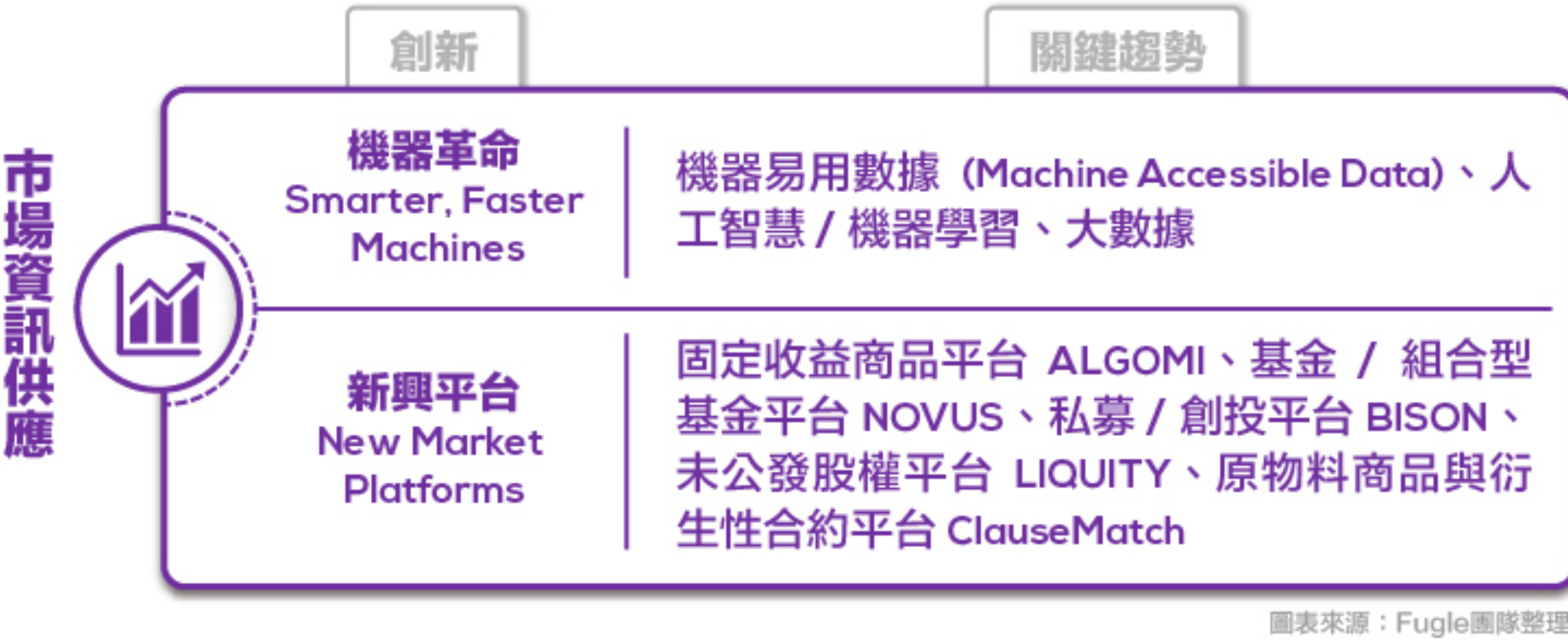
FinTech

功能	創新項目
支付 Payments	無現金世界 (Cashless World) 新興支付 (Emerging Payment Rails)
保險 Insurance	價值鏈裂解 (Insurance Disaggregation) 保險串接裝置 (Connected Insurance)
存貸 Deposit & Lending	替代管道 (Alternative Lending) 通路偏好移轉 (Shifting Customer Preferences)
籌資 Capital Raising	群眾募資 (Crowdfunding)
投資管理 Investment Management	賦權投資者 (Empowered Investors) 流程外部化 (Process Externalisation)
市場資訊供應 Market Provisioning	機器革命 (Smarter, Faster Machines) 新興平台 (New Market Platforms)

圖表來源：Fugle團隊整理

Source: https://www.stockfeel.com.tw/2015年世界經濟論壇－未來的金融服務/

FinTech: Market Provisioning Smarter, Faster Machines



Source: https://www.stockfeel.com.tw/2015年世界經濟論壇－未來的金融服務/

FinTech: Investment Management

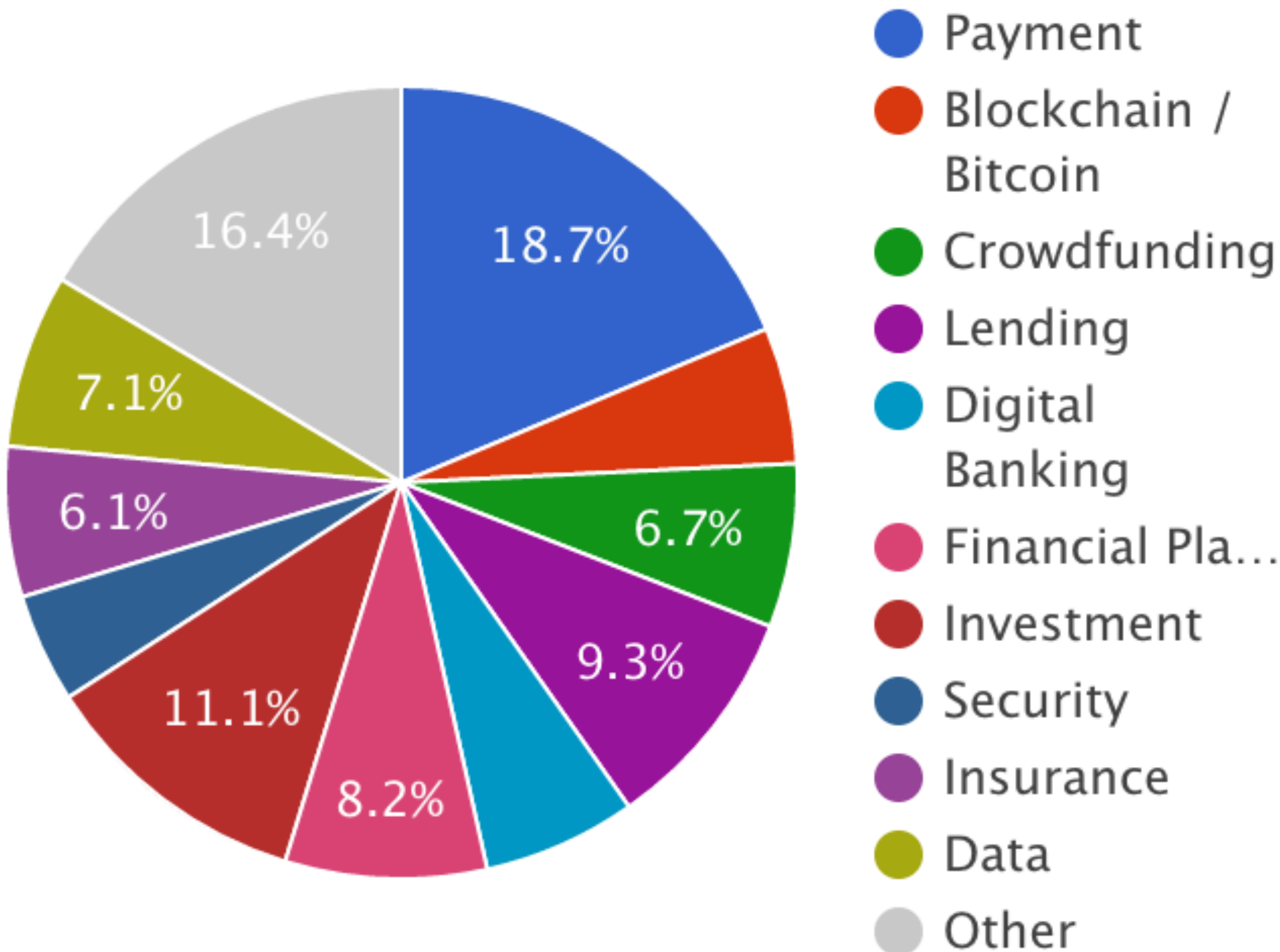
Source: https://www.stockfeel.com.tw/2015年世界經濟論壇一未來的金融服務/

FinTech for Financial Services

- Retail Banking
- Lending and Financing
- Payments and Transfers
- Wealth and Asset Management
- Markets and Exchanges
- Insurance
- Blockchain Transactions

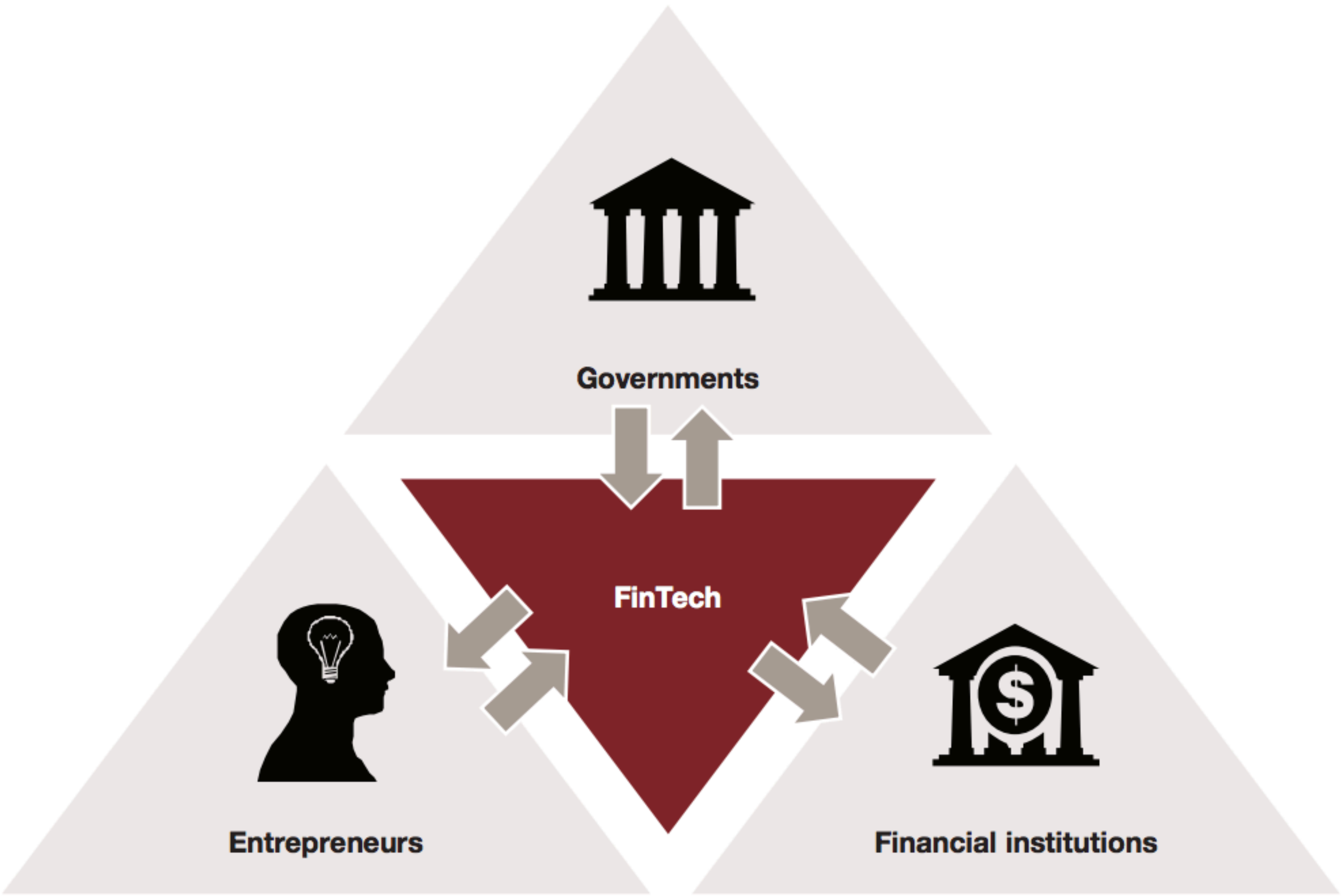
Source: https://fintechweekly.com/fintech-companies

Fintech Companies



Source: https://fintechweekly.com/fintech-companies

Major Participants in the FinTech Ecosystem



Source: http://www.strategyand.pwc.com/media/file/Developing-a-FinTech-ecosystem-in-the-GCC.pdf

FinTech Ecosystem Development Framework

1. Business environment/access to markets

- Cost advantages
- Labor availability and know-how
- Degree of clustering & integration
- Infrastructure quality & access

2. Government/regulatory support

- Policy setting
- Land ownership & development
- Ease of doing business
- Taxes & work permits

3. Access to capital

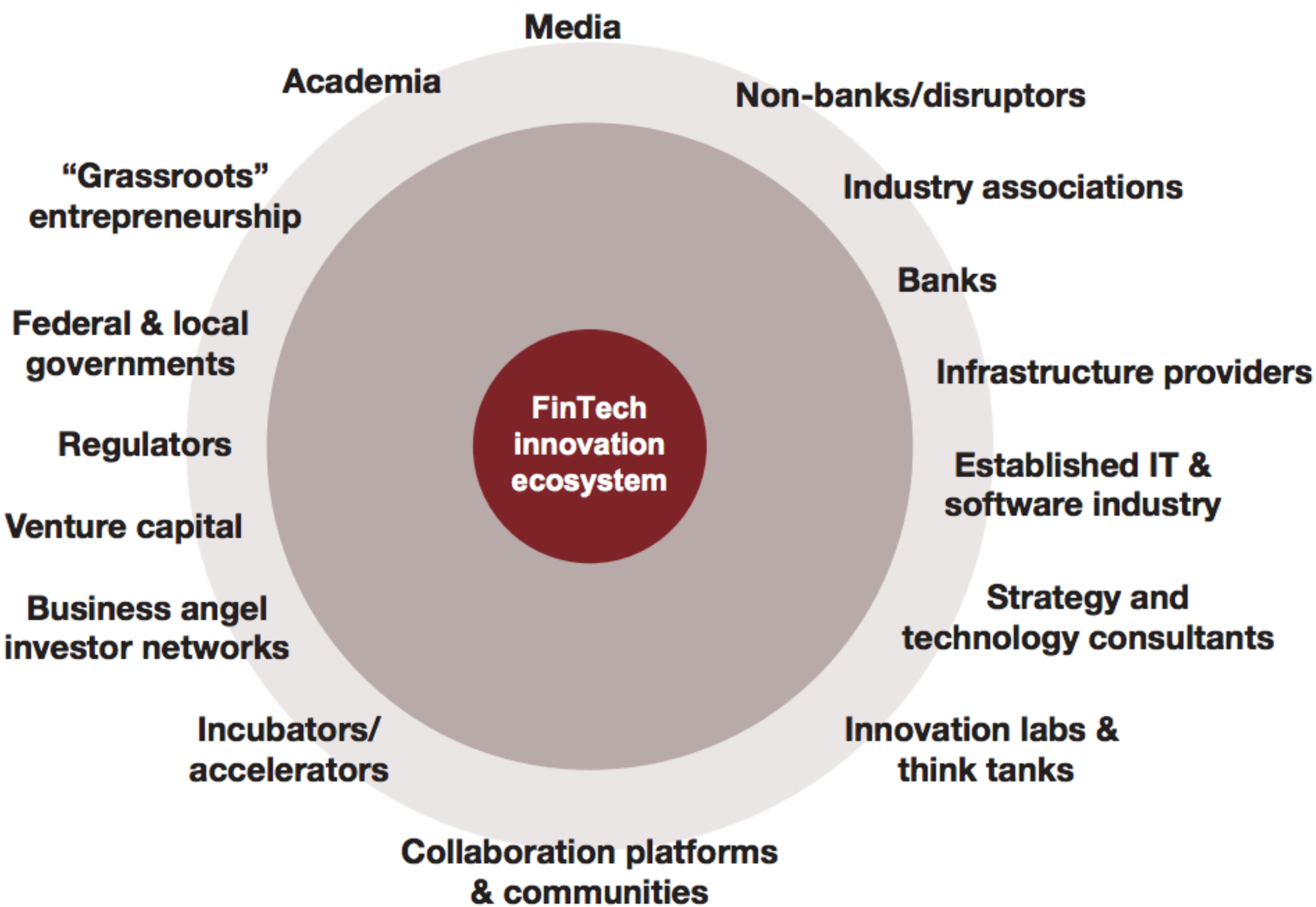
- Governmental funding
- Bank funding
- PE/VC funding
- Incubators / accelerators

4. Financial expertise

- Funding strategies
- "Know your customer" & regulatory requirements
- Deals structuring
- Due diligence

Source: http://www.strategyand.pwc.com/media/file/Developing-a-FinTech-ecosystem-in-the-GCC.pdf

The FinTech Innovation Ecosystem



Source: http://www.strategyand.pwc.com/media/file/Developing-a-FinTech-ecosystem-in-the-GCC.pdf

The U.S. FinTech landscape

Entrepreneurs

- New York is the fastest-growing FinTech ecosystem in the U.S.
- Talent feed from world's biggest financial center
- New York is a lifestyle choice for talented young entrepreneurs

Support structures

- Tax credits for business R&D and patents
- Incubators & accelerators (e.g., Partnership Fund for New York City)

Investment

- U.S. received 83 percent of global FinTech investments in 2013
- The financial services industry globally spent more than US$485 billion on ICT[1] in 2014

Customers

- Business to business: high density of financial services firms seeking support for digitalization
- Business to consumer: widespread mobile & e-commerce usage, low bank client "stickiness"

Source: http://www.strategyand.pwc.com/media/file/Developing-a-FinTech-ecosystem-in-the-GCC.pdf

Fintech Startups

USA Fintech Ecosystem

+ ADD STARTUP

Startups (74) | **News** (11)

Crowdfunding

	STARTUP	DESCRIPTION	LINKS	STATUS	MONEY RAISED
C	Crowdera				-
	AngelList	Community of start-ups & investors who make fundraising efficient	cb	alive	$24.1M
Circle Up	CircleUp	Online marketplace that links accredited investors with consumer product and retail companies.	cb	alive	$53M
ES	Early Shres	Real estate crowdfunding platform		alive	
GO	Indiegogo	Crowdfunding platform	cb	alive	$56.5M
	Kickfurther	Businesses finance inventory. Backers earn returns.		alive	
K	Kickstarter	Crowdfunding platform	cb	alive	$10M
	Local lift	Brings crowdfunding to your local area.	cb	alive	$160k
	Onevest (Rock The Post)	Equity crowd-funding platform	cb	alive	$2M
	Quirky	Community-led invention platform	cb	alive	

FS

Fintech Startups WorldWide

TEB FINTECH FUTURE FOUR

Subscribe to our Newsletter
fstartupsco
fintech@startups.watch

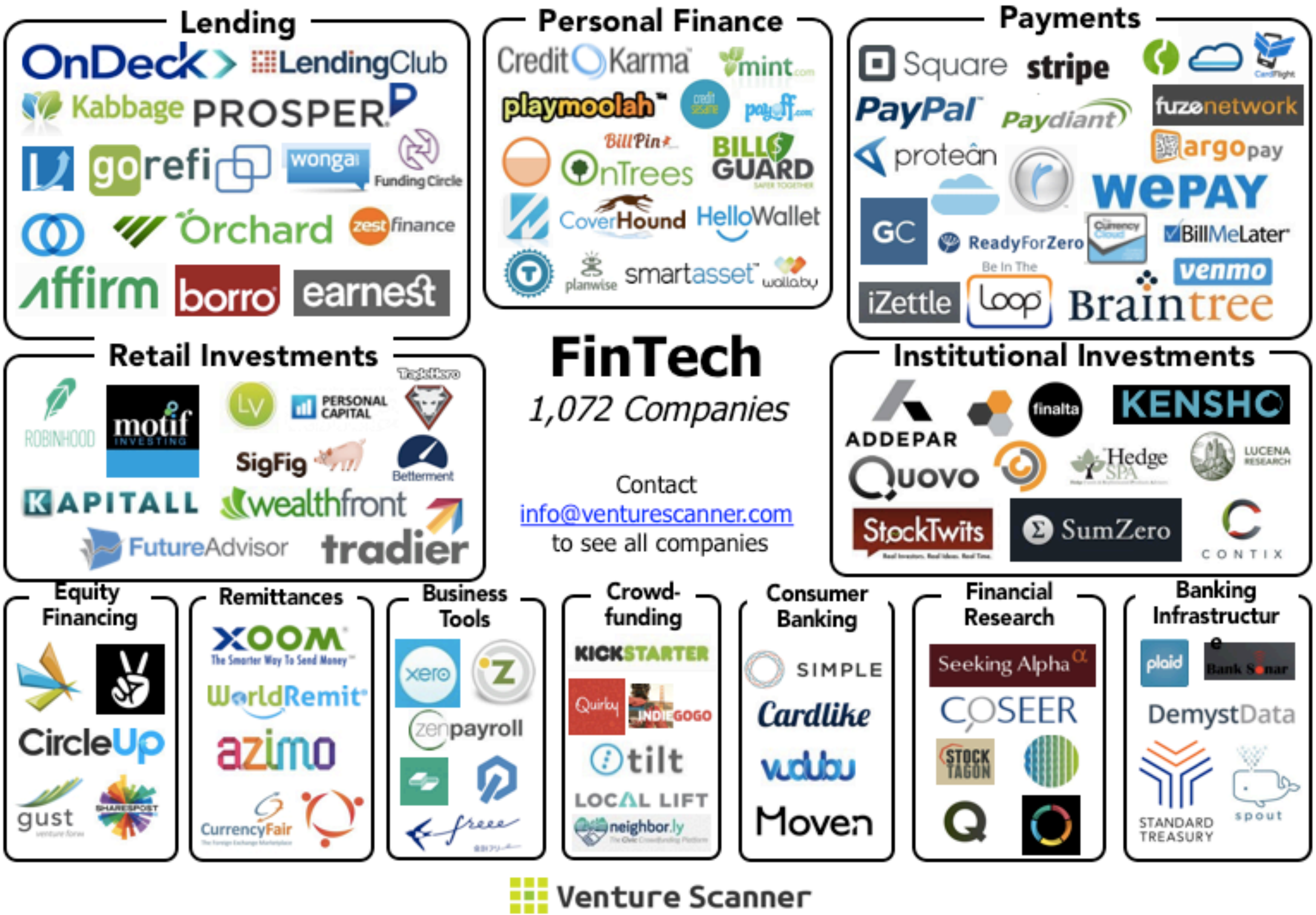
Source: https://fintechstartups.co

Financial Technology (Fintech) Categories

1. Banking Infrastructure
2. Business Lending
3. Consumer and Commercial Banking
4. Consumer Lending
5. Consumer Payments
6. Crowdfunding
7. Equity Financing
8. Financial Research and Data
9. Financial Transaction Security
10. Institutional Investing
11. International Money Transfer
12. Payments Backend and Infrastructure
13. Personal Finance
14. Point of Sale Payments
15. Retail Investing
16. Small and Medium Business Tools

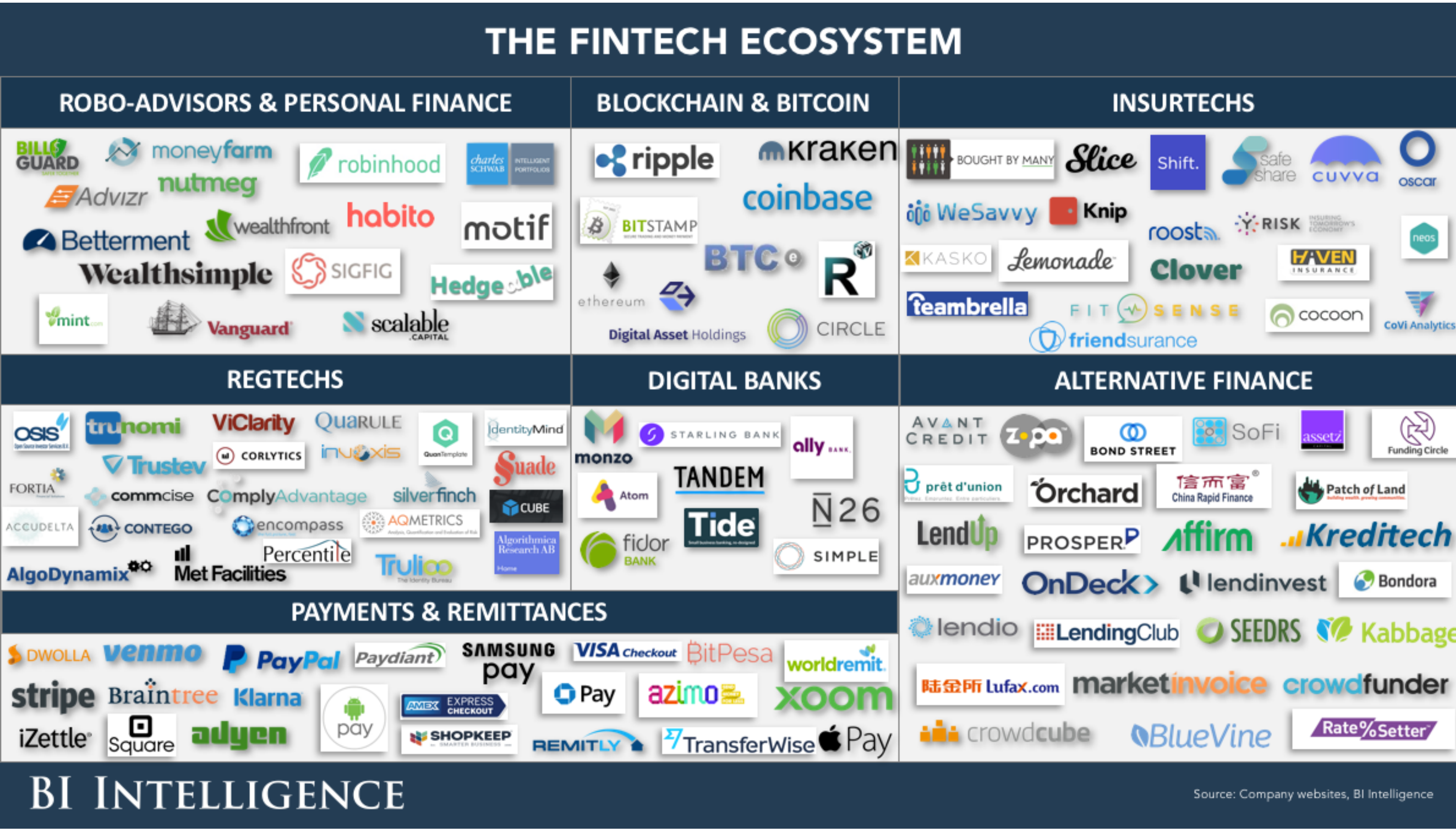
Source: http://www.venturescanner.com/financial-technology

FinTech Ecosystem (April 2015)



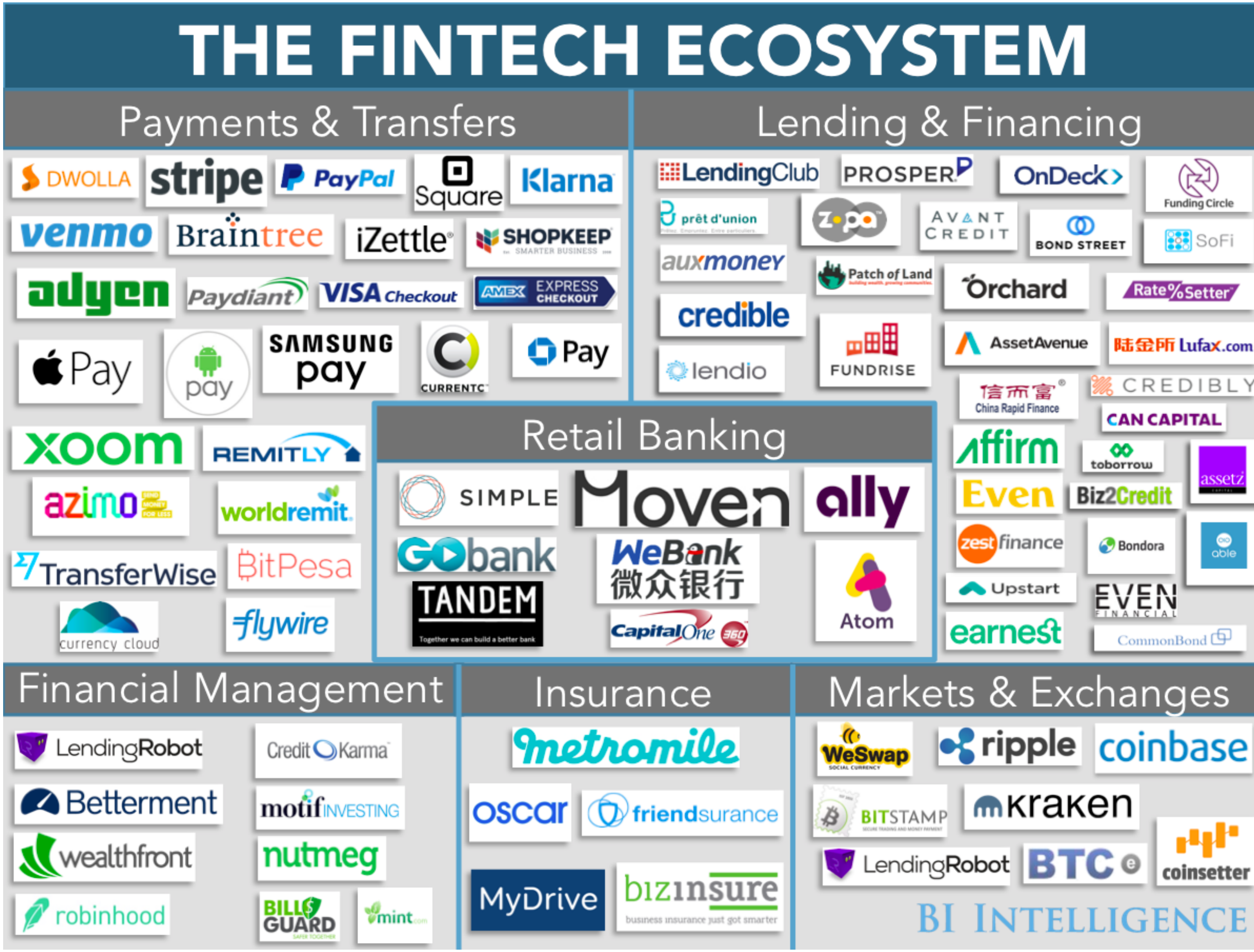
Source: https://venturescannerinsights.wordpress.com/2015/04/07/fintech-ecosystem-update-april-2015/

Financial Technology (FinTech)

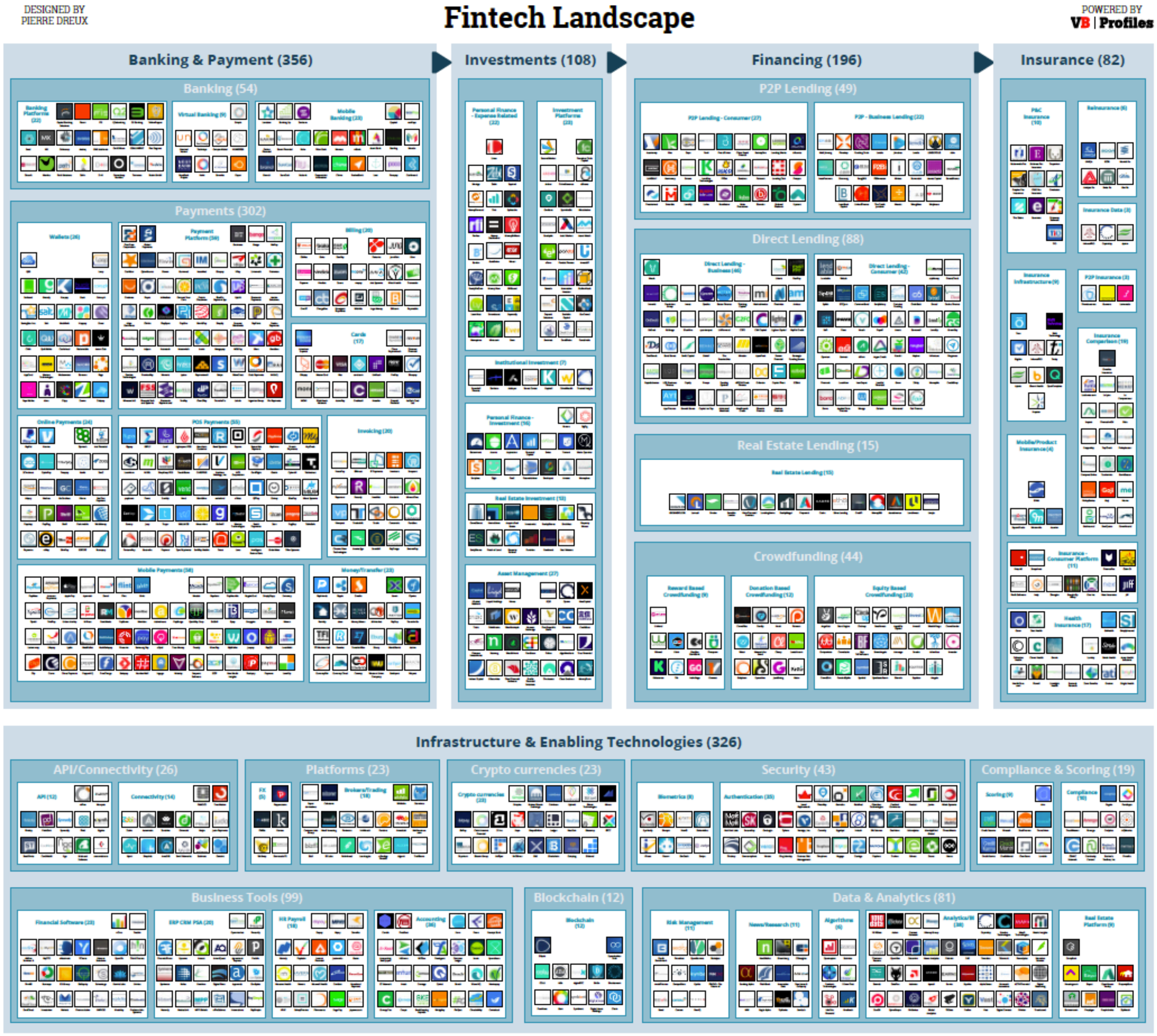


Source: http://www.businessinsider.com/the-fintech-ecosystem-explained-measuring-the-effects-of-technology-on-the-entire-financial-services-industry-2015-12

Financial Technology (FinTech)



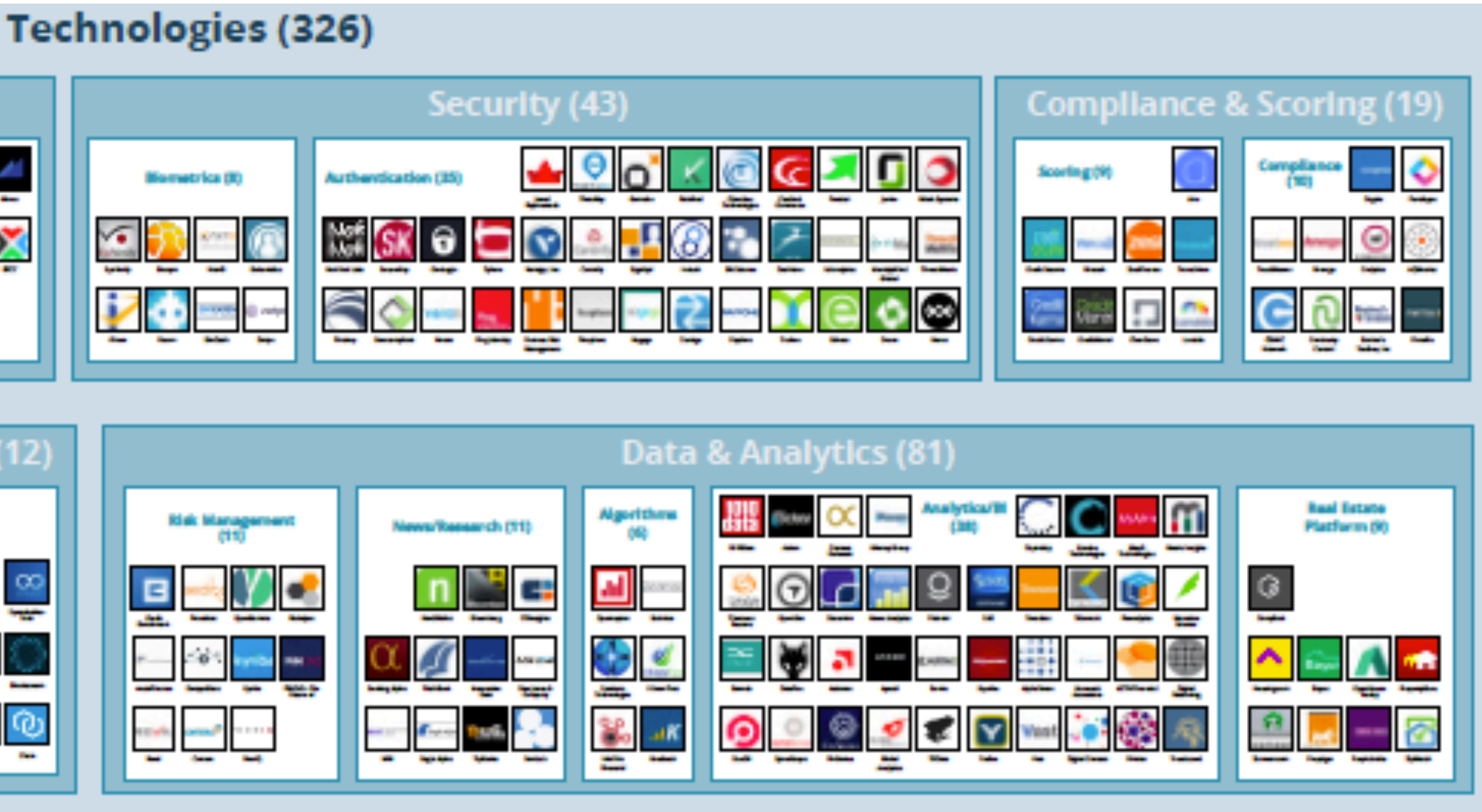
Source: http://www.businessinsider.com/fintech-ecosystem-financial-technology-research-plus-business-opportunities-2016-2



Source: http://www.forbes.com/sites/jeanbaptiste/2016/09/28/the-global-fintech-landscape-reaches-over-1000-companies-105b-in-funding-867b-in-value-report

FinTech Landscape Enabling Technologies

Data & Analytics



Source: https://www.vbprofiles.com/l/fintech

FINTECH | LANDSCAPE

everisDigital

Digital & Mobile payments

Braintree, Bitnet, CLINKLE, flint, JUMIO, CardFlight, adyen, TransferWise, KOPO KOPO, KASH, soldsie, coin, Zumigo, 2CHECKOUT.COM, affirm, DWOLLA, gumroad, cashstar, InstaMed, iZettle, LevelUp, mineraltree, PayNearMe, peerTransfer, POWA, SavingStar, Square, stripe, Tipalti, TRANSACTIS, sumup, WePAY, tyfone, worldremit, paytouch, ZenBanx, monitise, Yaap, Airbrite

Capital Markets & Investing

Betterment, EQUITYZEN, estimize, iex, sliced, eShares, INSIKT, INVESDOR, LexShares, nextcapital, REALCROWD, Robinhood, tradier, 6fusion, ADDEPAR, dough, KANTOX, MARKITX, Mercatus, motif INVESTING, NOVUS, Quantopian, secondmarket, wealthfront, Xambala

Banking & Corporate Finance

Continuity, BlueTarp, PLAID, coupa, ALKAMI, CHROMERIVER, orderwithme, ncino, APTTUS, hostanalytics, OnDeck, MarketInvoice, Bill.com, JUSTWORKS, nettit, FASTPAY, Avalara, Intacct, SWITCHFLY

Big Data & Analytics

CONTIX, KENSHO, OPENGOV, feedzai, DRAFT, METRIC INSIGHTS, Eagle Alpha, cardlytics, Swipely, 1010data, DemystData, zestfinance, xignite, CONFLUENCE, tidemark, wikinvest, Credit Karma

Financial platforms

Biz2Credit, CAN CAPITAL, Upstart, LOAN LOGISTICS, Kabbage, tilt, OCX, borro, SoFi, UCX, payoff, C2FO, Even, OPORTUN, ApplePie CAPITAL, wonga, LendUp

Crowdfunding & peer-to-peer lending

eMoneyPool, Dealstruck, fundera, Funding Circle, crowdfunder, KICKSTARTER, OurCrowd, PROSPER, AVANT, CommonBond, indiegogo, LendingHome, LendingClub, MytripleA

Personal financial management

activehours, CoPatient, MileIQ, REMITLY, Expensify, FutureAdvisor, PERSONAL CAPITAL, smartasset, DRIVERUP

Bitcoin & Cryptocurrency

pagobit, AlphaPoint, bitpay, CRYPTOPAY, airBitz, CIRCLE, coinbase, xapo, rippleLabs

Blockchain technology

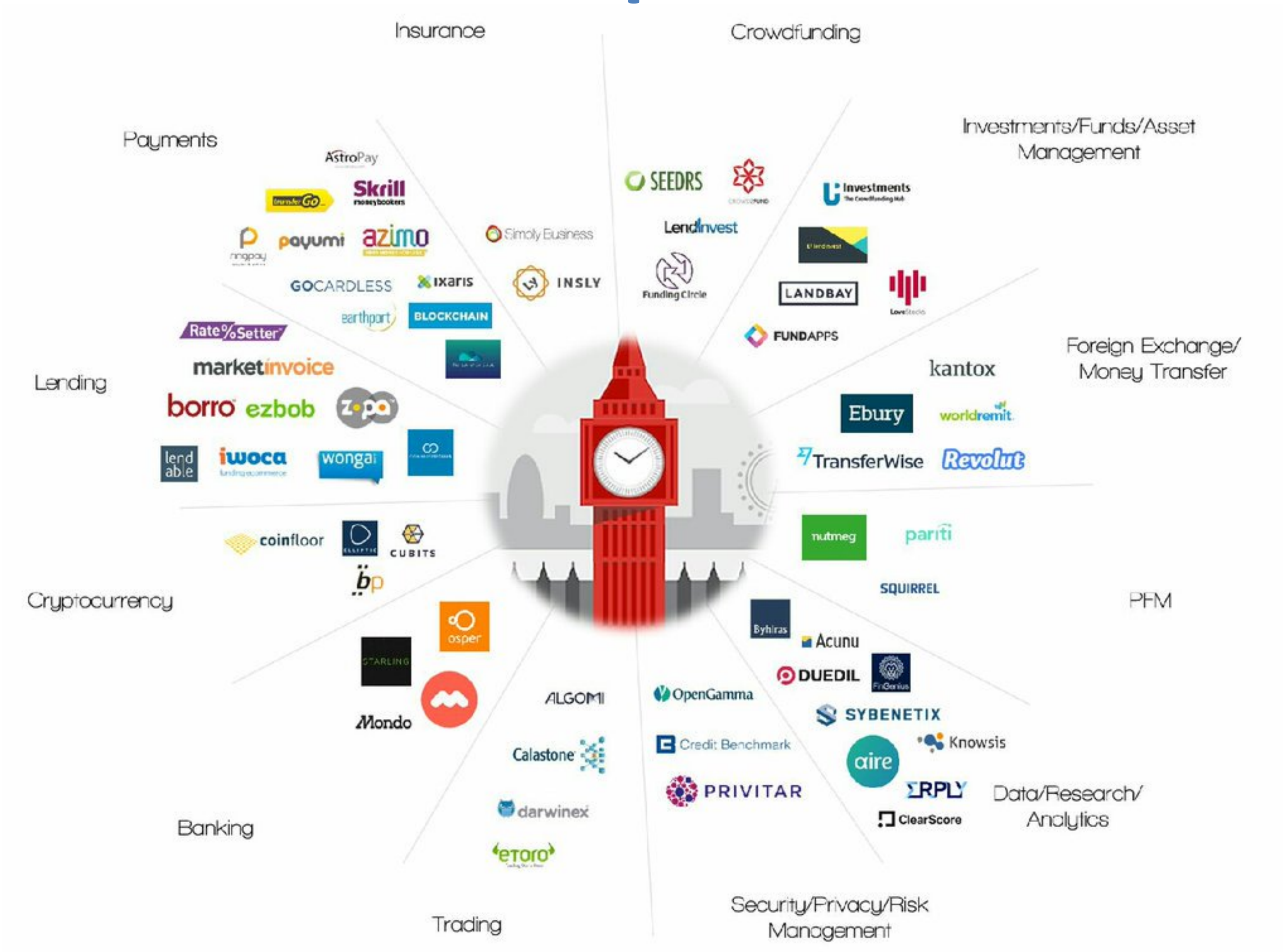
BitGo, Mirror, PEERNOVA, BLOCKCHAIN, coinsetter

 | **Source:** "OnFinance Top 100 private companies " by AlwaysOn & everisDigital database.

Source: https://everisnext.com/2015/06/02/top-9-verticals-within-the-fintech-landscape-for-large-corporations/

FinTech Startups Worldwide

Fintech Startups WorldWide



Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/

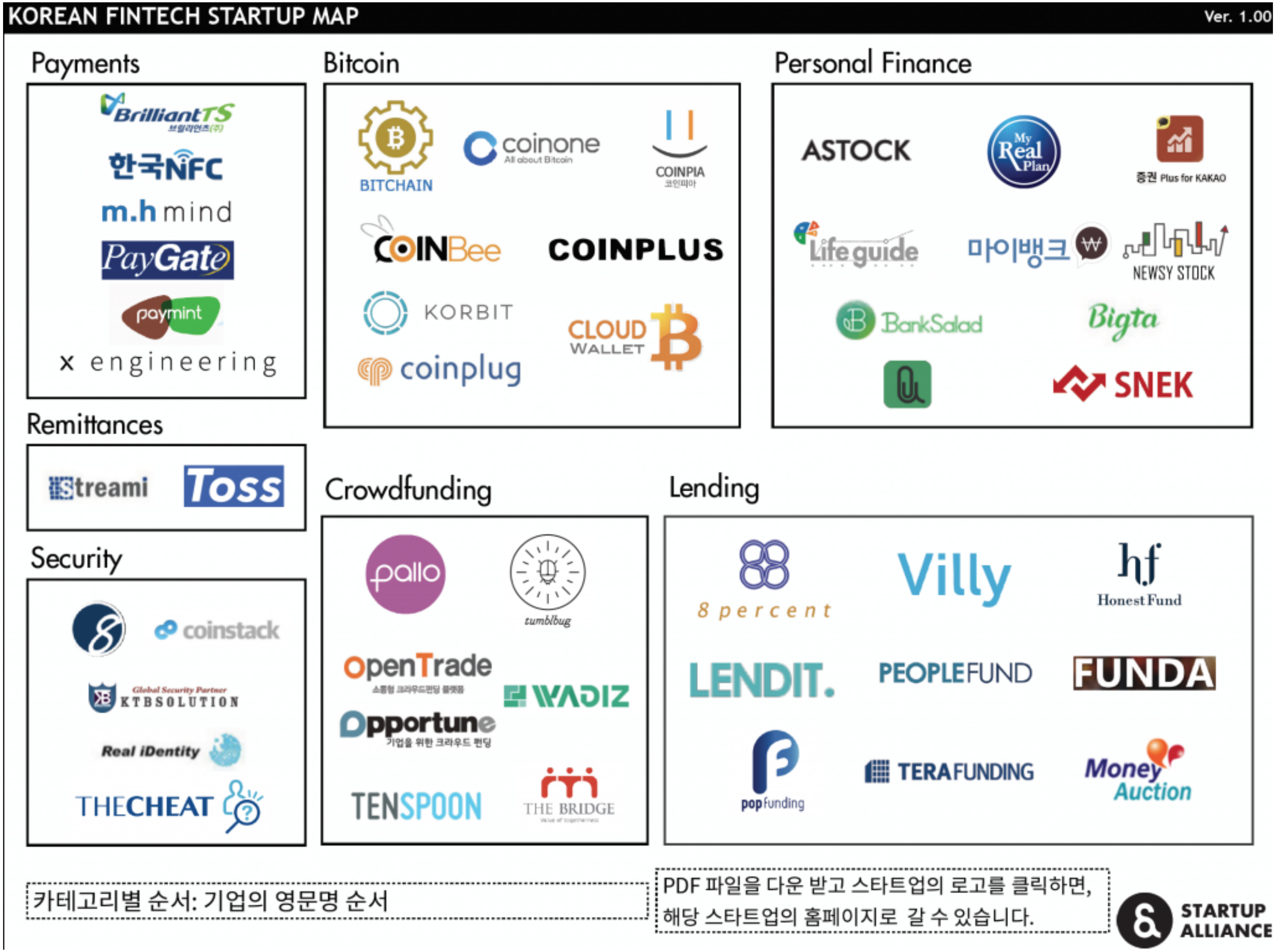
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide

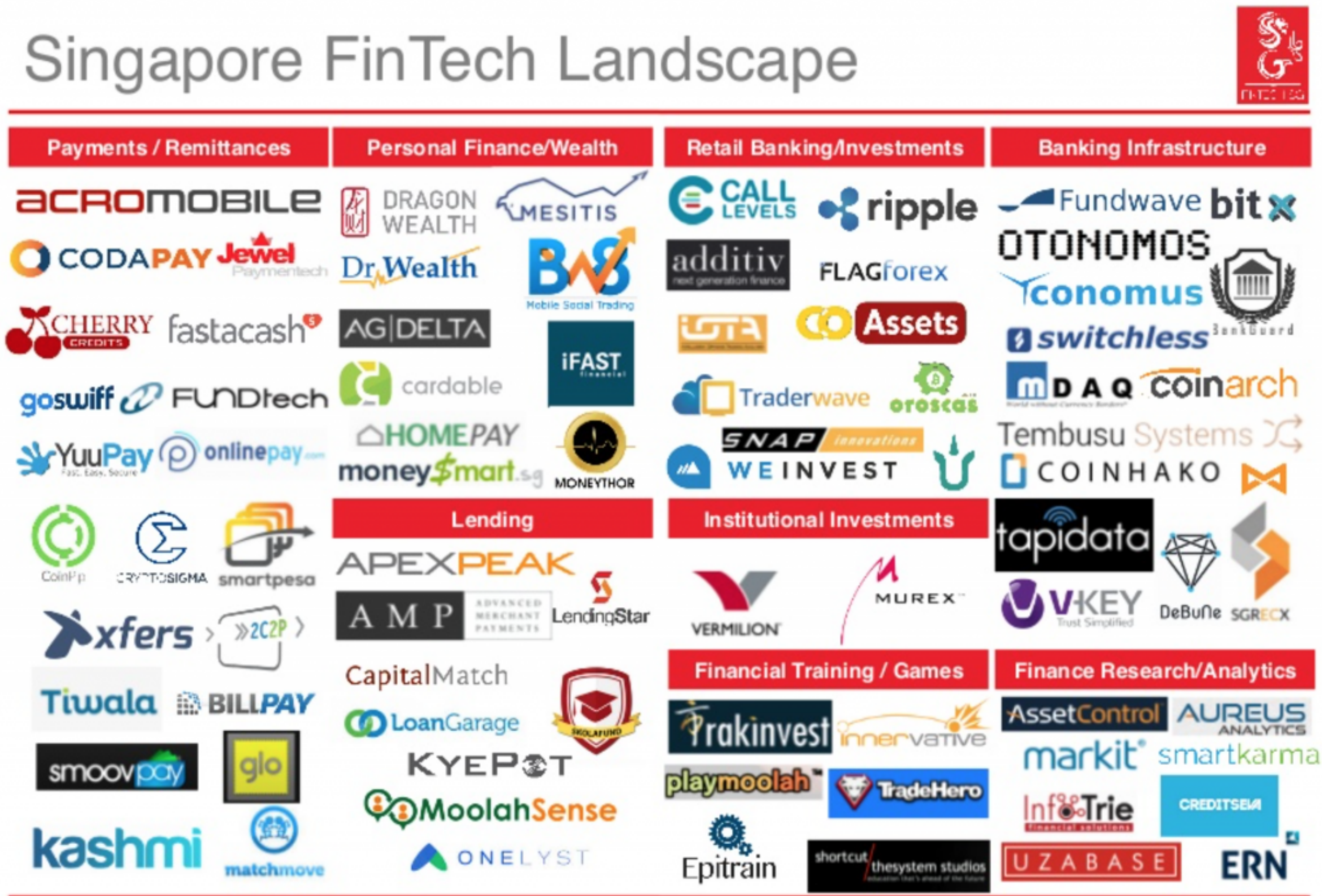
Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide



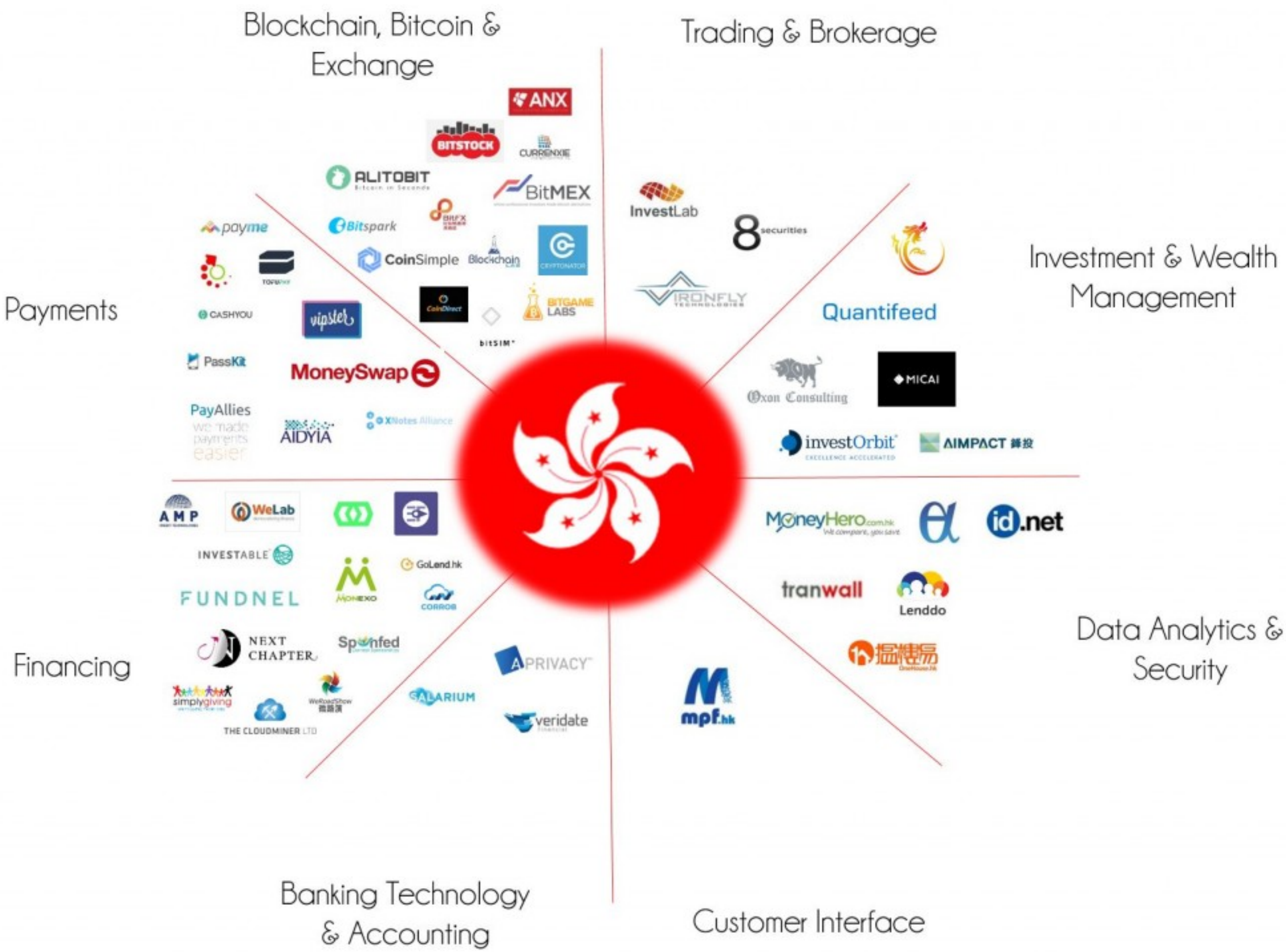
Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide



Source: https://startups.watch/fintech-ecosystems-worldwide/

Fintech Startups WorldWide



Source: https://startups.watch/fintech-ecosystems-worldwide/

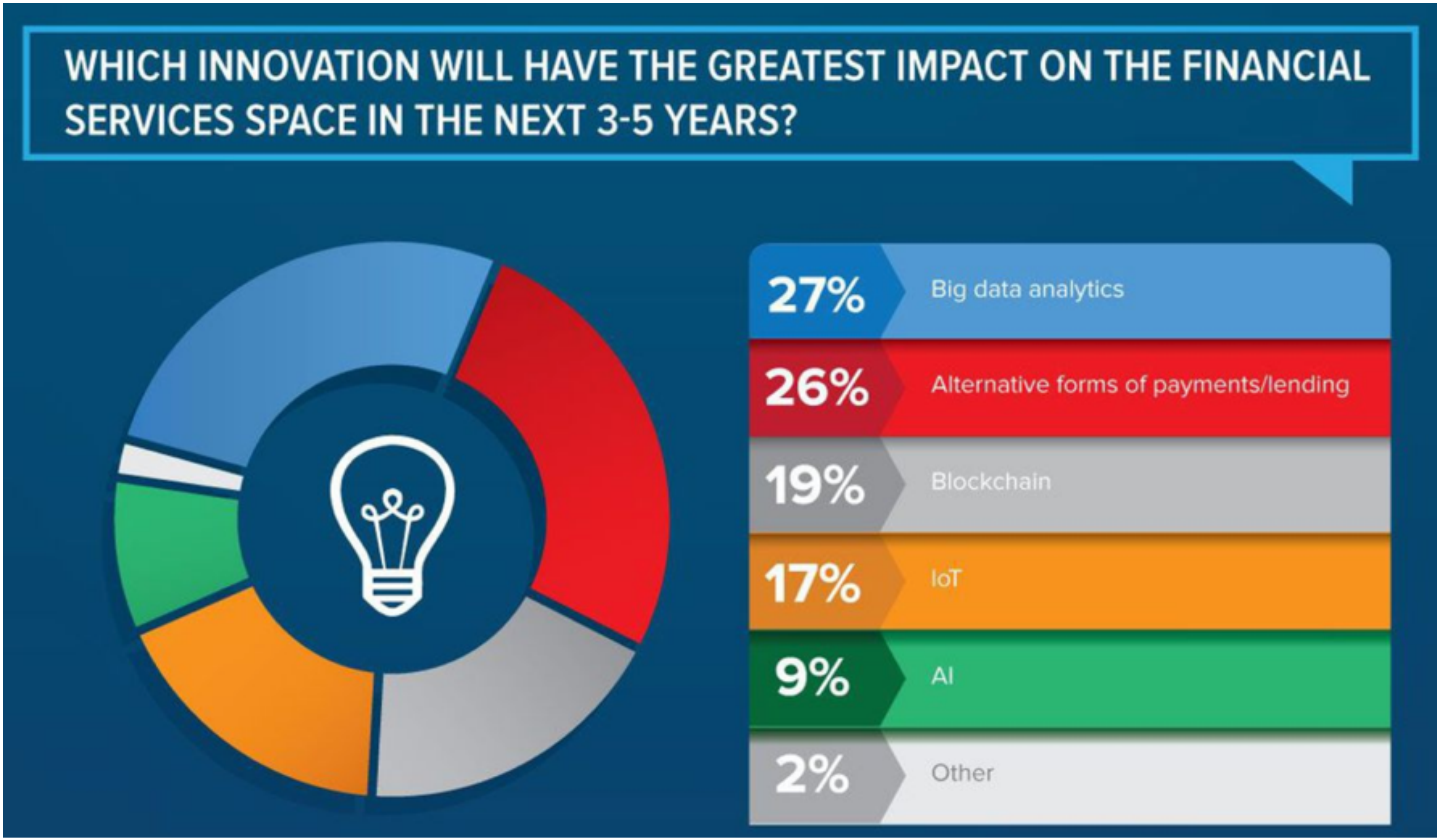
Fintech Startups WorldWide

Source: https://startups.watch/fintech-ecosystems-worldwide/

FinTech Trends

Source: http://mashable.com/2016/01/27/financial-tech-brandspeak'

FinTech
Big Data Analytics



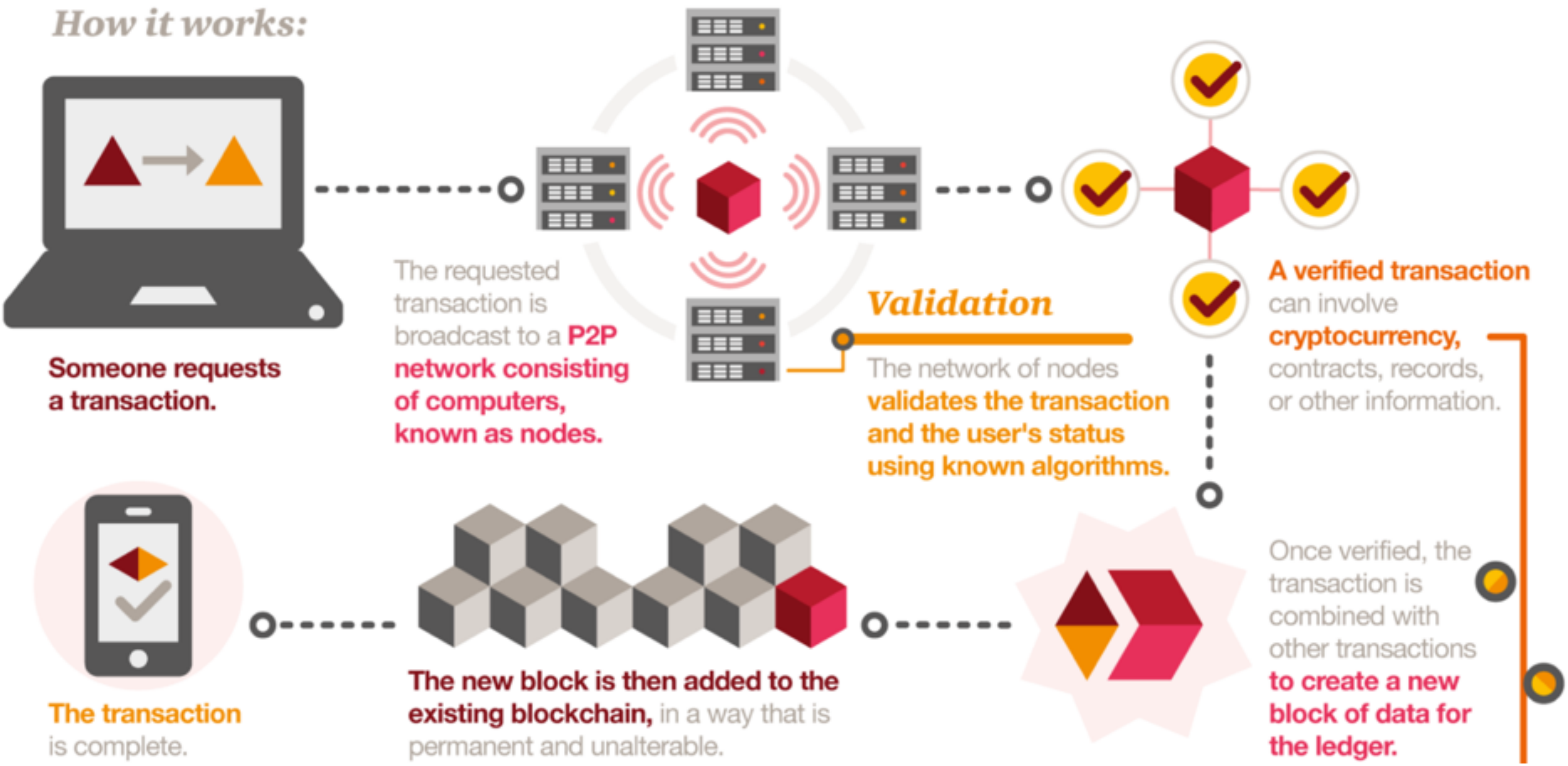
Source: http://mashable.com/2016/01/27/financial-tech-brandspeak'

Blockchain Technology

The blockchain is a
decentralized ledger
of all transactions
across
a peer-to-peer network.

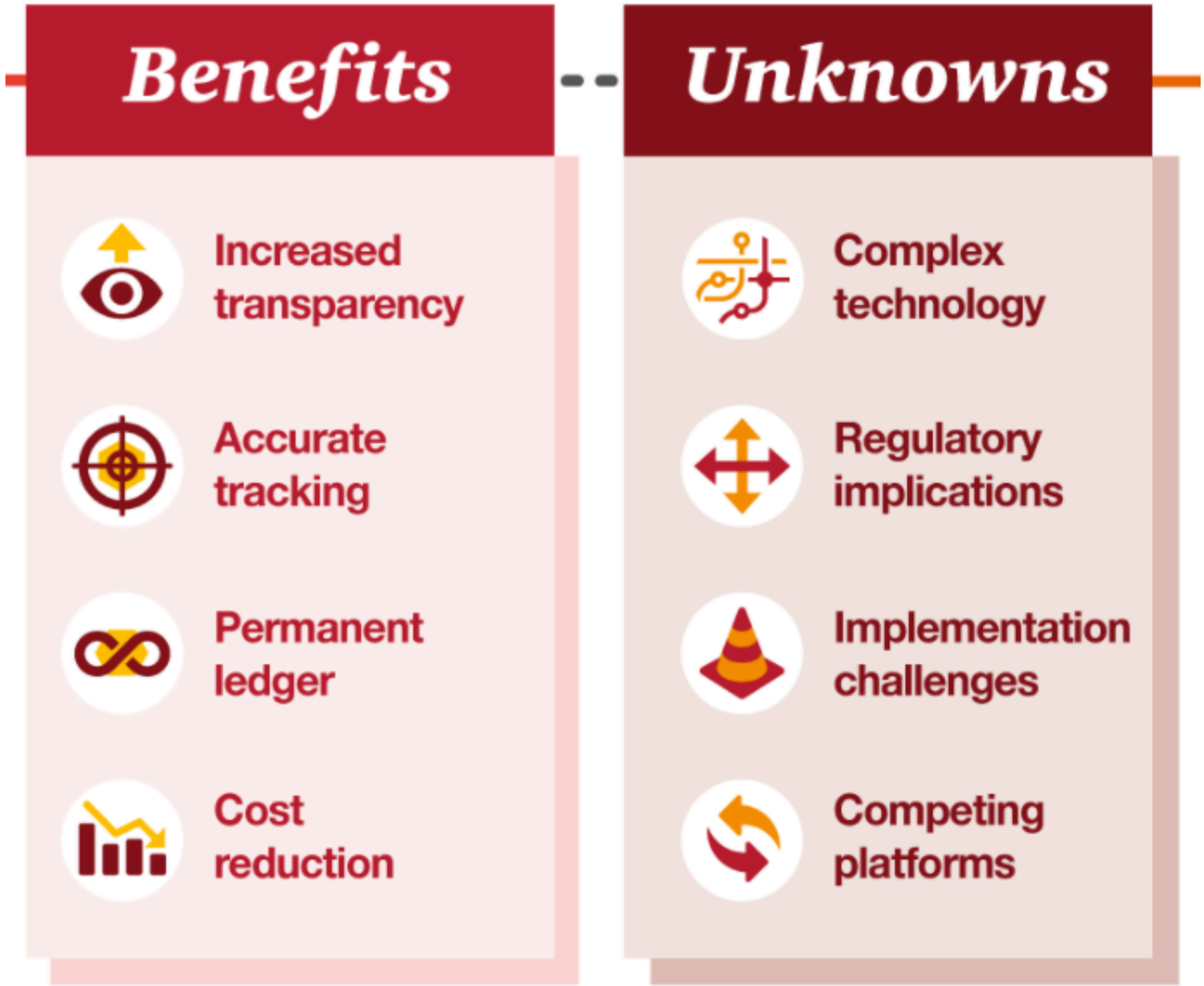
Source: https://www.pwc.com/us/en/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

Blockchain Technology



Source: https://www.pwc.com/us/en/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

Blockchain Technology



Source: https://www.pwc.com/us/en/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

Blockchain Technology Potential Applications

Automotive

Consumers could use the blockchain to manage fractional ownership in autonomous cars.

Financial services

Faster, cheaper settlements could shave billions of dollars from transaction costs while improving transparency.

Source: https://www.pwc.com/us/en/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

Blockchain Technology Potential Applications

Voting

Using a blockchain code, constituents could cast votes via smartphone, tablet or computer, **resulting in immediately verifiable results.**

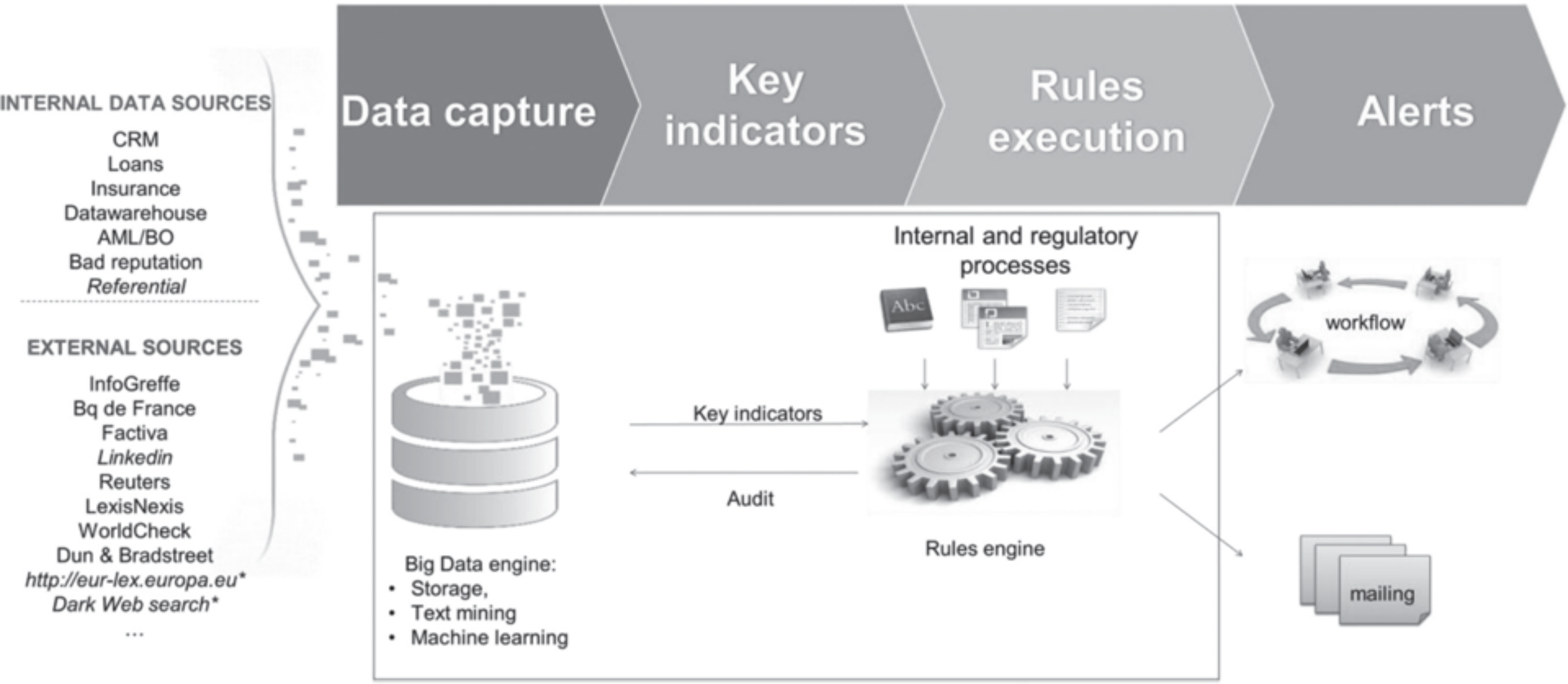
Healthcare

Patients' encrypted health information could be shared with multiple providers without the risk of privacy breaches.

Source: https://www.pwc.com/us/en/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

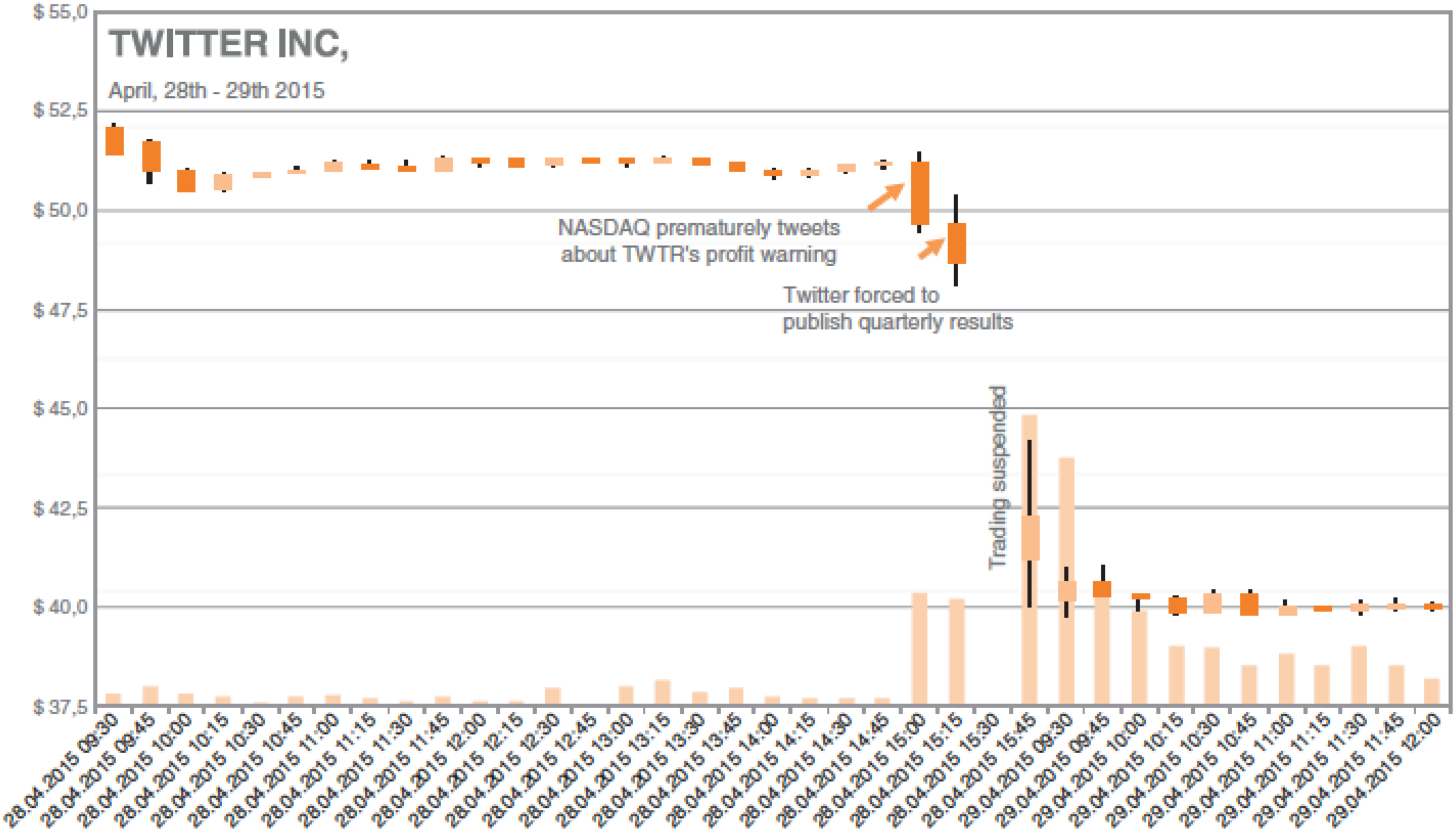
Financial Sentiment Analysis

Big Data Approach to Combining Internal and External Data



Source: Susanne Chishti and Janos Barberis,
The FINTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries, Wiley, 2016

Ultra-Fast Text Analytics in Trading Strategies

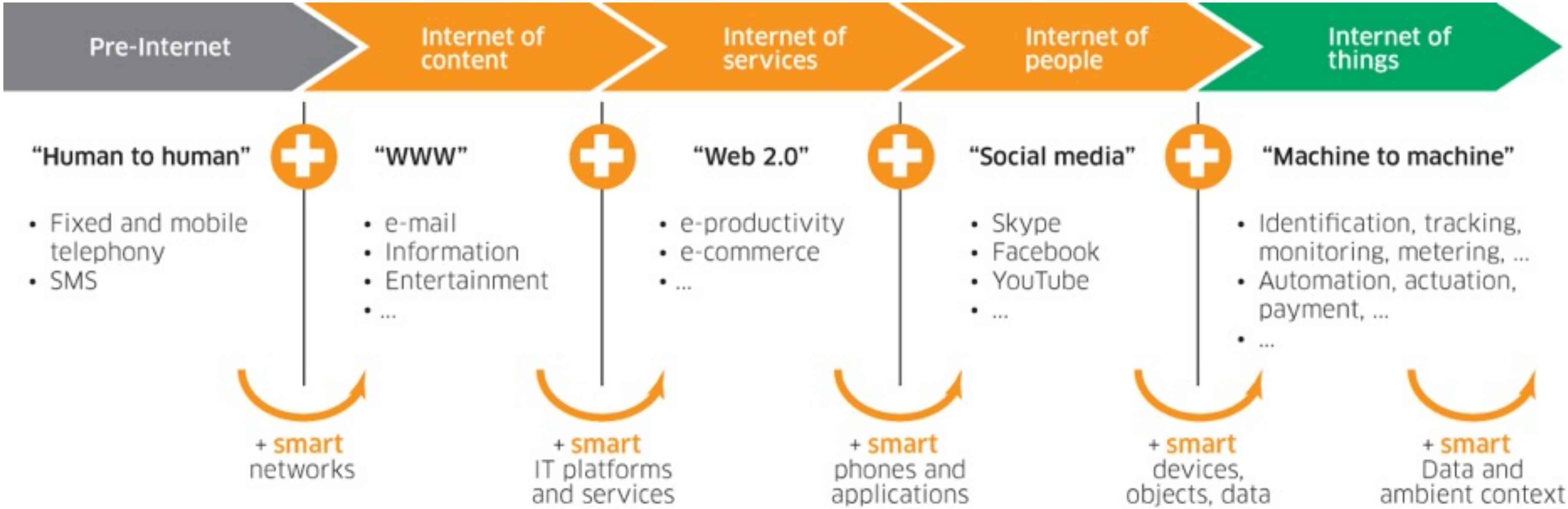


Twitter stock prices affected by news, Source: econob

Source: Susanne Chishti and Janos Barberis,
The FINTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries, Wiley, 2016

Internet Evolution

Internet of People (IoP): Social Media
Internet of Things (IoT): Machine to Machine



Source: Marc Jadoul (2015), The IoT: The next step in internet evolution, March 11, 2015
http://www2.alcatel-lucent.com/techzine/iot-internet-of-things-next-step-evolution/

Social Media

Source: http://hungrywolfmarketing.com/2013/09/09/what-are-your-social-marketing-goals/

Emotions

Love

Anger

Joy

Sadness

Surprise

Fear

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"I bought an iPhone a few days ago.

It was such a nice phone.

The touch screen was really cool.

The voice quality was clear too.

However, my mother was mad with me as I did not tell her before I bought it.

She also thought the phone was too expensive, and wanted me to return it to the shop. … "

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"(1) I bought an iPhone a few days ago.

(2) It was such a **nice** phone.

(3) The touch screen was really **cool**.

(4) The voice quality was **clear** too.

+Positive Opinion

(5) However, my mother was mad with me as I did not tell her before I bought it.

(6) She also thought the phone was too **expensive**, and wanted me to return it to the shop. … "

-Negative Opinion

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

How consumers **think, feel, and act**

Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Emotions

Love

Anger

Joy

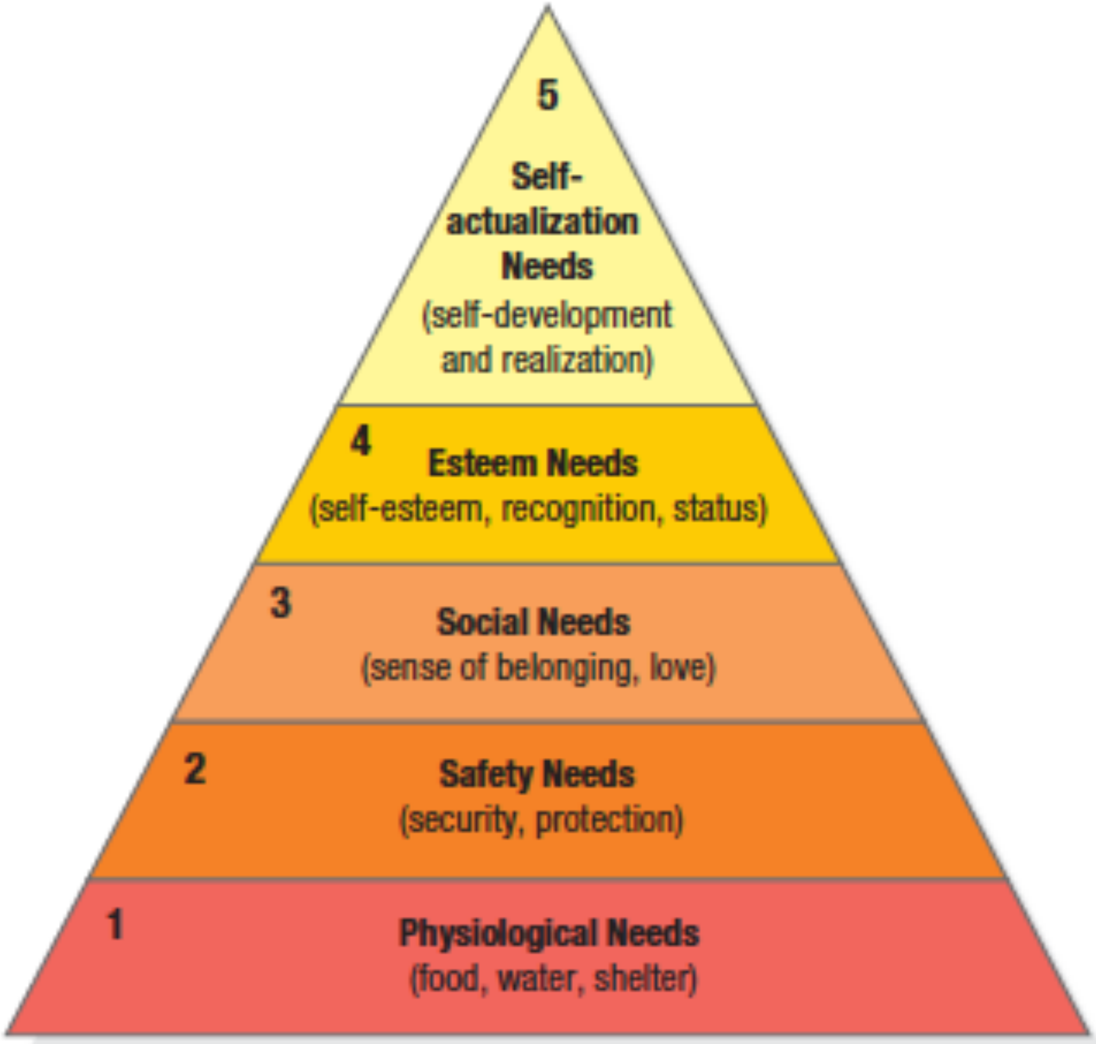
Sadness

Surprise

Fear

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

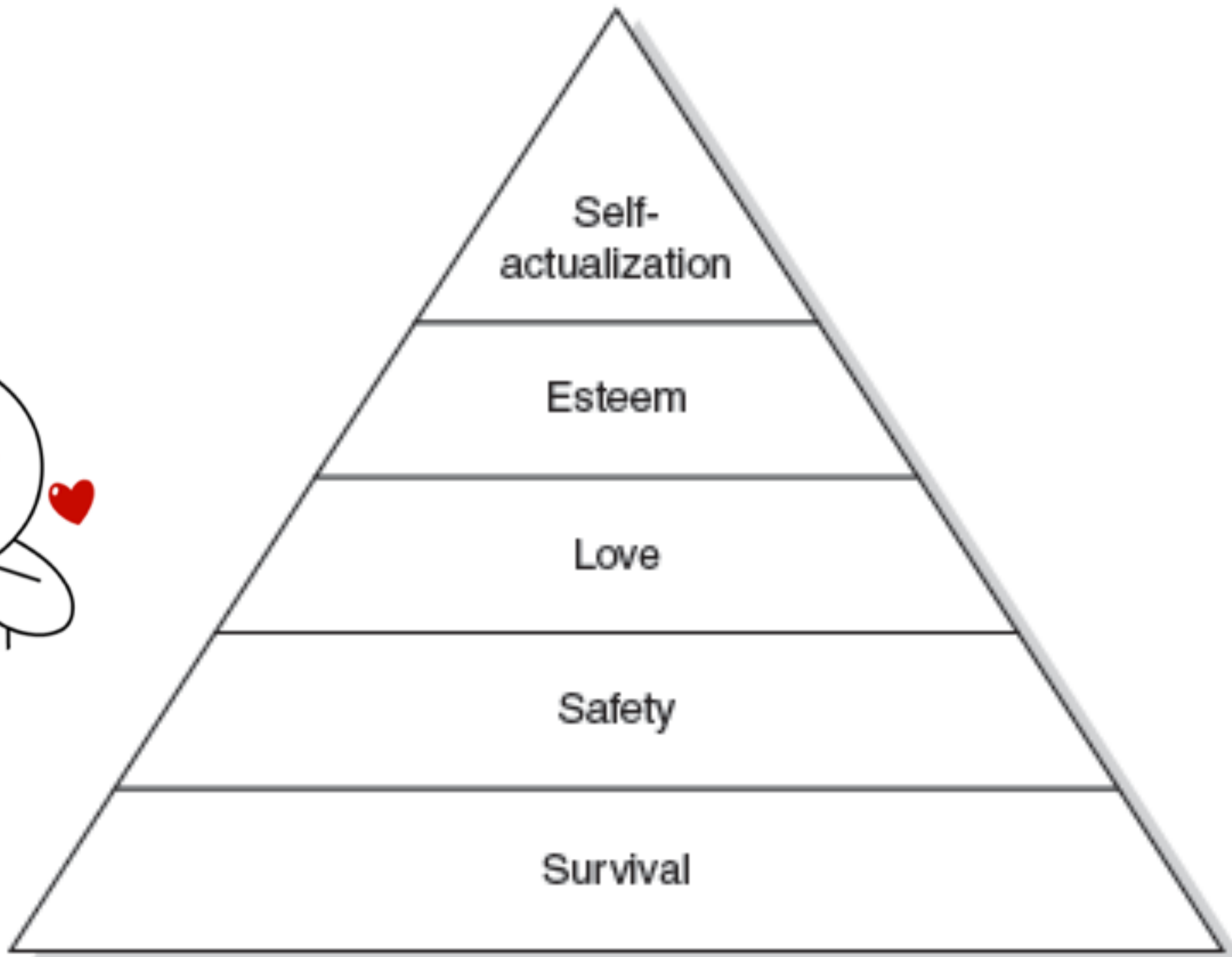
Maslow's Hierarchy of Needs



Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

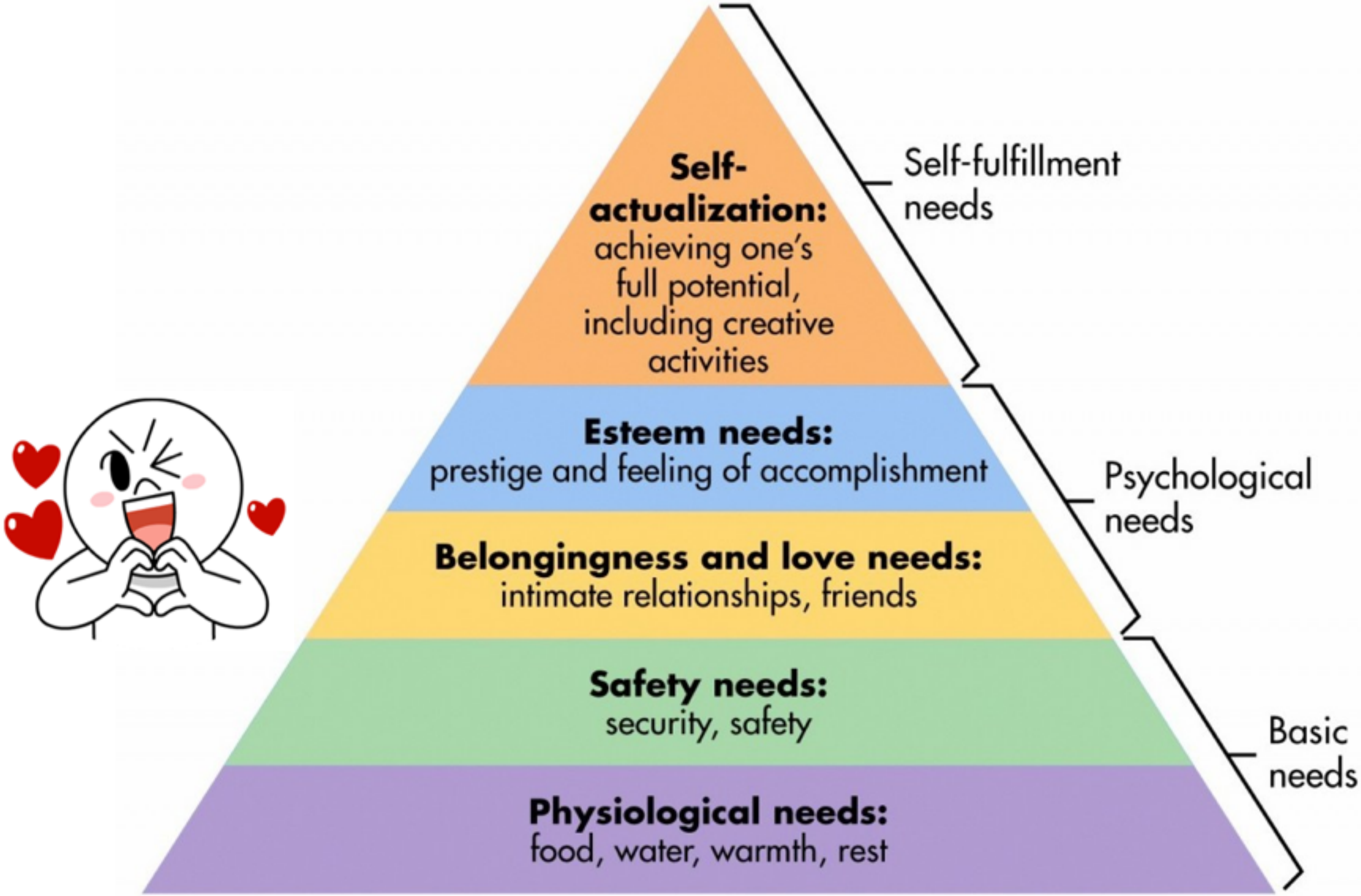
Maslow's hierarchy of human needs

(Maslow, 1943)



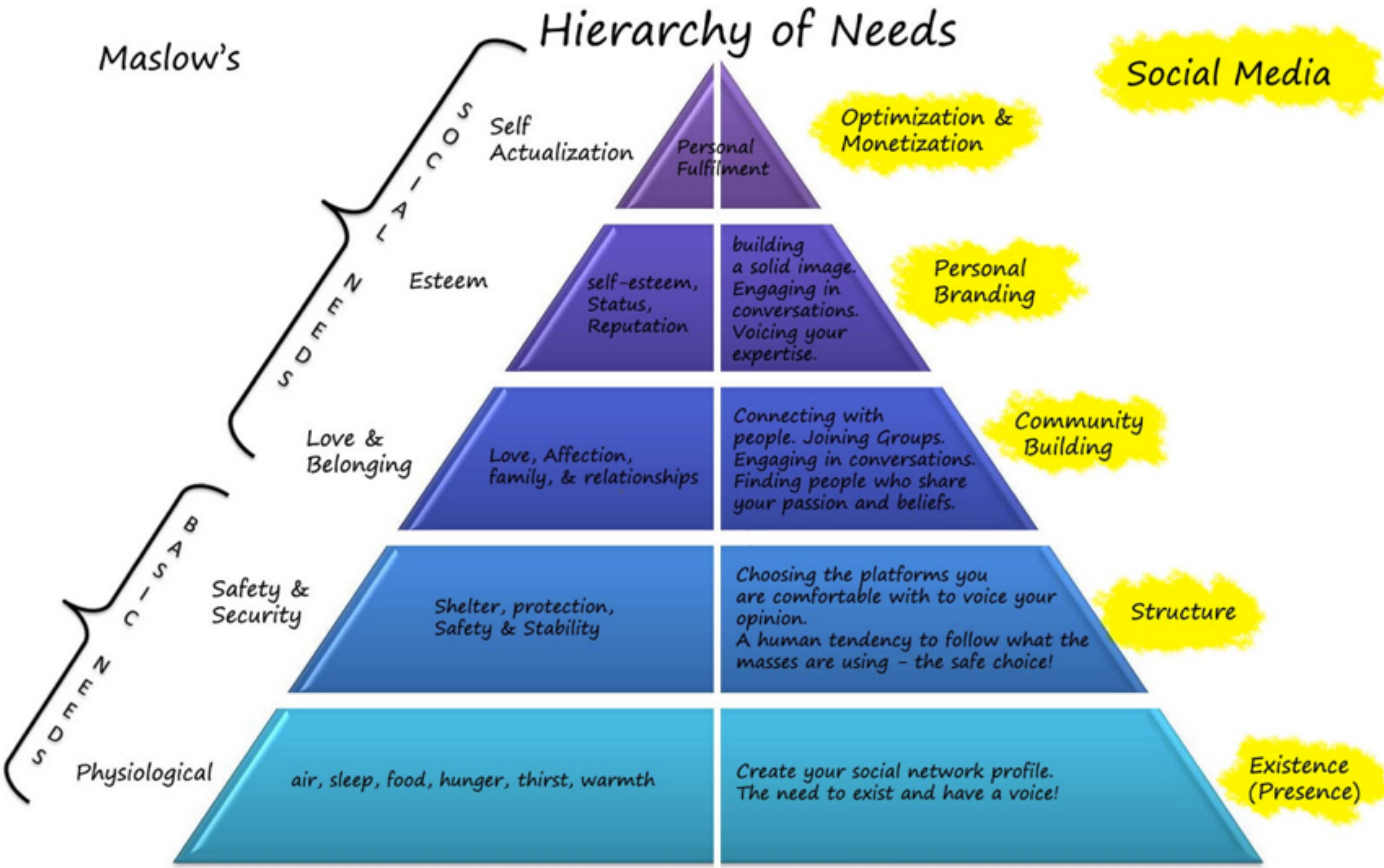
Source: Backer & Saren (2009), Marketing Theory: A Student Text, 2nd Edition, Sage

Maslow's Hierarchy of Needs



Source: http://sixstoriesup.com/social-psyche-what-makes-us-go-social/

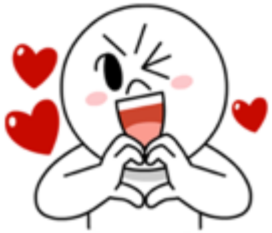
Social Media Hierarchy of Needs

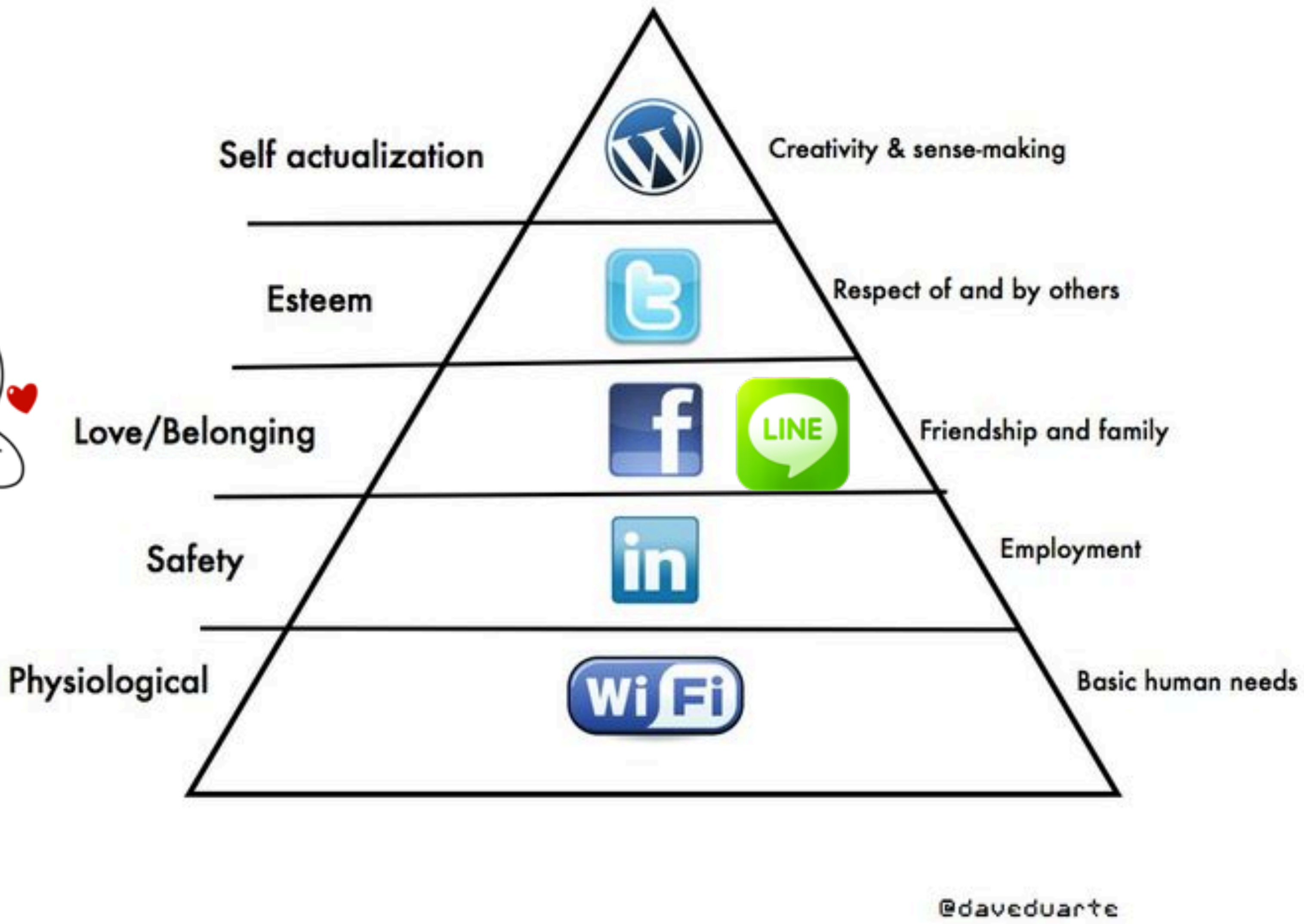


Social Media Hierarchy of Needs - by John Antonios

Source: http://2.bp.blogspot.com/_Rta1VZltiMk/TPavcanFtfl/AAAAAAAAACo/OBGnRL5arSU/s1600/social-media-heirarchy-of-needs1.jpg

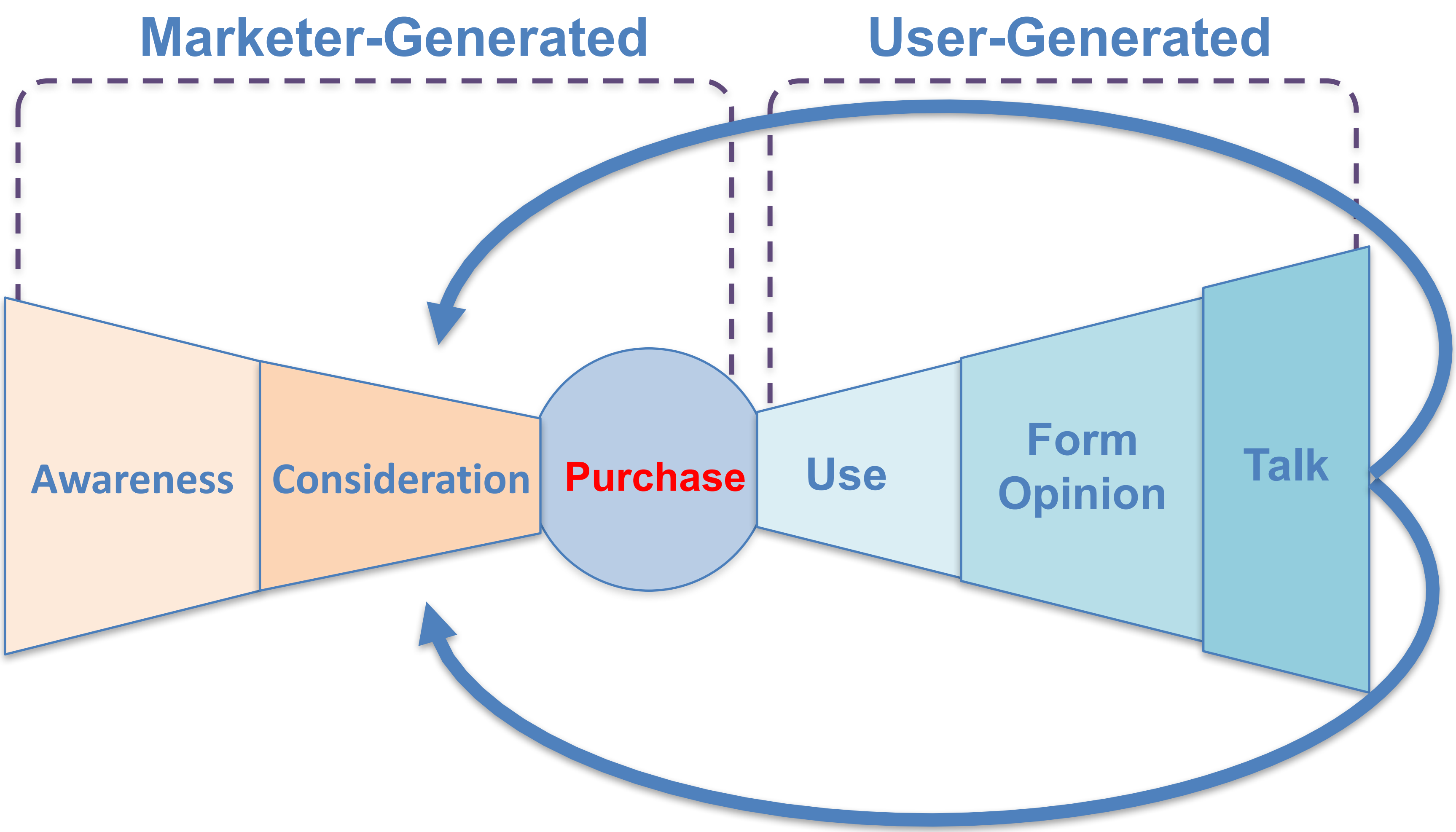
Social Media Hierarchy of Needs





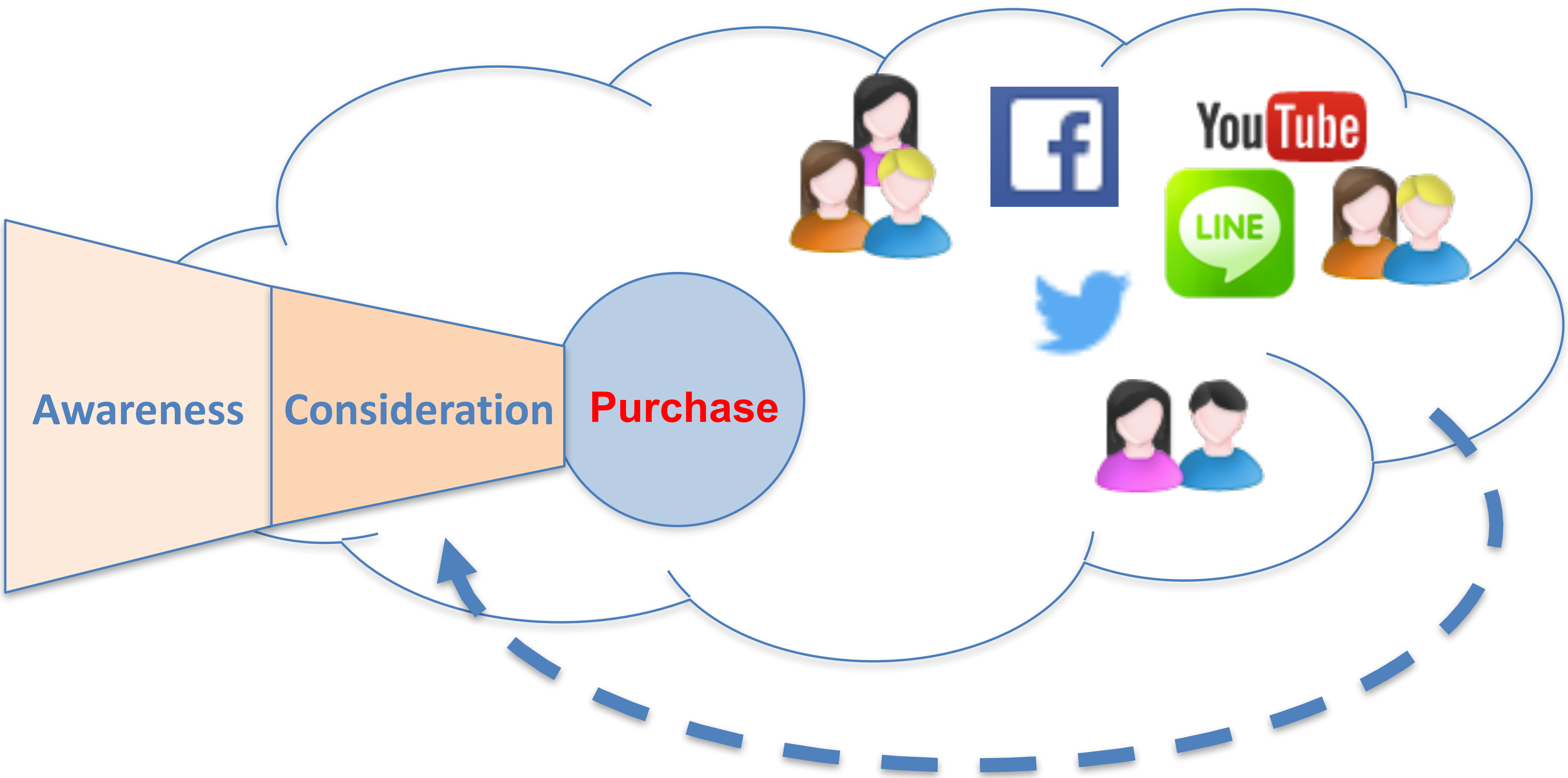
Source: http://www.pinterest.com/pin/186477859309903585/

The Social Feedback Cycle
Consumer Behavior on Social Media



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

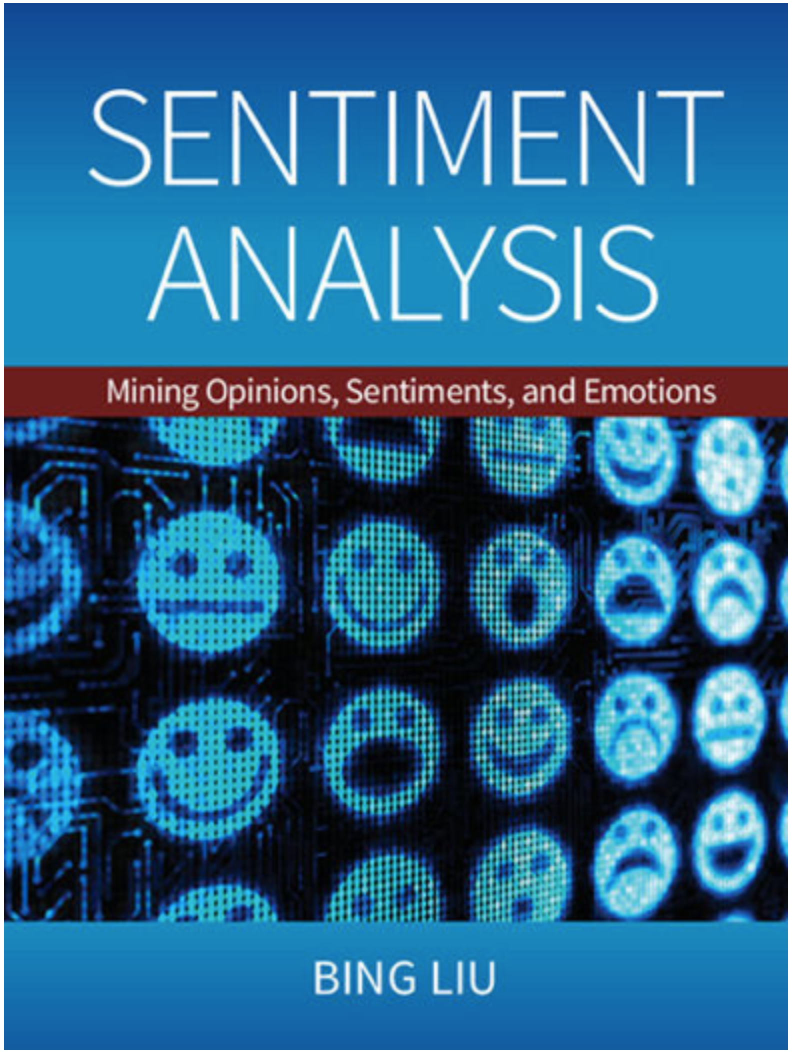
The New Customer Influence Path



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

Architectures of Sentiment Analytics

Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press



http://www.amazon.com/Sentiment-Analysis-Opinions-Sentiments-Emotions/dp/1107017890

Sentiment Analysis and Opinion Mining

- Computational study of
  opinions,
  sentiments,
  subjectivity,
  evaluations,
  attitudes,
  appraisal,
  affects,
  views,
  emotions,
  ets., expressed in text.
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Research Area of Opinion Mining

- Many names and tasks with difference objective and models
  - Sentiment analysis
  - Opinion mining
  - Sentiment mining
  - Subjectivity analysis
  - Affect analysis
  - Emotion detection
  - Opinion spam detection

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Sentiment Analysis

- Sentiment
  - A thought, view, or attitude, especially one based mainly on emotion instead of reason
- Sentiment Analysis
  - opinion mining
  - use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text

Applications of Sentiment Analysis

- Consumer information
  - Product reviews
- Marketing
  - Consumer attitudes
  - Trends
- Politics
  - Politicians want to know voters’ views
  - Voters want to know policitians’ stances and who else supports them
- Social
  - Find like-minded individuals or communities

Sentiment detection

- How to interpret features for sentiment detection?
  - Bag of words (IR)
  - Annotated lexicons (WordNet, SentiWordNet)
  - Syntactic patterns
- Which features to use?
  - Words (unigrams)
  - Phrases/n-grams
  - Sentences

Problem statement of Opinion Mining

- Two aspects of abstraction
  - Opinion definition
    - What is an opinion?
    - What is the structured definition of opinion?
  - Opinion summarization
    - Opinion are subjective
      - An opinion from a single person (unless a VIP) is often not sufficient for action
    - We need opinions from many people, and thus opinion summarization.

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

What is an opinion?

- Id: **Abc123** on **5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"
- One can look at this review/blog at the
  - Document level
    - Is this review + or -?
  - Sentence level
    - Is each sentence + or -?
  - Entity and feature/aspect level

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Entity and aspect/feature level

- Id: **Abc123** on **5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"
- What do we see?
  - Opinion targets: entities and their features/aspects
  - Sentiments: positive and negative
  - Opinion holders: persons who hold the opinions
  - Time: when opinion are expressed

Two main types of opinions

- Regular opinions: Sentiment/Opinion expressions on some target entities
  - Direct opinions: sentiment expressions on one object:
    - "The touch screen is really cool."
    - "The picture quality of this camera is great"
  - Indirect opinions: comparisons, relations expressing similarities or differences (objective or subjective) of more than one object
    - "phone X is cheaper than phone Y." (objective)
    - "phone X is better than phone Y." (subjective)
- Comparative opinions: comparisons of more than one entity.
  - "iPhone is better than Blackberry."

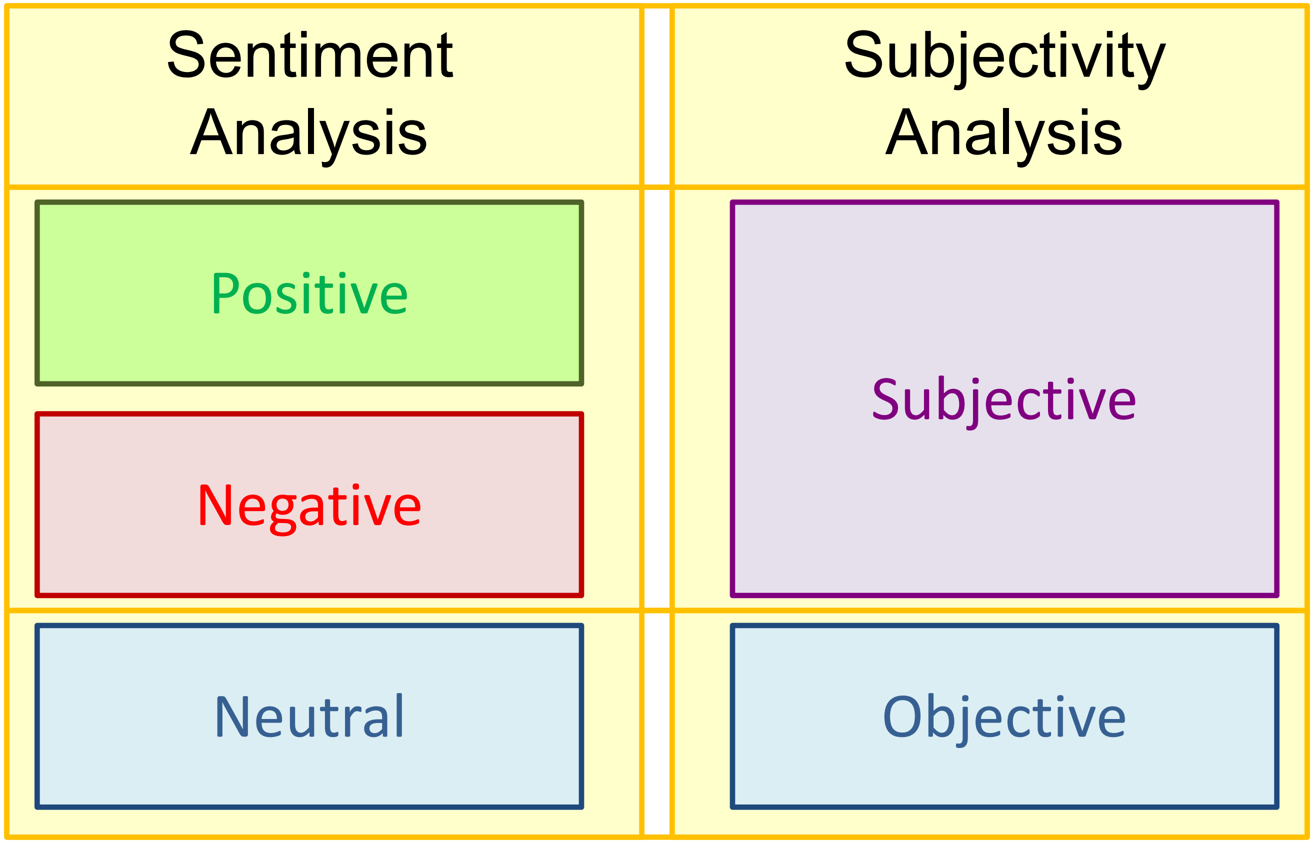
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Subjective and Objective

- Objective
  - An objective sentence expresses some factual information about the world.
  - "I returned the phone yesterday."
  - Objective sentences can implicitly indicate opinions
    - "The earphone broke in two days."
- Subjective
  - A subjective sentence expresses some personal feelings or beliefs.
  - "The voice on my phone was not so clear"
  - Not every subjective sentence contains an opinion
    - "I wanted a phone with good voice quality"
- → Subjective analysis

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Sentiment Analysis
vs.
Subjectivity Analysis



A (regular) opinion

- Opinion (a restricted definition)
  - An opinion (regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.
- Sentiment orientation of an opinion
  - Positive, negative, or neutral (no opinion)
  - Also called:
    - Opinion orientation
    - Semantic orientation
    - Sentiment polarity

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Entity and aspect

- Definition of Entity:
  - An *entity e* is a product, person, event, organization, or topic.
  - e is represented as
    - A hierarchy of components, sub-components.
    - Each node represents a components and is associated with a set of attributes of the components
- An opinion can be expressed on any node or attribute of the node
- Aspects(features)
  - represent both components and attribute

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Opinion Definition

- An opinion is a quintuple
  (e_j, a_{jk}, so_{ijkl}, h_i, t_l)
  where
  - e_j is a target entity.
  - a_{jk} is an aspect/feature of the entity e_j .
  - so_{ijkl} is the sentiment value of the opinion from the opinion holder on feature of entity at time.
    so_{ijkl} is +ve, -ve, or neu, or more granular ratings
  - h_i is an opinion holder.
  - t_l is the time when the opinion is expressed.
- (e_j, a_{jk}) is also called opinion target

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Terminologies

- Entity: object
- Aspect: feature, attribute, facet
- Opinion holder: opinion source

- Topic: entity, aspect

- Product features, political issues

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Subjectivity and Emotion

- Sentence subjectivity
  - An objective sentence presents some factual information, while a subjective sentence expresses some personal feelings, views, emotions, or beliefs.
- Emotion
  - Emotions are people’s subjective feelings and thoughts.

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Classification Based on Supervised Learning

- Sentiment classification
  - Supervised learning Problem
  - Three classes
    - *Positive*
    - *Negative*
    - *Neutral*

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Opinion words in Sentiment classification

- topic-based classification
  - topic-related words are important
    - e.g., *politics, sciences, sports*
- Sentiment classification
  - topic-related words are unimportant
  - **opinion words** (also called **sentiment words)**
    - **that indicate positive or negative opinions** are important,
      e.g., *great, excellent, amazing, horrible, bad, worst*

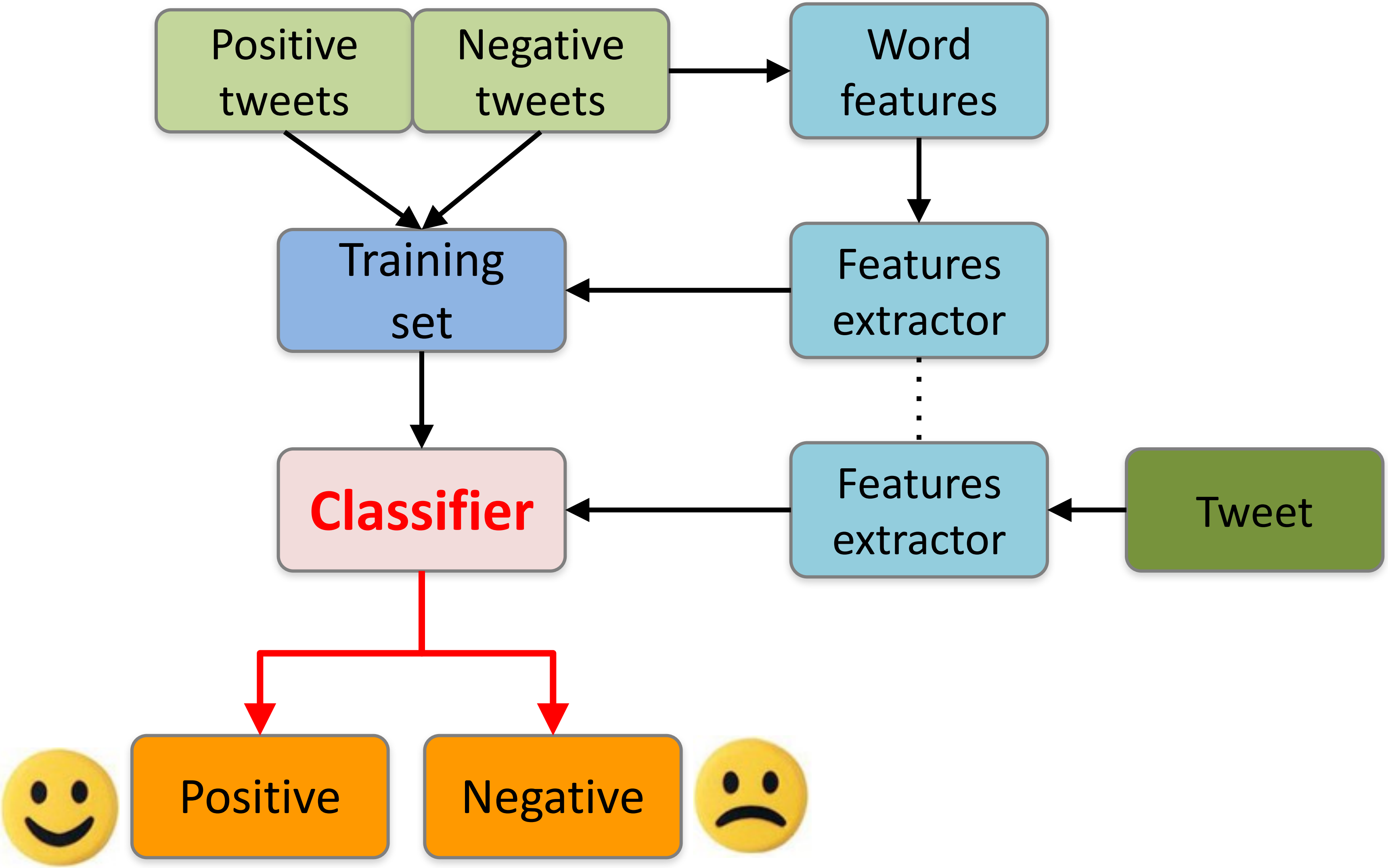
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Features in Opinion Mining

- *Terms and their frequency*
  - *TF-IDF*
- *Part of speech (POS)*
  - Adjectives
- *Opinion words and phrases*
  - *beautiful, wonderful, good, and amazing are positive opinion words*
  - *bad, poor, and terrible are negative opinion words.*
  - opinion phrases and idioms,
    e.g., *cost someone an arm and a leg*
- *Rules of opinions*
- *Negations*
- *Syntactic dependency*

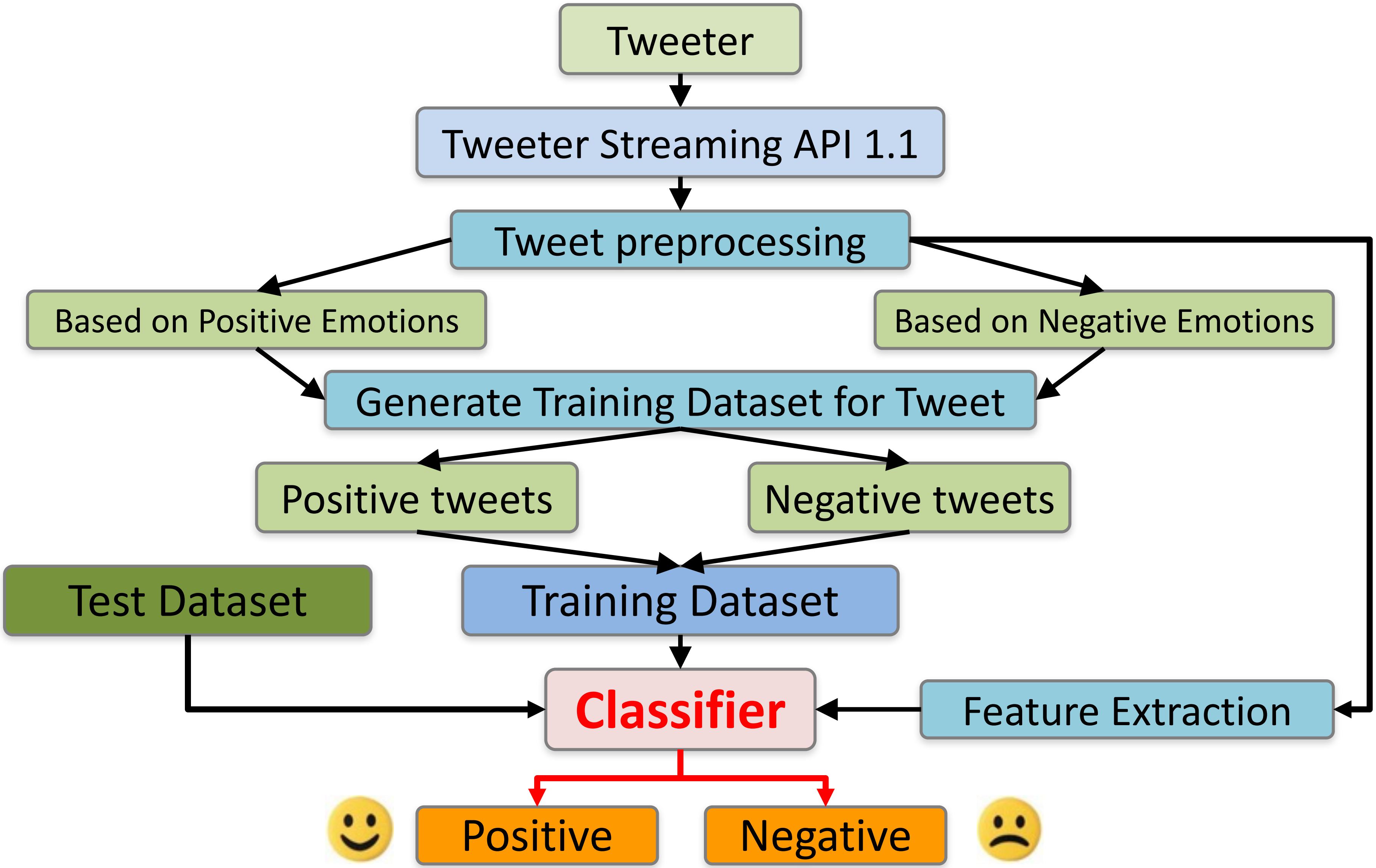
Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Sentiment Analysis Architecture



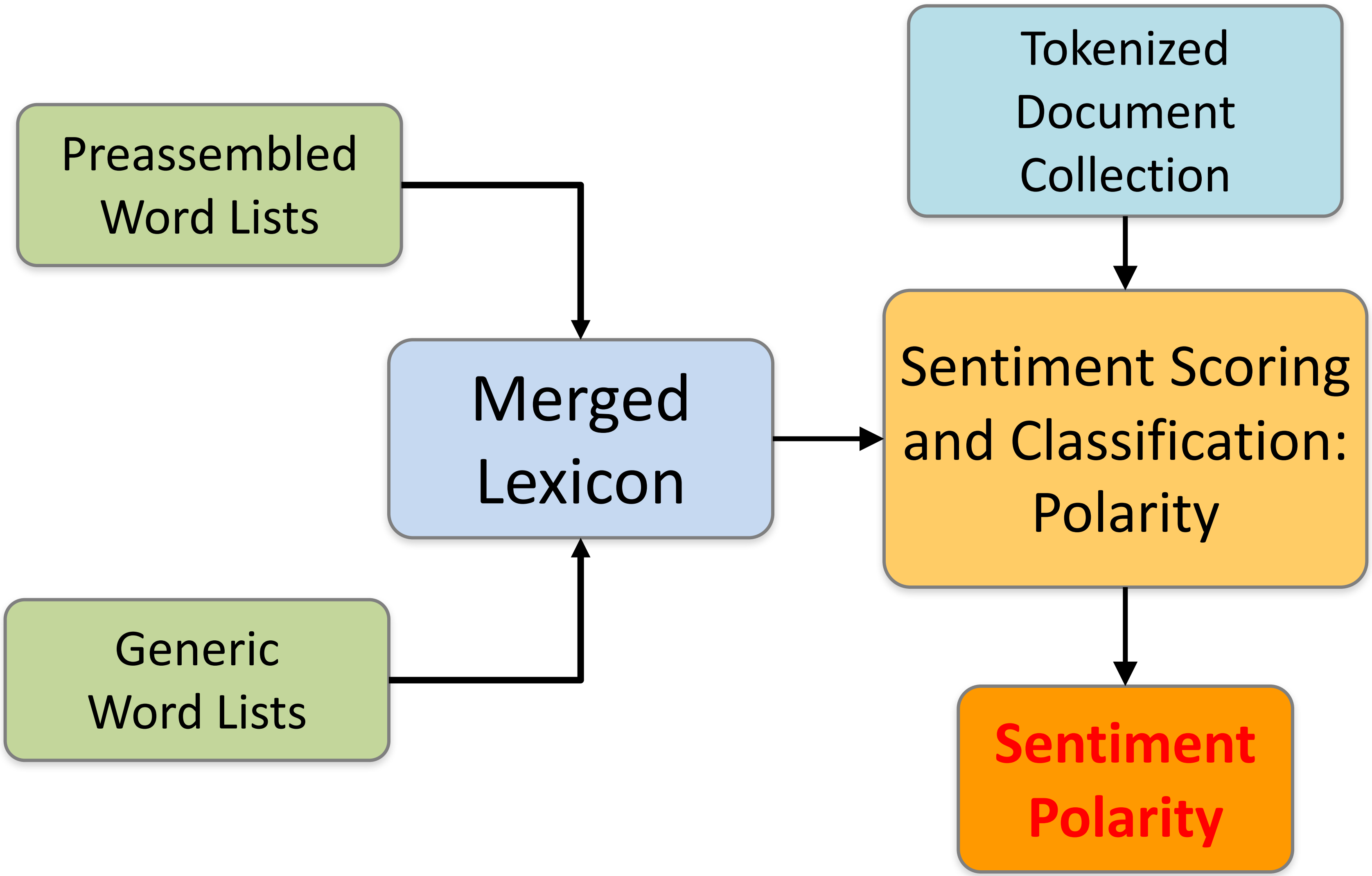
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Classification Based on Emoticons



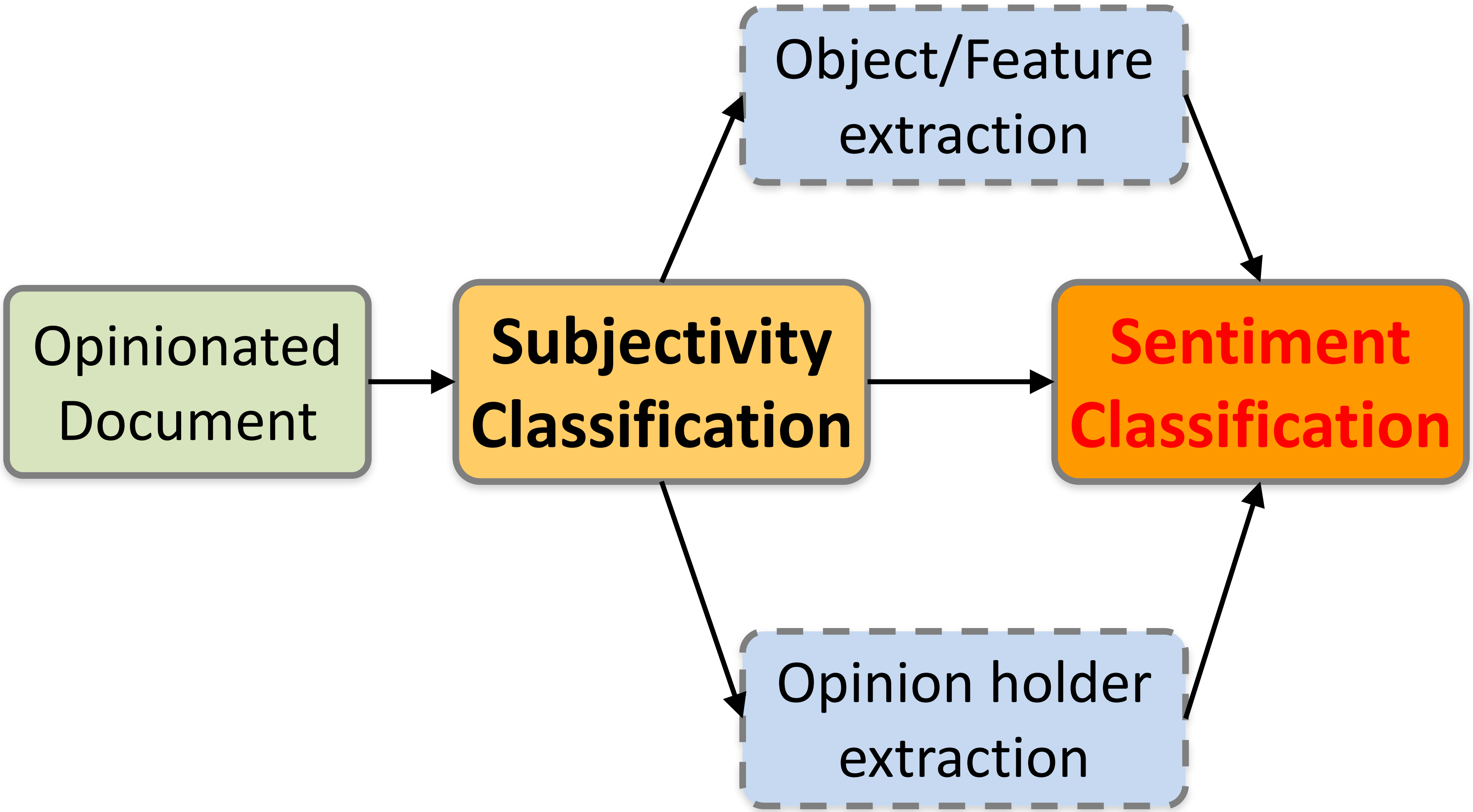
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Lexicon-Based Model



Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis Tasks

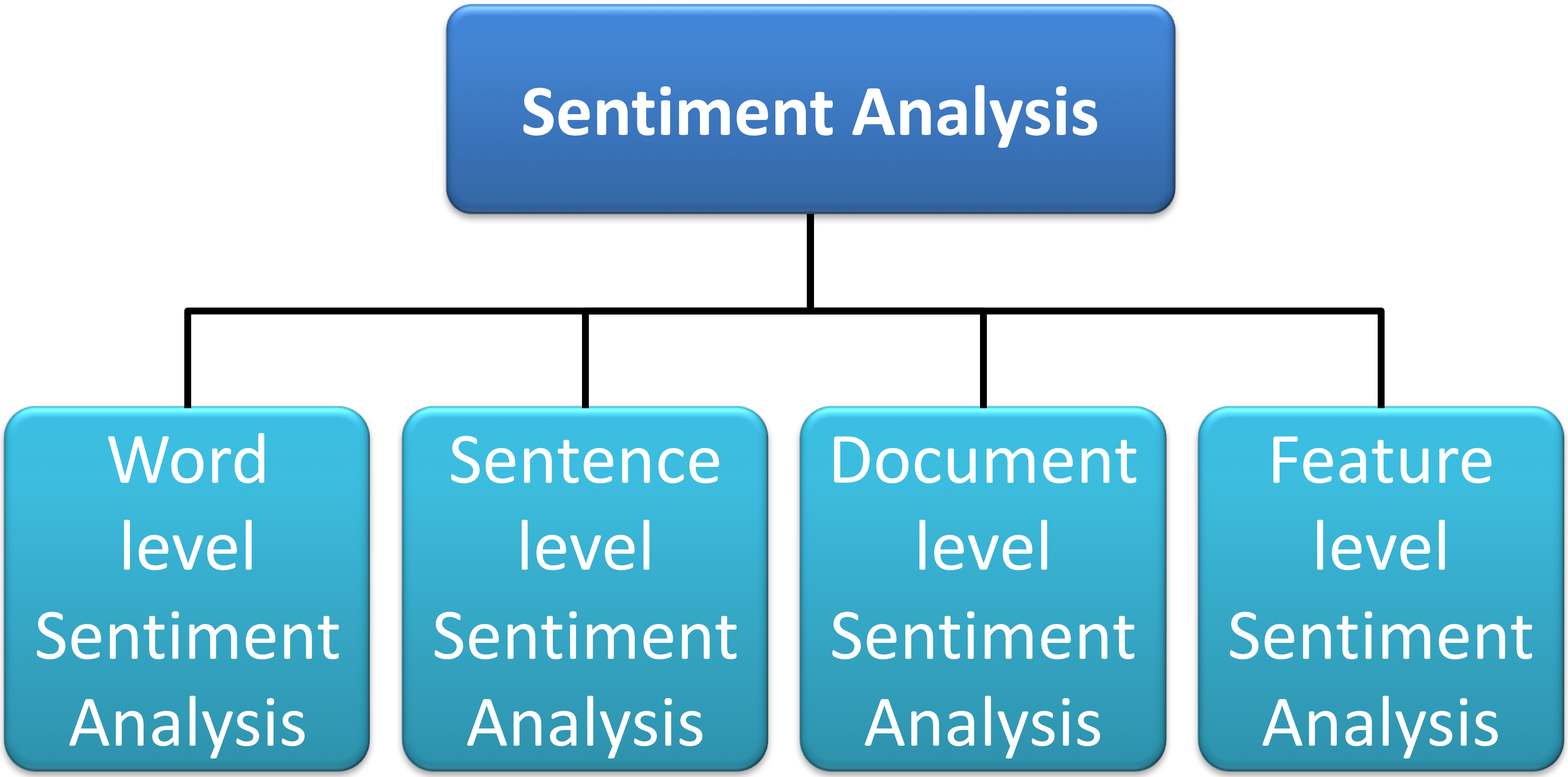


Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis vs. Subjectivity Analysis

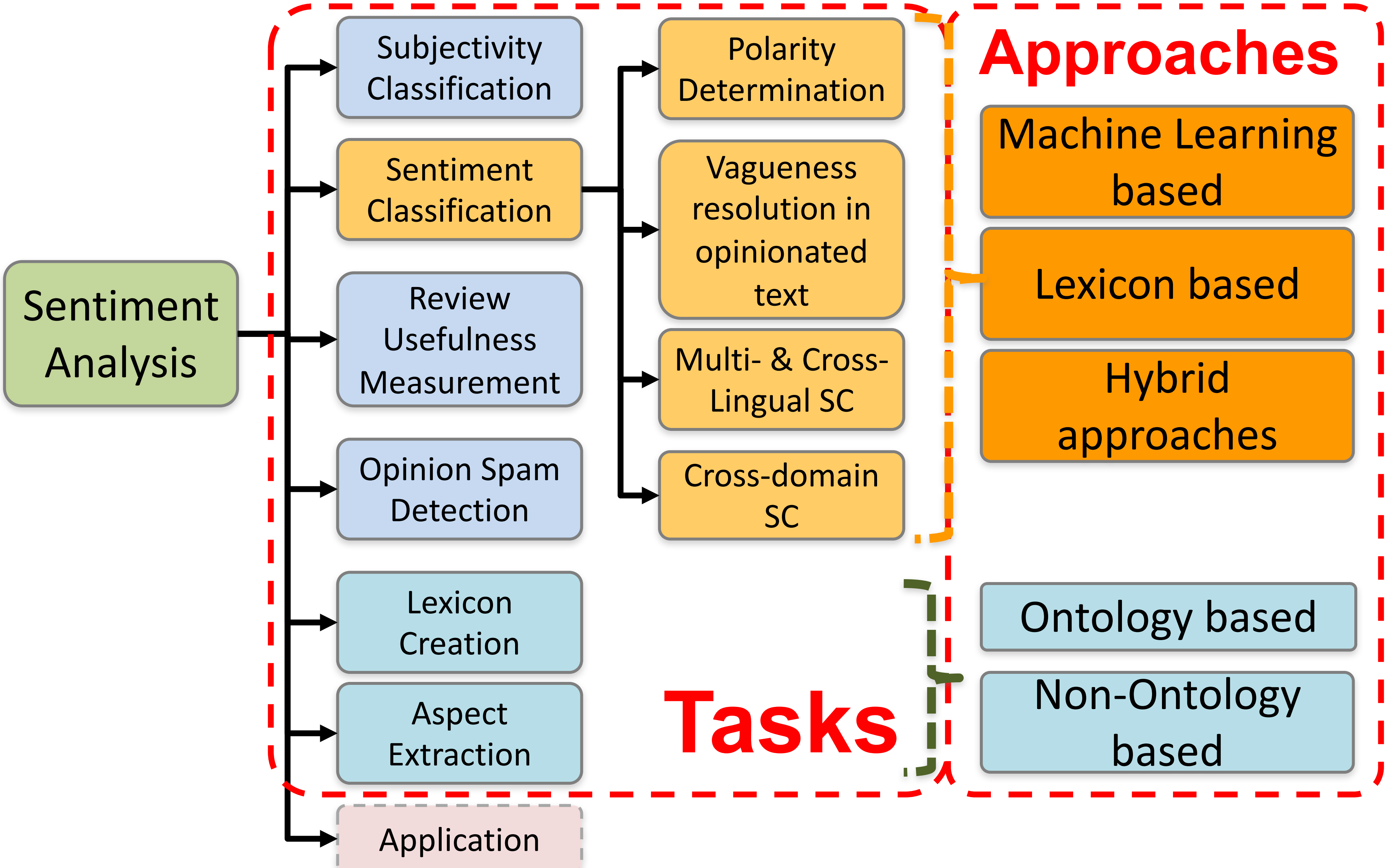
Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	
Neutral	Objective

Levels of Sentiment Analysis



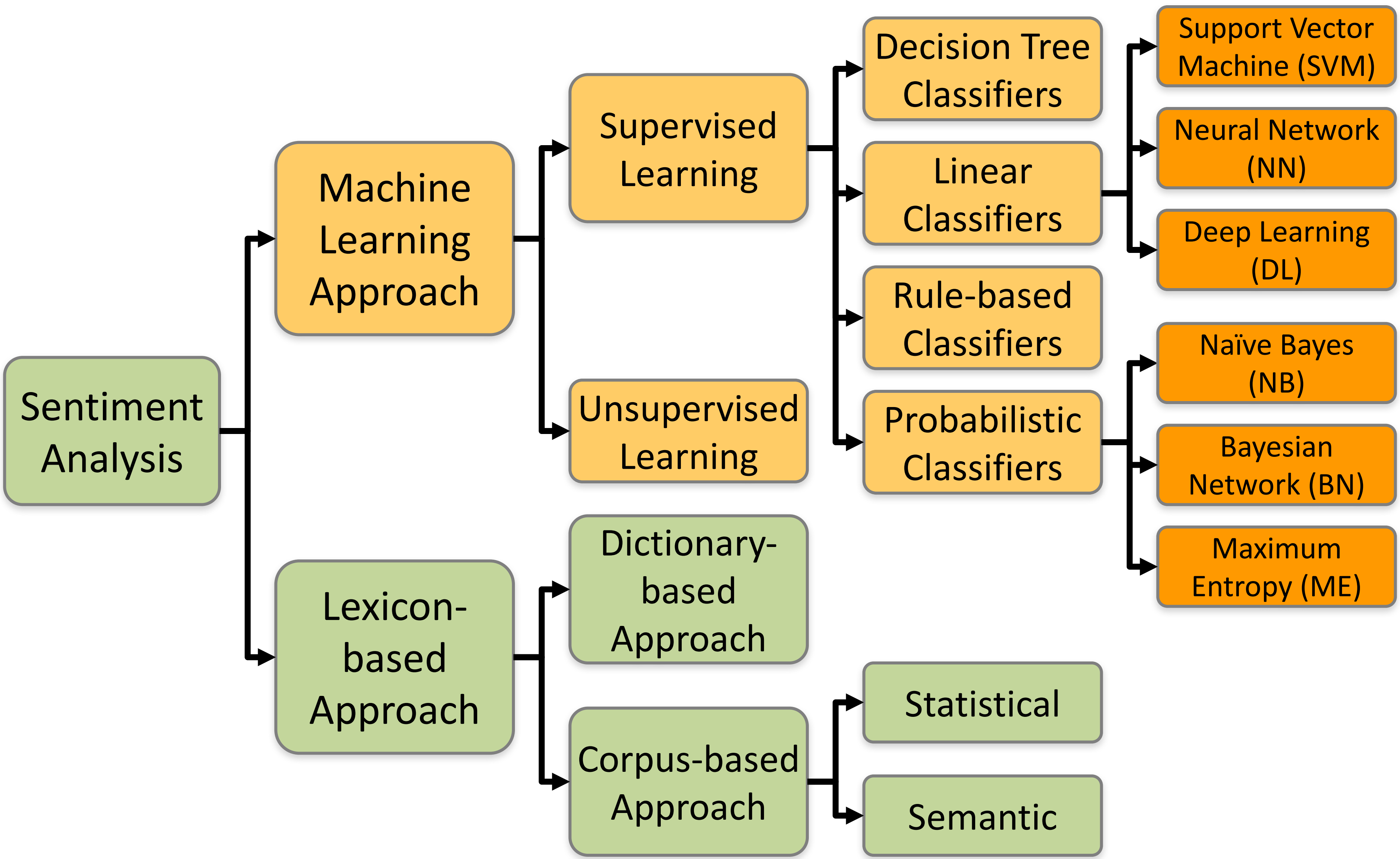
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis



Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Techniques



Source: Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.

A Brief Summary of Sentiment Analysis Methods

Study	Analysis Task	Sentiment Identification		Sentiment Aggregation		Nature of Measure
		Method	Level	Method	Level	
Hu and Li, 2011	Polarity	ML (Probabilistic model)	Snippet			Valence
Li and Wu, 2010	Polarity	Lexicon/Rule	Phrase	Sum	Snippet	Valence
Thelwall et al., 2010	Polarity	Lexicon/Rule	Sentence	Max & Min	Snippet	Range
Boiy and Moens, 2009	Both	ML (Cascade ensemble)	Sentence			Valence
Chung 2009	Polarity	Lexicon	Phrase	Average	Sentence	Valence
Wilson, Wiebe, and Hoffmann, 2009	Both	ML (SVM, AdaBoost, Rule, etc.)	Phrase			Valence
Zhang et al., 2009	Polarity	Lexicon/Rule	Sentence	Weighted average	Snippet	Valence
Abbasi, Chen, and Salem, 2008	Polarity	ML (GA + feature selection)	Snippet			Valence
Subrahmanian and Reforgiato, 2008	Polarity	Lexicon/Rule	Phrase	Rule	Snippet	Valence
Tan and Zhang 2008	Polarity	ML (SVM, Winnow, NB, etc.)	Snippet			Valence
Airoldi, Bai, and Padman, 2007	Polarity	ML (Markov Blanket)	Snippet			Valence
Das and Chen, 2007	Polarity	ML (Bayesian, Discriminate, etc.)	Snippet	Average	Daily	Valence
Liu et al., 2007	Polarity	ML (PLSA)	Snippet			Valence
Kennedy and Inkpen, 2006	Polarity	Lexicon/Rule, ML (SVM)	Phrase	Count	Snippet	Valence
Mishne 2006	Polarity	Lexicon	Phrase	Average	Snippet	Valence
Liu et al., 2005	Polarity	Lexicon/Rule	Phrase	Distribution	Object	Range
Mishne 2005	Polarity	ML (SVM)	Snippet			Valence
Popescu and Etzioni 2005	Polarity	Lexicon/Rule	Phrase			Valence
Efron 2004	Polarity	ML (SVN, NB)	Snippet			Valence
Wilson, Wiebe, and Hwa, 2004	Both	ML (SVM, AdaBoost, Rule, etc.)	Sentence			Valence
Nigam and Hurst 2004	Polarity	Lexicon/Rule	Chunk	Rule	Sentence	Valence
Dave, Lawrence, and Pennock, 2003	Polarity	ML (SVM, Rainbow, etc.)	Snippet			Valence
Nasukawa and Yi 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yi et al., 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yu and Hatzivassiloglou 2003	Both	ML (NB) + Lexicon/Rule	Phrase	Average	Sentence	Valence
Pang, Lee, and Vaithyanathan 2002	Polarity	ML (SVM, MaxEnt, NB)	Snippet			Valence
Subasic and Huettner 2001	Polarity	Lexicon/Fuzzy logic	Phrase	Average	Snippet	Valence
Turney 2001	Polarity	Lexicon/Rule	Phrase	Average	Snippet	Valence

(Both = Subjectivity and Polarity; ML= Machine Learning; Lexicon/Rule= Lexicon enhanced by linguistic rules)

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Word-of-Mouth (WOM)

- “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

- “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

This
book
is
the
best
written
documentary
thus
far
,
yet
sadly
,
there
is
no
soft
cover
edition
.

Word	POS
This	DT
book	NN
is	VBZ
the	DT
best	JJS
written	VBN
documentary	NN
thus	RB
far	RB
,	,
yet	RB
sadly	RB
,	,
there	EX
is	VBZ
no	DT
soft	JJ
cover	NN
edition	NN
.	.

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Conversion of text representation

Word Vector (WV)	Polarity Score Vector (PSV) *pscore*	*nscore*		Microstate Sequence (MS)
This	0	0	Neutral (0)	0
book	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
the	0	0	Neutral (0)	0
best	0.75	0	Positive (0.75)	1
written	0	0	Neutral (0)	0
documentary	0	0	Neutral (0)	0
thus	0.375	0	Positive (0.375)	1
far	0.375	0	Positive (0.375)	1
,				
yet	0	0.125	Negative (0.125)	-1
sadly	0.25	0.5	Negative (0.25)	-1
,				
there	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
no	0	0.75	Negative (0.75)	-1
soft	0	0	Neutral (0)	0
cover	0	0	Neutral (0)	0
edition	0	0	Neutral (0)	0
.				

SentiWordNet Lookup

Microstate Mapping

Probability Mapping

Probability Distribution (P)

P("1")=3/17

P("-1")=3/17

P("0")=11/17

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

SenticNet

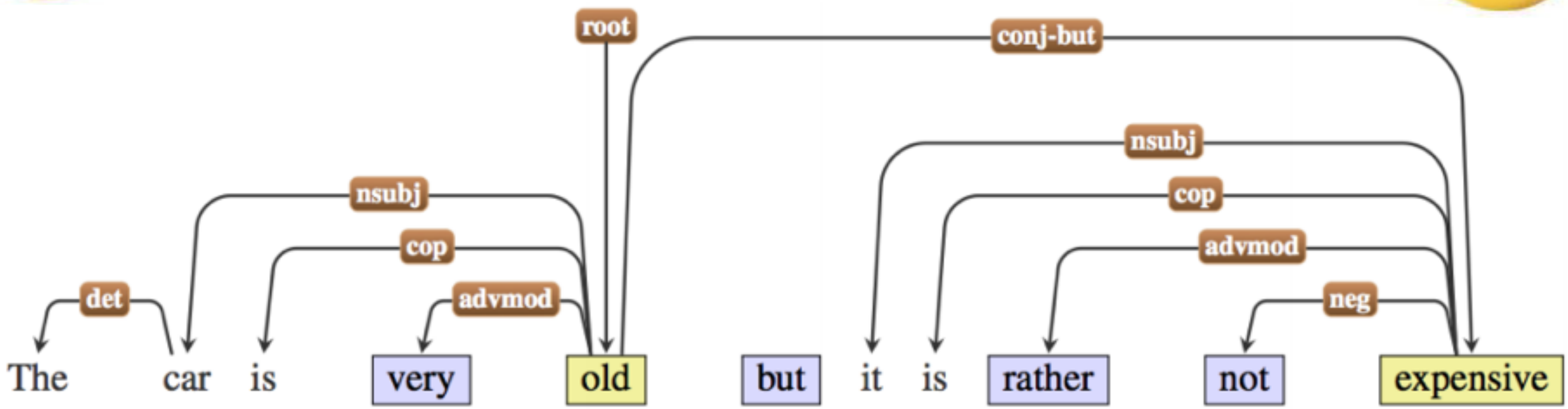
The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet

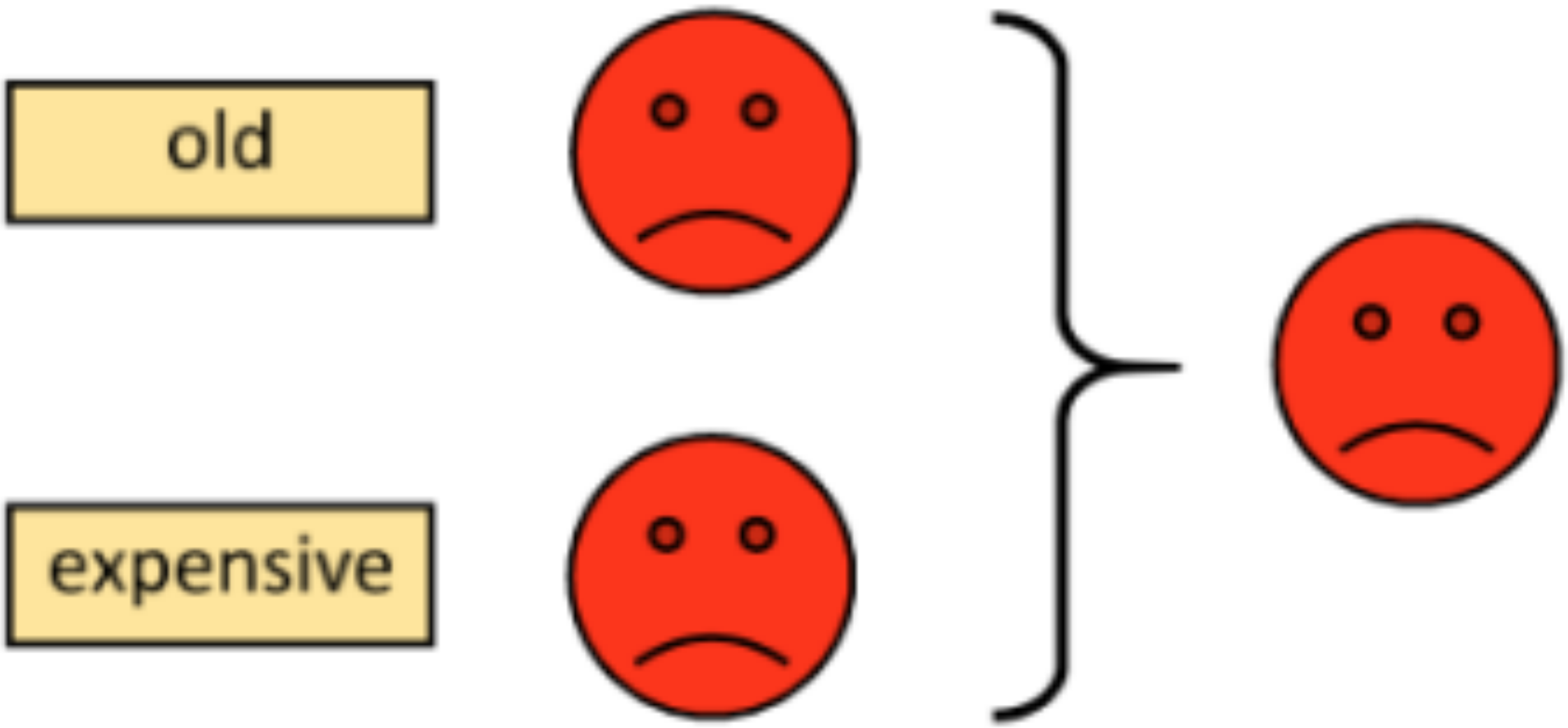


The car is very old but it is rather not expensive.

The car is <u>very</u> old <u>but</u> it is <u>rather</u> <u>not</u> expensive.

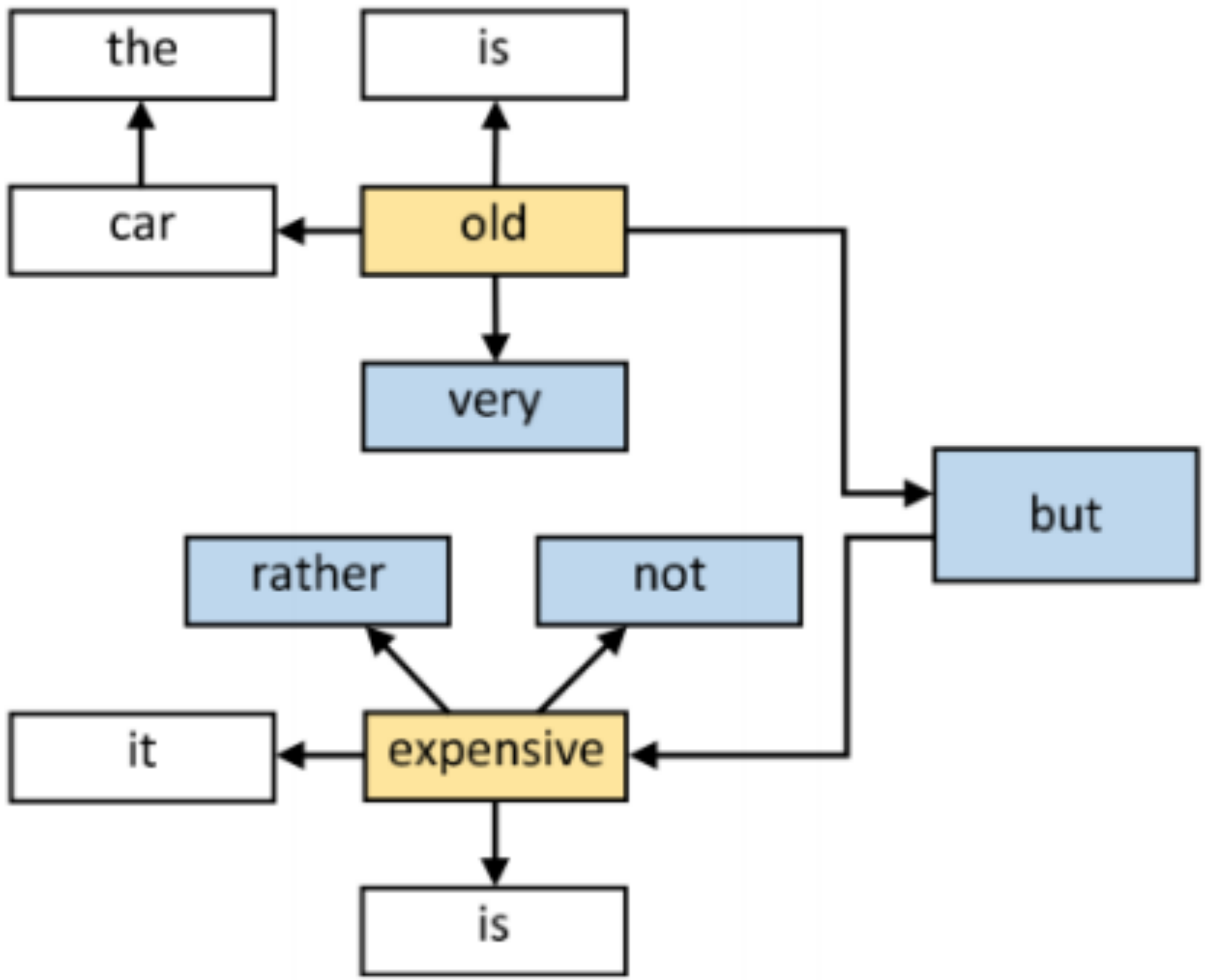
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



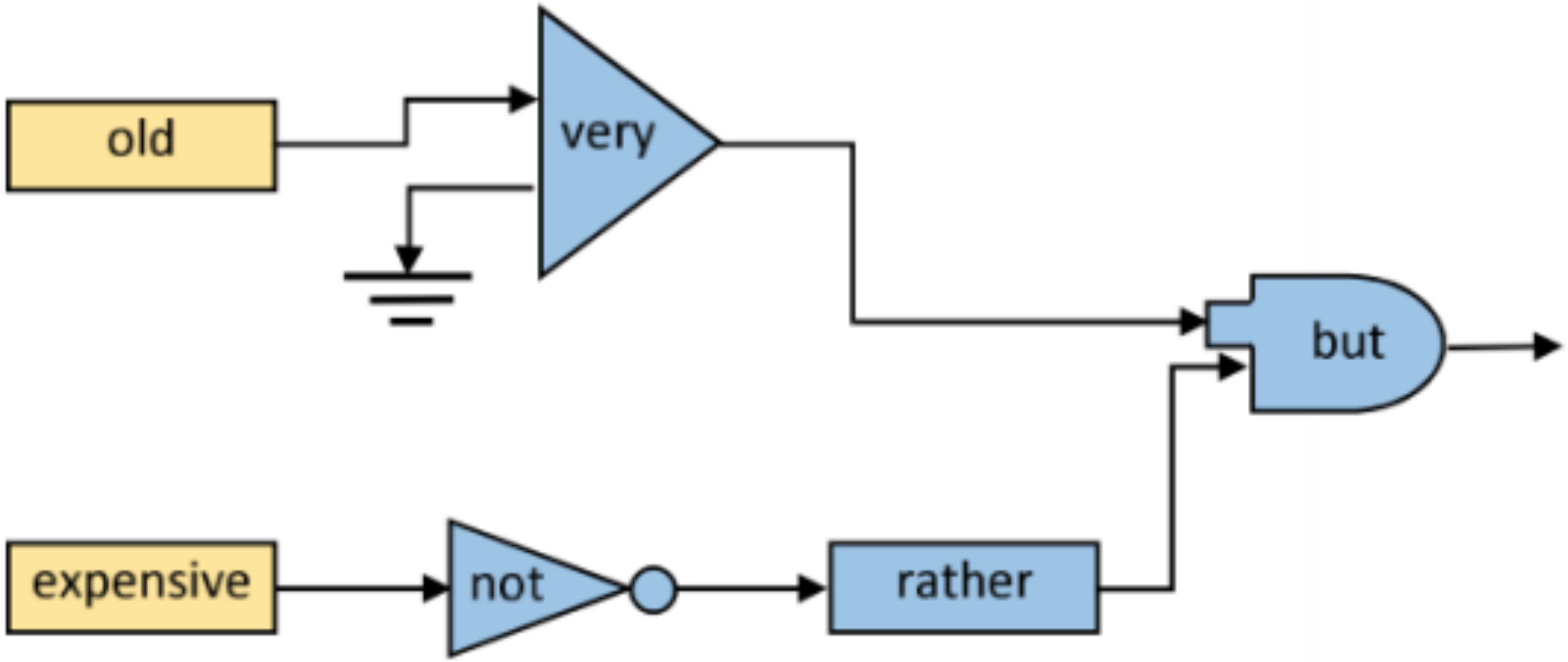
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



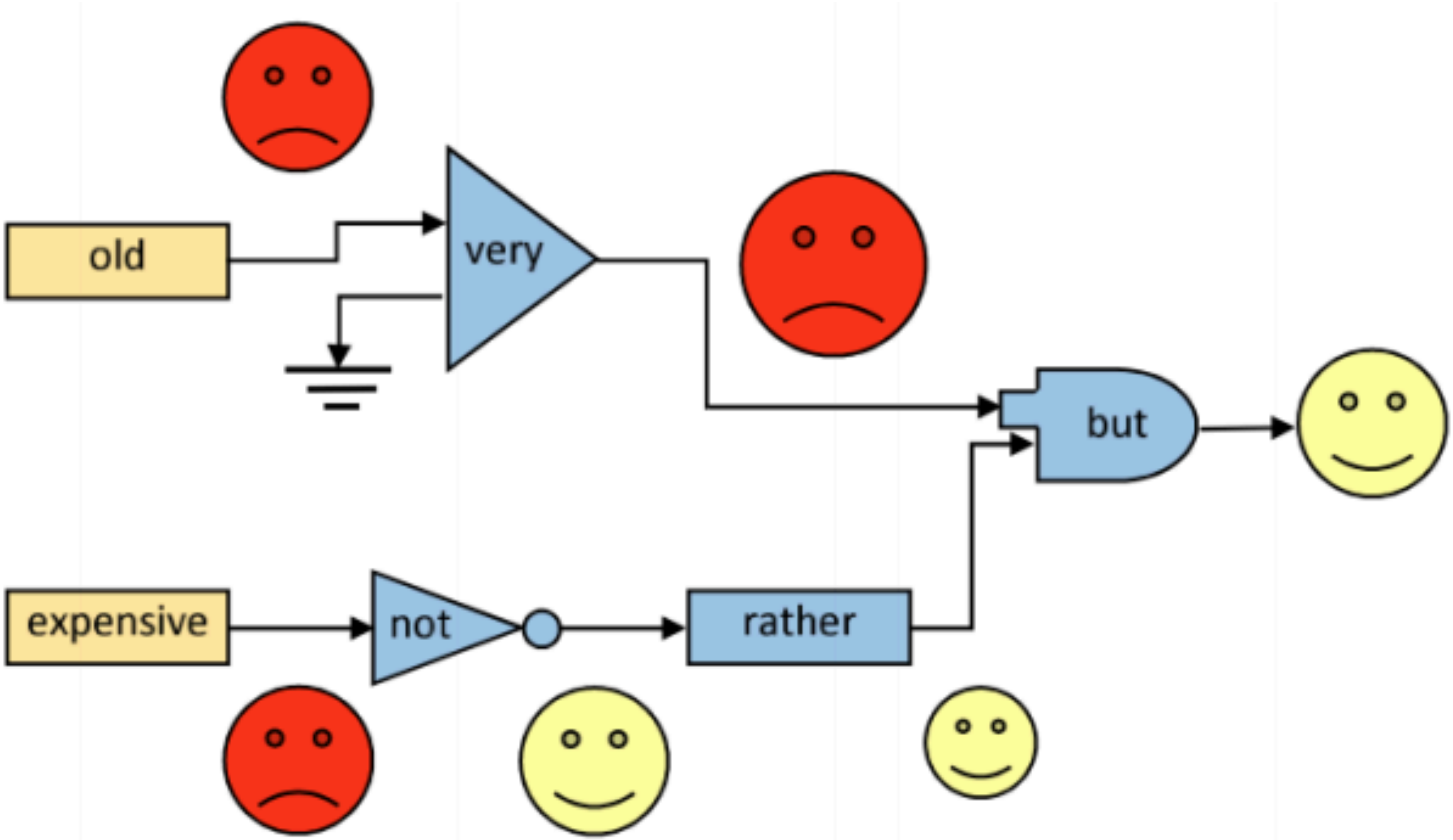
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Evaluation of Text Mining and Sentiment Analysis

- Evaluation of Information Retrieval
- Evaluation of Classification Model (Prediction)
  - Accuracy
  - Precision
  - Recall
  - F-score

Deep Learning for Sentiment Analytics

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts
Stanford University, Stanford, CA 94305, USA
richard@socher.org,{aperelyg,jcchuang,ang}@cs.stanford.edu
{jeaneis,manning,cgpotts}@stanford.edu

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model out-

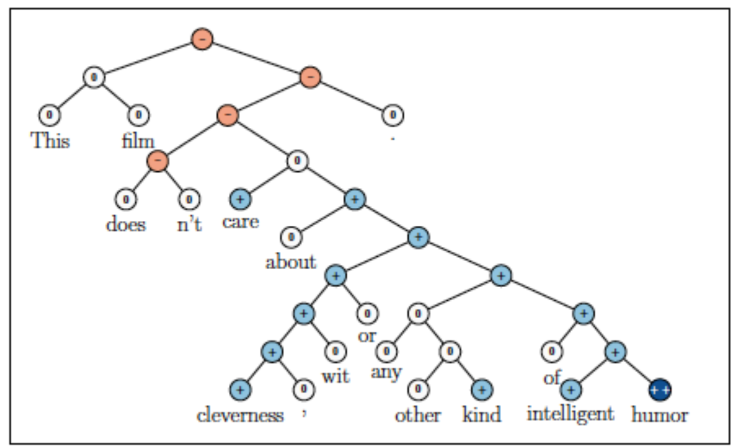
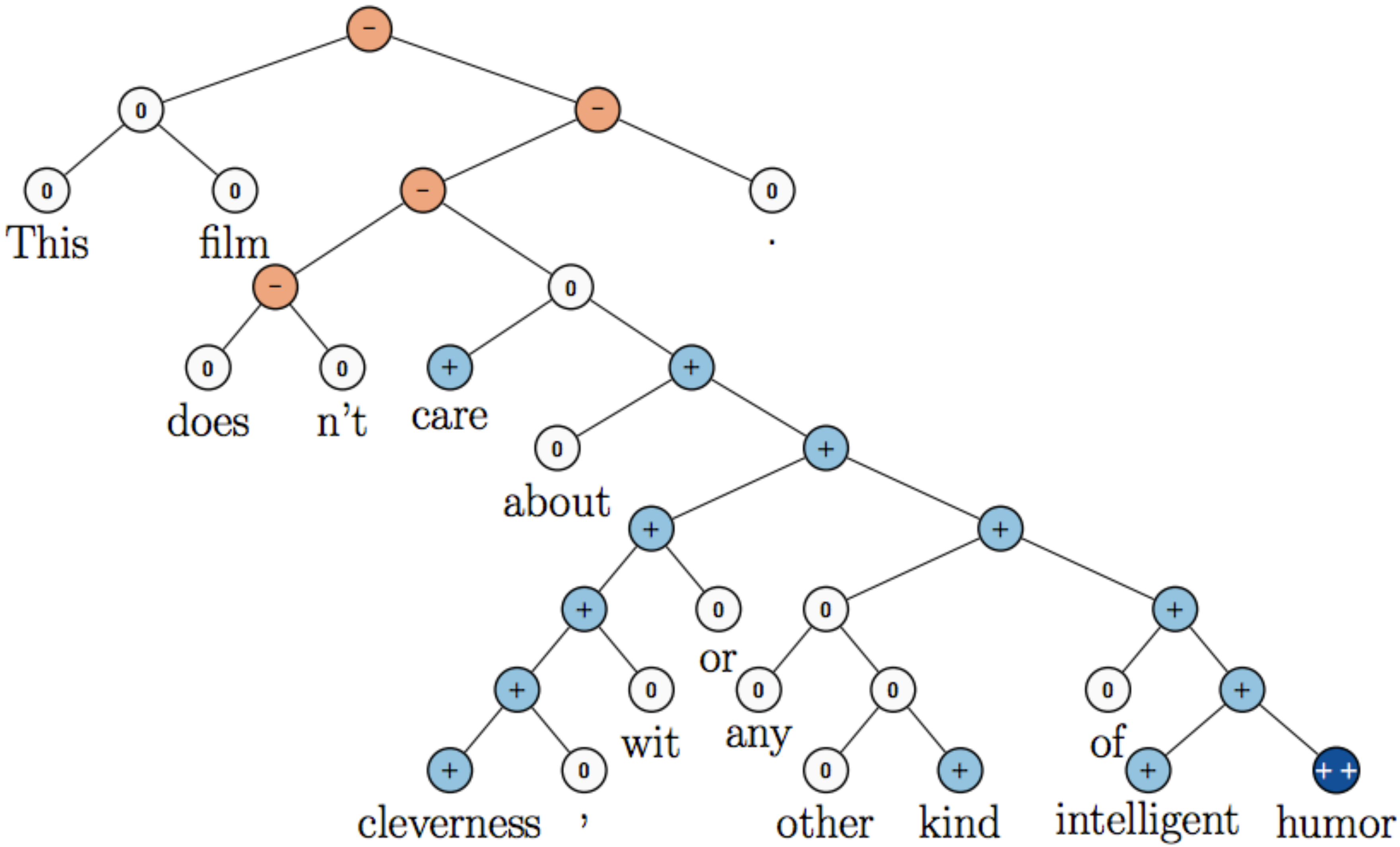


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

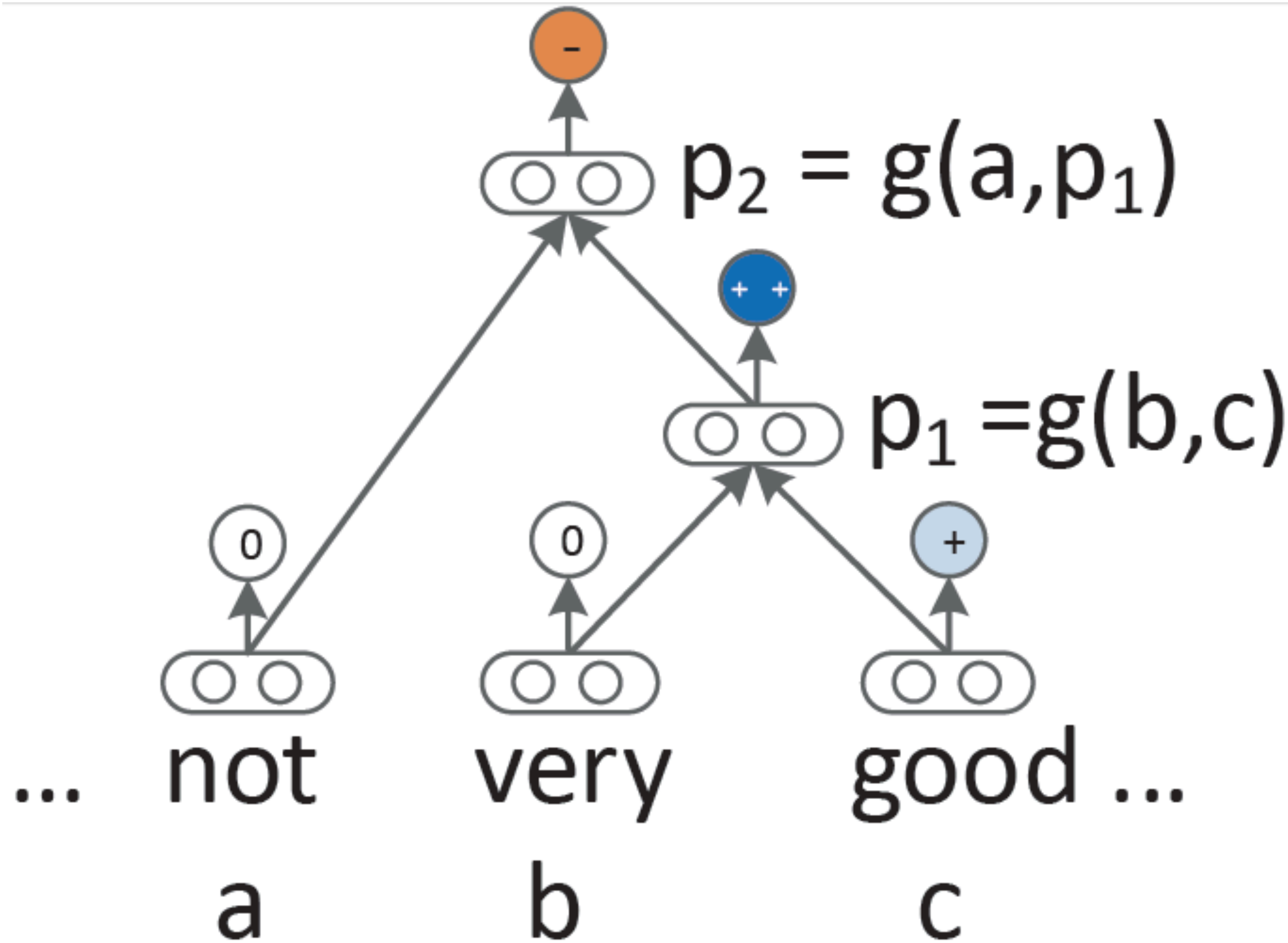
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



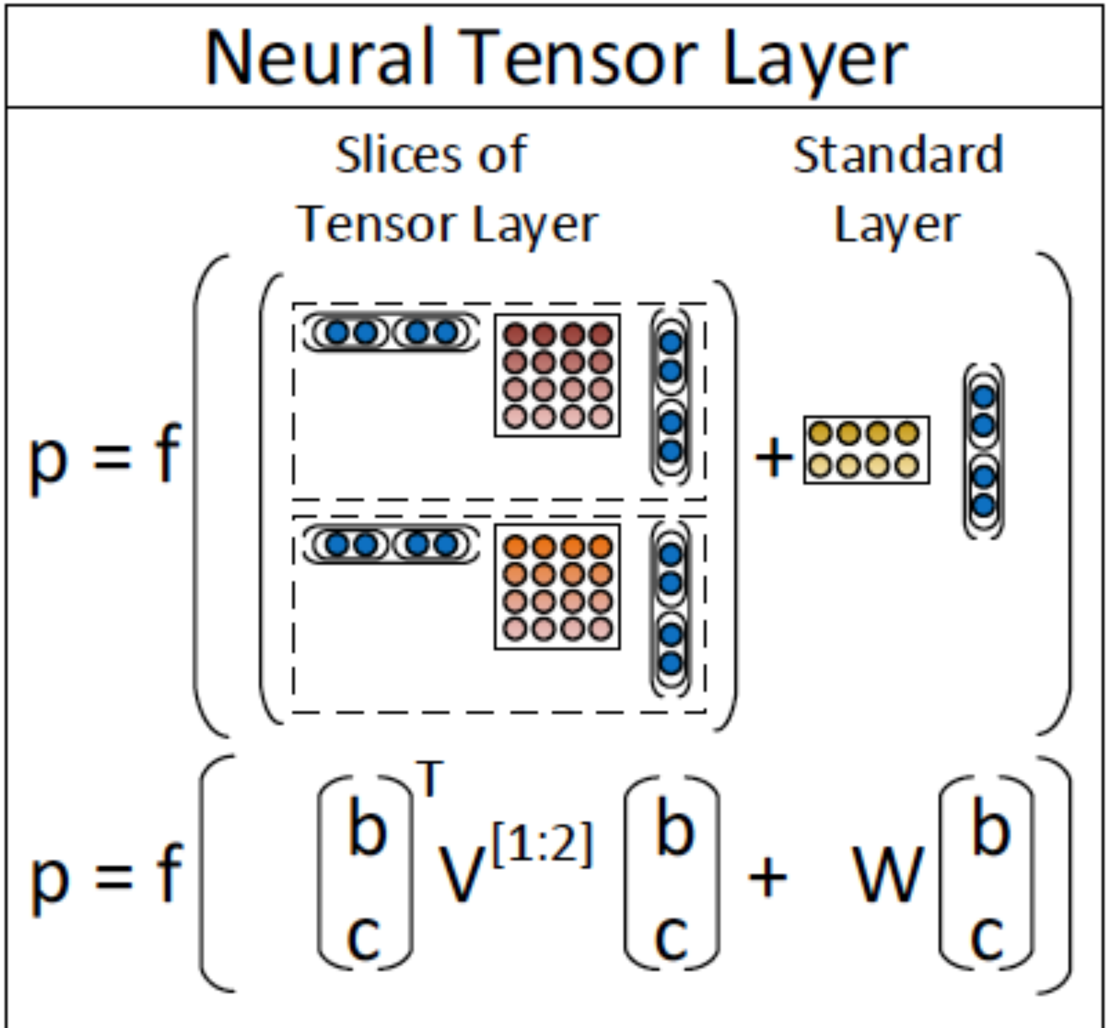
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Network (RNN) models for sentiment



Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



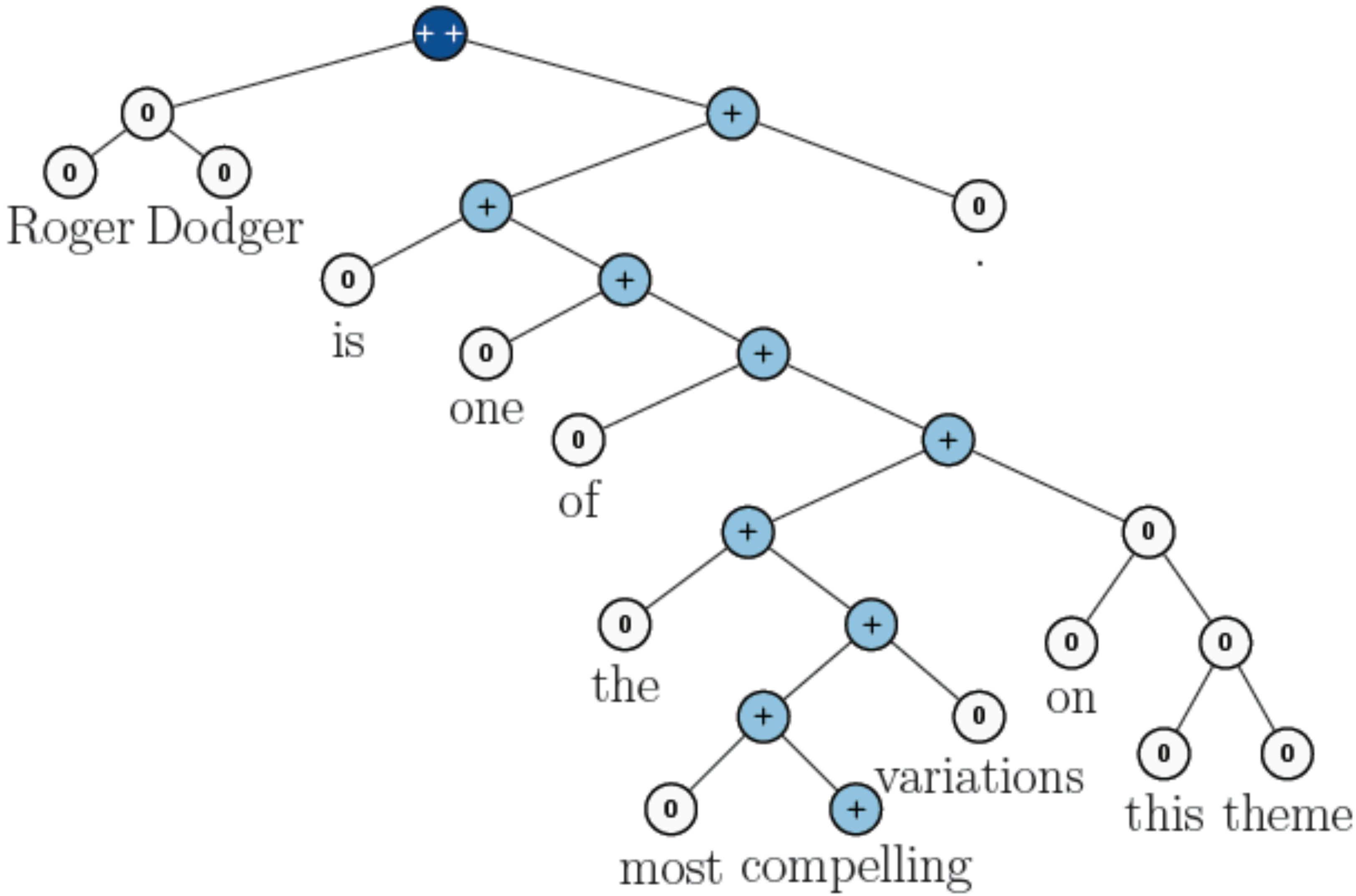
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

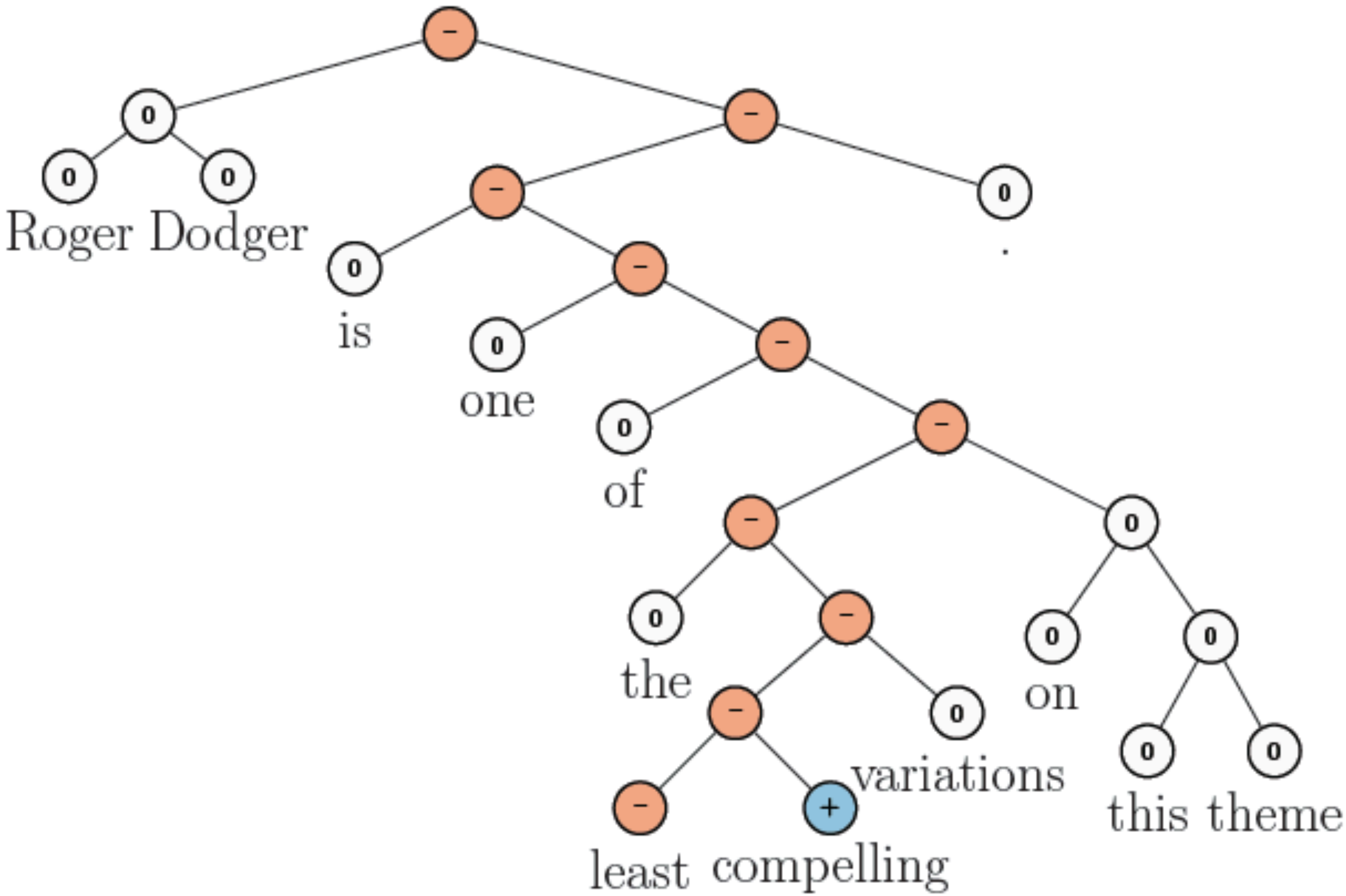
RNTN for Sentiment Analysis



Roger Dodger is one of the most compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

RNTN for Sentiment Analysis



Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes

Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

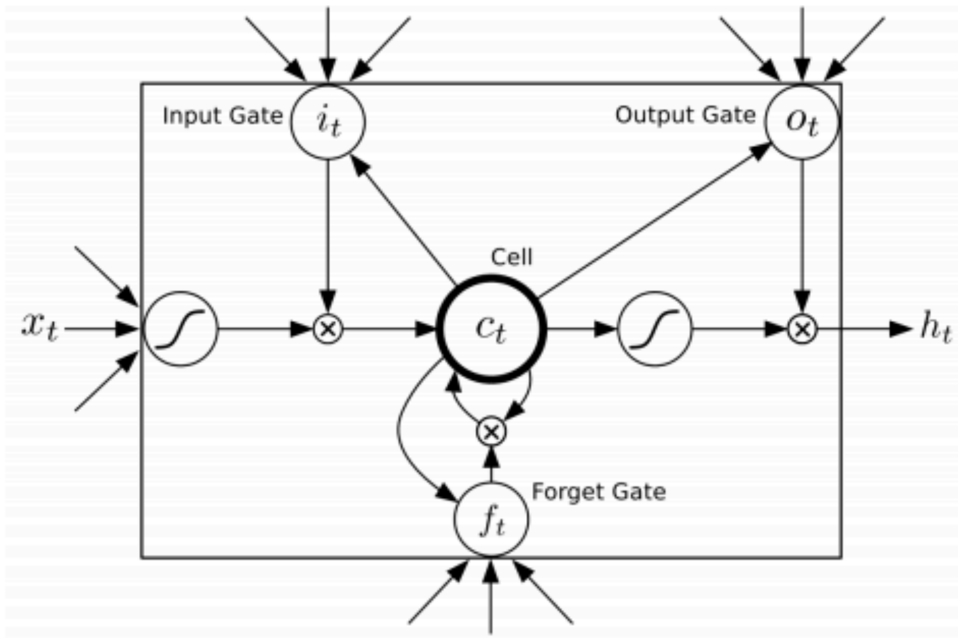
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

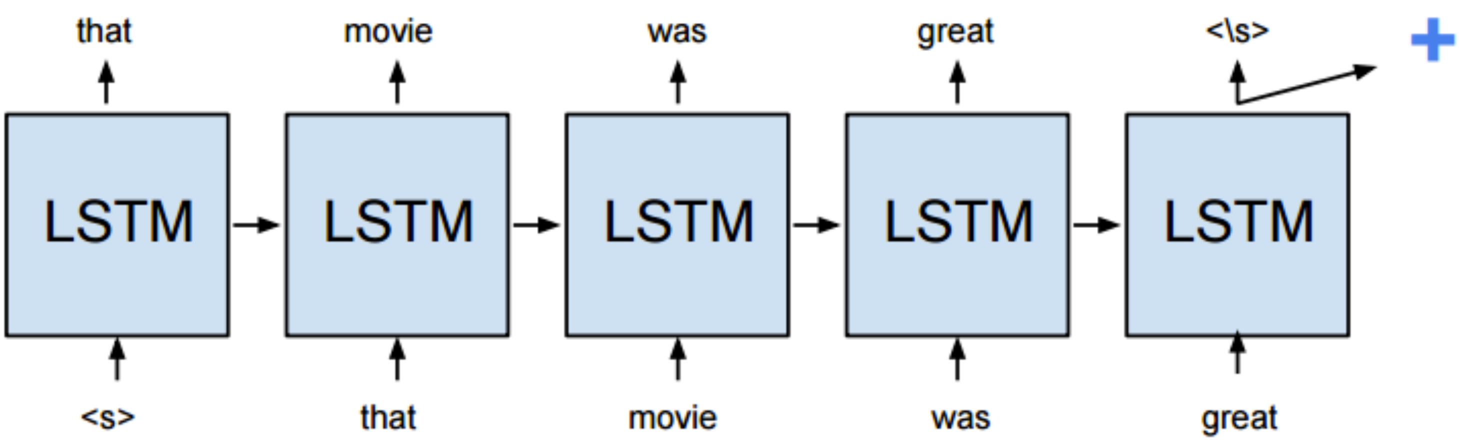
Accuracy of negation detection

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Long Short-Term Memory (LSTM)





Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Deep Learning for Sentiment Analysis CNN RNTN LSTM

Model	Fine (5-class)	Binary
DCNN (Blunsom, et al. 2014)	0.485	0.868
RNTN (Socher, et al. 2013)	0.457	0.854
CNN-non-static (Kim, 2014)	0.480	0.872
CNN-multi-channel (Kim, 2014)	0.474	0.881
DRNN w. pretrained word-embeddings (Irsoy and Cardie, 2014)	0.498	0.866
Paragraph Vector (Le and Mikolov. 2014)	0.487	0.878
Dependency Tree-LSTM (Tai, et al, 2015)	0.484	0.857
Constituency Tree-LSTM (Tai, et al, 2015)	0.439	0.820
Constituency Tree-LSTM (Glove vectors) (Tai, et al, 2015)	0.510	0.880
Paragraph Vector	0.391	0.798
LSTM	0.456	0.843
Deep Recursive-NN	0.469	0.847

Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Performance Comparison of Sentiment Analysis Methods

	Method	Data Set	Acc.	Author
Machine Learning	SVM	Movie reviews	86.40%	Pang, Lee[23]
	CoTraining SVM	Twitter	82.52%	Liu[14]
	Deep learning	Stanford Sentiment Treebank	80.70%	Richard[18]
Lexical based	Corpus	Product reviews	74.00%	Turkey
	Dictionary	Amazon's Mechanical Turk	---	Taboada[20]
Cross-lingual	Ensemble	Amazon	81.00%	Wan,X[16]
	Co-Train	Amazon, ITI68	81.30%	Wan,X.[16]
	EWGA	IMDb movie review	>90%	Abbasi,A.
	CLMM	MPQA,NTCIR,ISI	83.02%	Mengi
Cross-domain	Active Learning	Book, DVD, Electronics, Kitchen	80% (avg)	Li, S
	Thesaurus			Bollegala[22]
	SFA			Pan S J[15]

Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Resources of Opinion Mining

Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - http://ir.dcs.gla.ac.uk/test collections/access to data.html
- Cornell movie-review datasets
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/
- Customer review datasets
  - http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip
- Multiple-aspect restaurant reviews
  - http://people.csail.mit.edu/bsnyder/naacl07
- NTCIR multilingual corpus
  - NTCIR Multilingual Opinion-Analysis Task (MOAT)

Source: Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval

Lexical Resources of Opinion Mining

- SentiWordnet
  - http://sentiwordnet.isti.cnr.it/
- General Inquirer
  - http://www.wjh.harvard.edu/~inquirer/
- OpinionFinder's Subjectivity Lexicon
  - http://www.cs.pitt.edu/mpqa/
- NTU Sentiment Dictionary (NTUSD)
  - http://nlg18.csie.ntu.edu.tw:8080/opinion/
- Hownet Sentiment
  - http://www.keenage.com/html/c_bulletin_2007.htm

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

《知網》情感分析用詞語集（beta版）

- “中英文情感分析用詞語集”
  - 包含詞語約 17887
- “中文情感分析用詞語集”
  - 包含詞語約 9193
- “英文情感分析用詞語集”
  - 包含詞語 8945

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

中文正面情感詞語	836
中文負面情感詞語	1254
中文正面評價詞語	3730
中文負面評價詞語	3116
中文程度級別詞語	219
中文主張詞語	38
Total	9193

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面情感”詞語
  - 如：
    愛，讚賞，快樂，感同身受，好奇，喝彩，魂牽夢縈，嘉許 ...
- “負面情感”詞語
  - 如：
    哀傷，半信半疑，鄙視，不滿意，不是滋味兒，後悔，大失所望 ...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面評價”詞語
  - 如：
    不可或缺，部優，才高八斗，沉魚落雁，催人奮進，動聽，對勁兒 ...
- “負面評價”詞語
  - 如：
    醜，苦，超標，華而不實，荒涼，混濁，畸輕畸重，價高，空洞無物 ...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “程度級別” 詞語
  - 1. “極其|extreme / 最|most”
    - 非常，極，極度，無以倫比，最為
  - 2. “很|very”
    - 多麼，分外，格外，著實
  - …
- “主張” 詞語
  - 1. {perception|感知}
    - 感覺，覺得，預感
  - 2. {regard|認為}
    - 認為，以為，主張

Source: http://www.keenage.com/html/c_bulletin_2007.htm

References

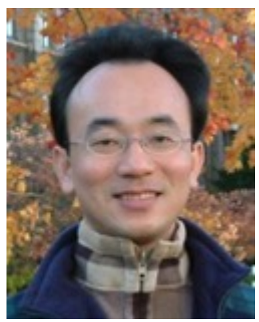
- Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” 2nd Edition, Springer. http://www.cs.uic.edu/~liub/WebMiningBook.html
- Bing Liu (2013), Opinion Spam Detection: Detecting Fake Reviews and Reviewers, http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
- Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.
- Wiltrud Kessler (2012), Introduction to Sentiment Analysis, http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf
- Z. Zhang, X. Li, and Y. Chen (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.
- Efraim Turban, Ramesh Sharda, Dursun Delen (2011), Decision Support and Business Intelligence Systems, Ninth Edition, 2011, Pearson.
- Guandong Xu, Yanchun Zhang, Lin Li (2011), Web Mining and Social Networking: Techniques and Applications, 2011, Springer

References

- Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In the 26th International Conference on Computational Linguistics (COLING), Osaka. 2016.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts (2013), "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," In Proceedings of the conference on empirical methods in natural language processing (EMNLP), vol. 1631, p. 1642 http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf
- Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.
- Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol 139, no. 11, 2016. pp.5-15.
- Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.
- Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page
- Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press

Tamkang University

Deep Learning for Financial Sentiment Analysis on Finance News Providers




Min-Yuh Day Chia-Chou Lee

Department of Information Management

Tamkang University, Taiwan

myday@mail.tku.edu.tw

The 7th International Workshop on Mining and Analyzing Social Networks for Decision Support (MSNDS 2016)

MSNDS 2016, IEEE/ACM ASONAM 2016, San Francisco, CA, USA, August 18-21, 2016

Outline

- Introduction
- System Architecture
- Experimental Results and Discussion
- Conclusion

Introduction

Financial Sentiment Analysis

Motivation

- Investors have always been interested in stock price forecasting.
- The rapid development of electronic media, the big data of financial news are released on different media every day.

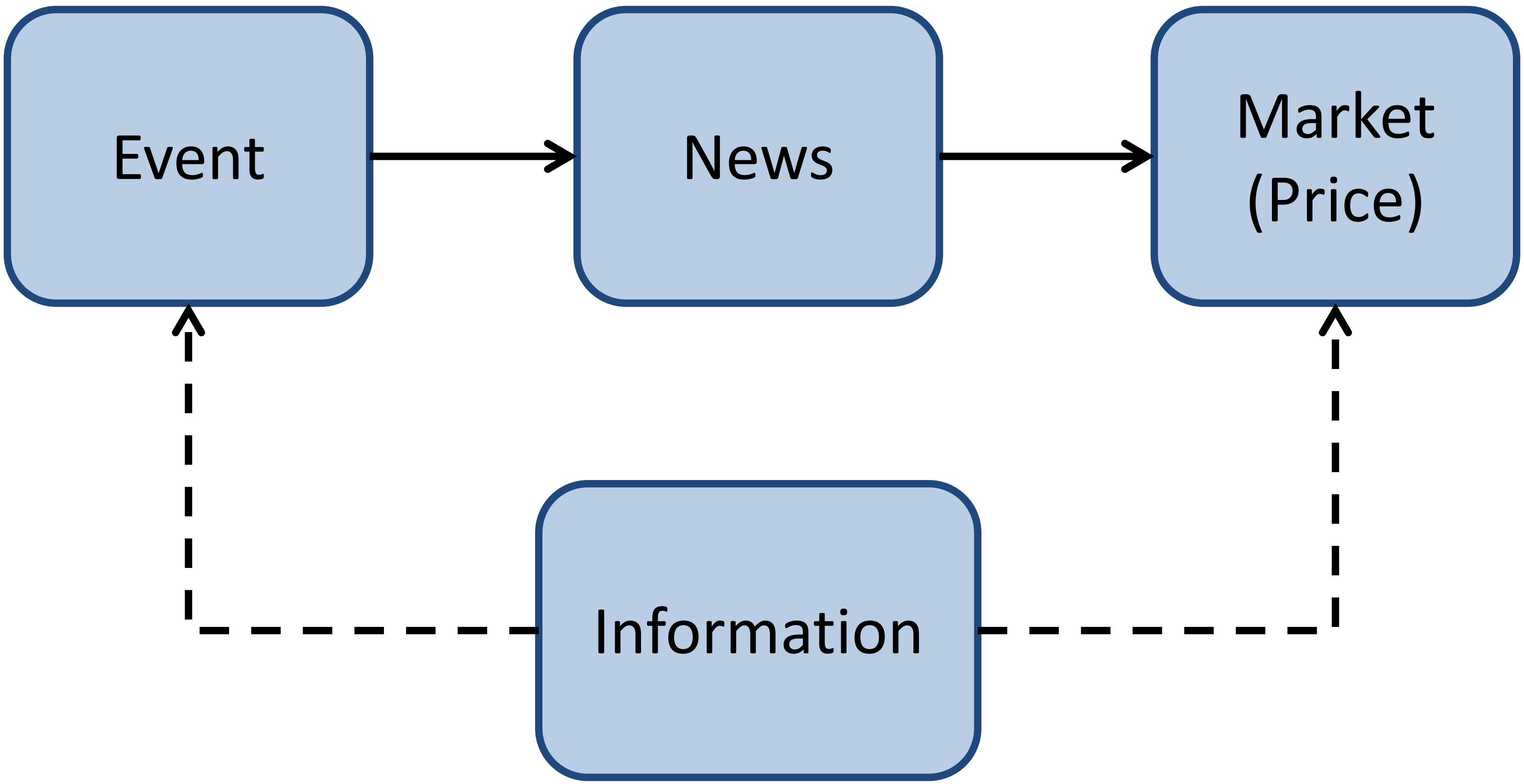
Research Gap

- Few research involved the discussion on whether using different media could affect forecasting results.
- Financial sentiment analysis is an important research area of financial technology (FinTech).

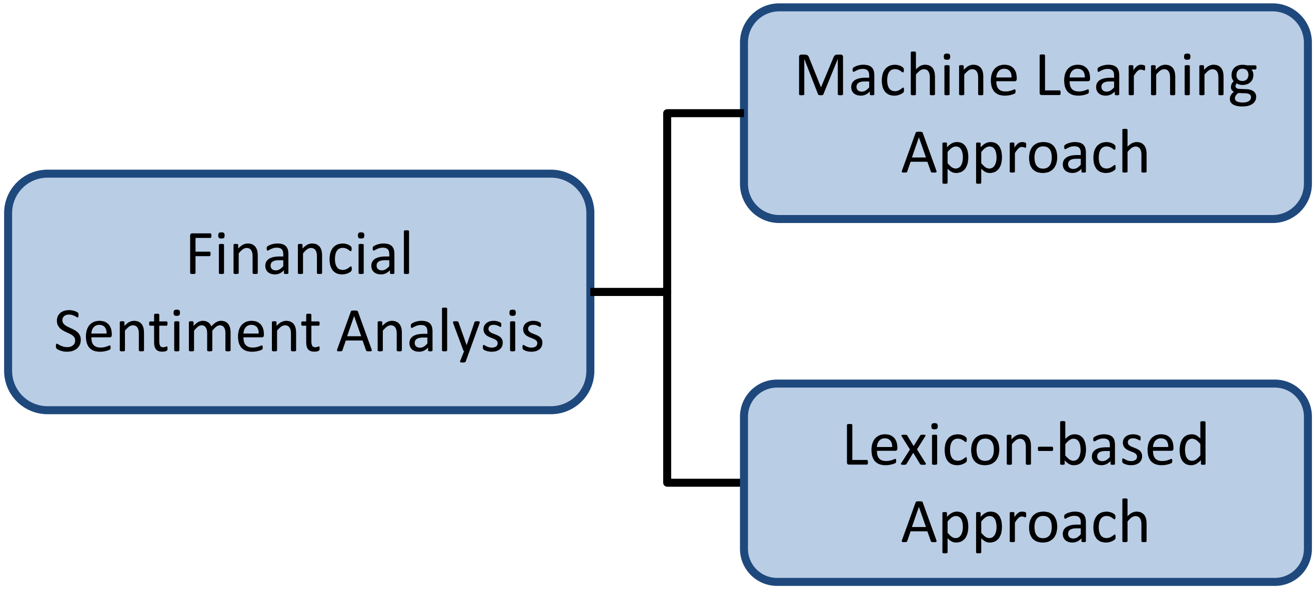
Highlights

- This research focuses on investigating the influence of using different financial resources to investment and how to improve the accuracy of forecasting through deep learning.
- The experimental result shows various financial resources have significantly different effects to investors and their investments, while the accuracy of news categorization could be improved through deep learning.

The relationship between Events, News and Markets (price) through Information.

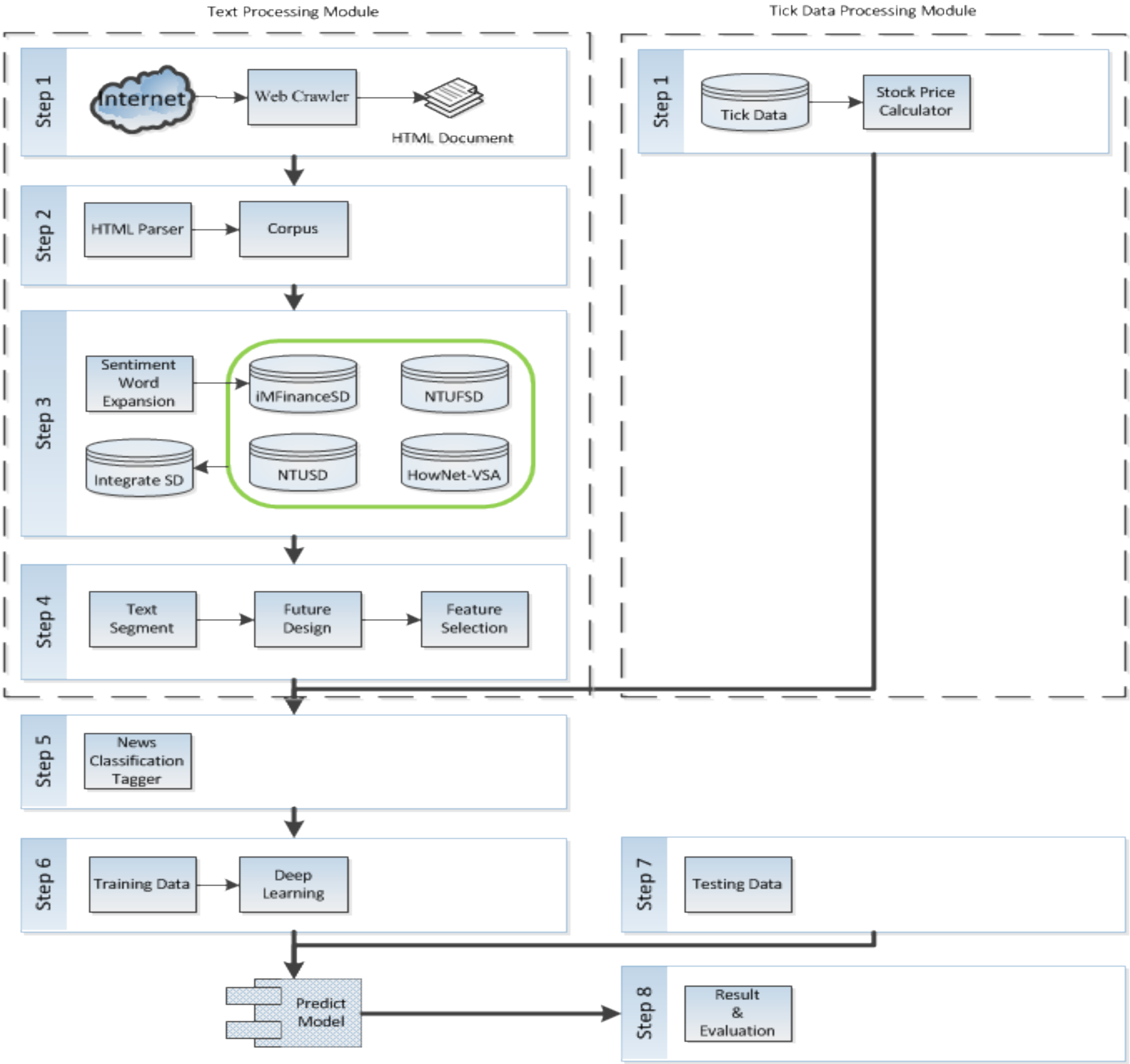


Financial Sentiment Analysis

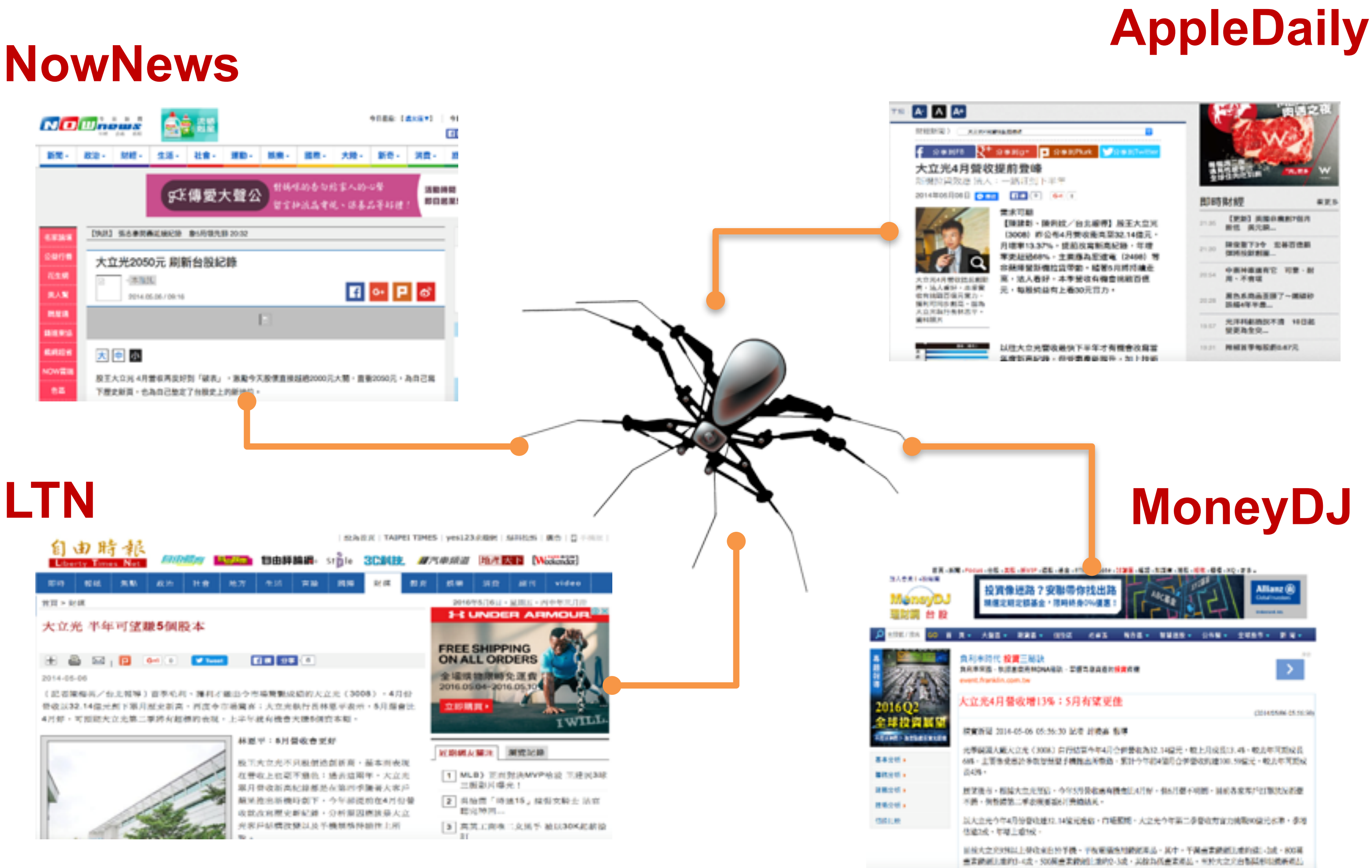


System Architecture

System Architecture



Finance News Data



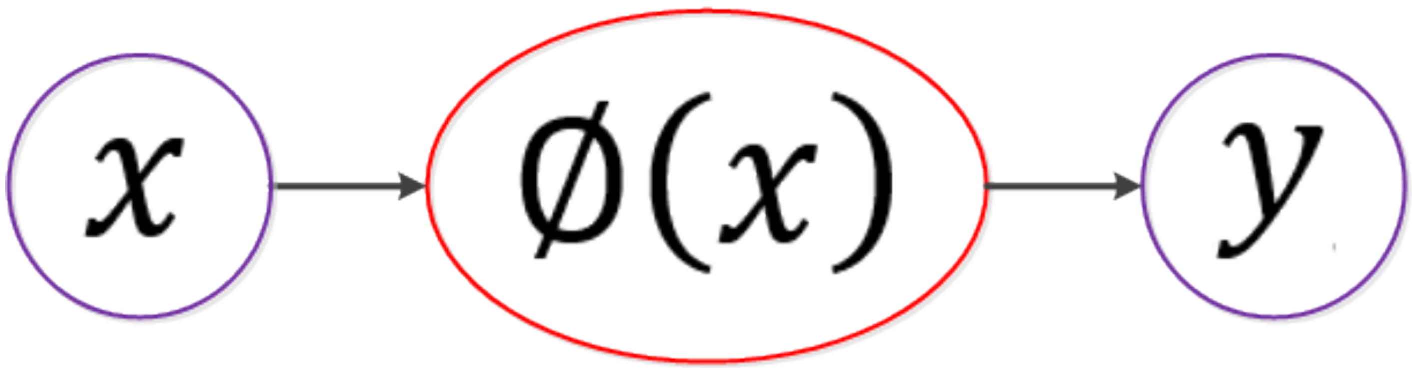
Feature used for financial sentiment analysis

ID	Future Name	Description
F01	NewsCharacters	Total word number of news text
F02	NewsTokens	Number of news words
F03	NTUSD_Positive	NTUSD positive word
F04	NTUSD_Negative	NTUSD negative word
F05	NTUSD_PNDiff	NTUSD difference of positive and negative word
F06	HowNet_Positive	HowNet positive sentiment word
F07	HowNet_Negative	HowNet negative sentiment word
F08	HowNet_PNDiff	HowNet difference of positive and negative word
F09	FinanceSD_Positive	NTUFSD+iMFinanceSD positive word
F10	FinanceSD_Negative	NTUFSD+iMFinanceSD negative word
F11	FinanceSD_PNDiff	NTUFSD+iMFinanceSD difference of positive and negative word

Example of feature used for Financial Sentiment Analysis

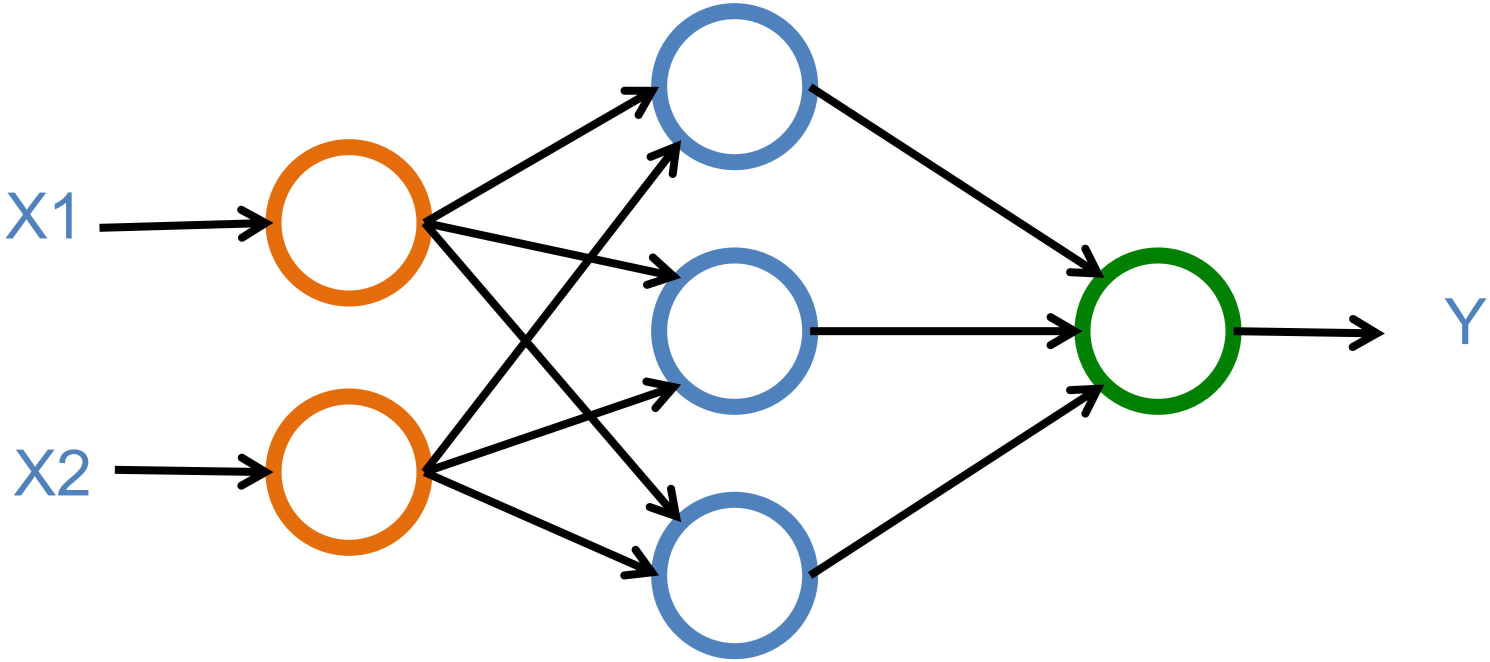
D00001, 516, 185, 3, 8, -5, 0, 1, -1, 14, 8, 6, 3008, 2013-01-04
D00002, 534, 185, 4, 2, 2, 0, 0, 0, 13, 2, 11, 3008, 2013-01-07
D00003, 846, 296, 6, 9, -3, 0, 1, -1, 25, 15, 10, 3008, 2013-01-09
D00004, 1495, 489, 55, 19, 36, 3, 0, 3, 23, 16, 7, 3008, 2013-01-10
D00005, 872, 282, 3, 5, -2, 0, 0, 0, 12, 17, -5, 3008, 2013-01-14
D00006, 573, 183, 2, 5, -3, 0, 0, 0, 8, 9, -1, 3008, 2013-01-21

Deep Learning and Neural Network

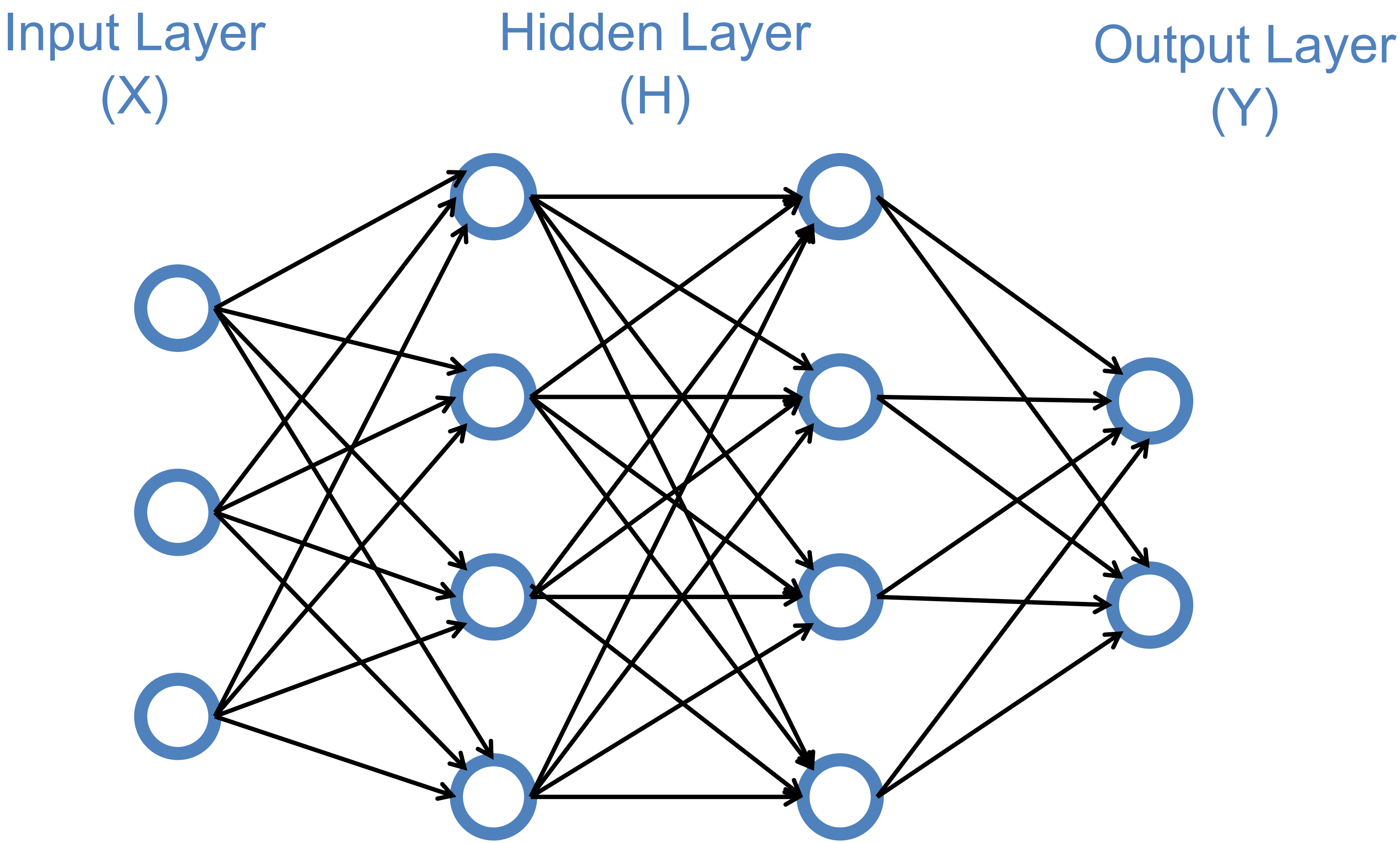


Deep Learning and Neural Networks

Input Layer (X) Hidden Layer (H) **Output Layer (Y)**



Deep Learning and Neural Networks



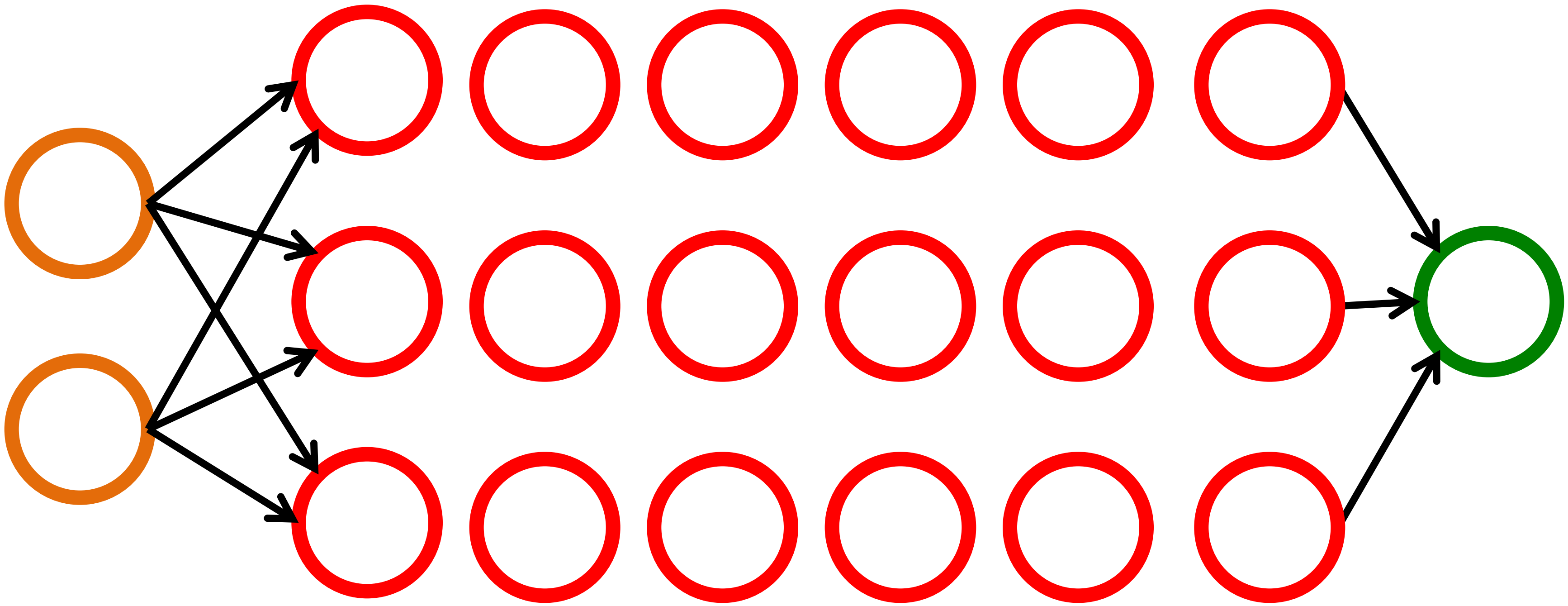
Deep Learning and Neural Networks

Input Layer
(X)

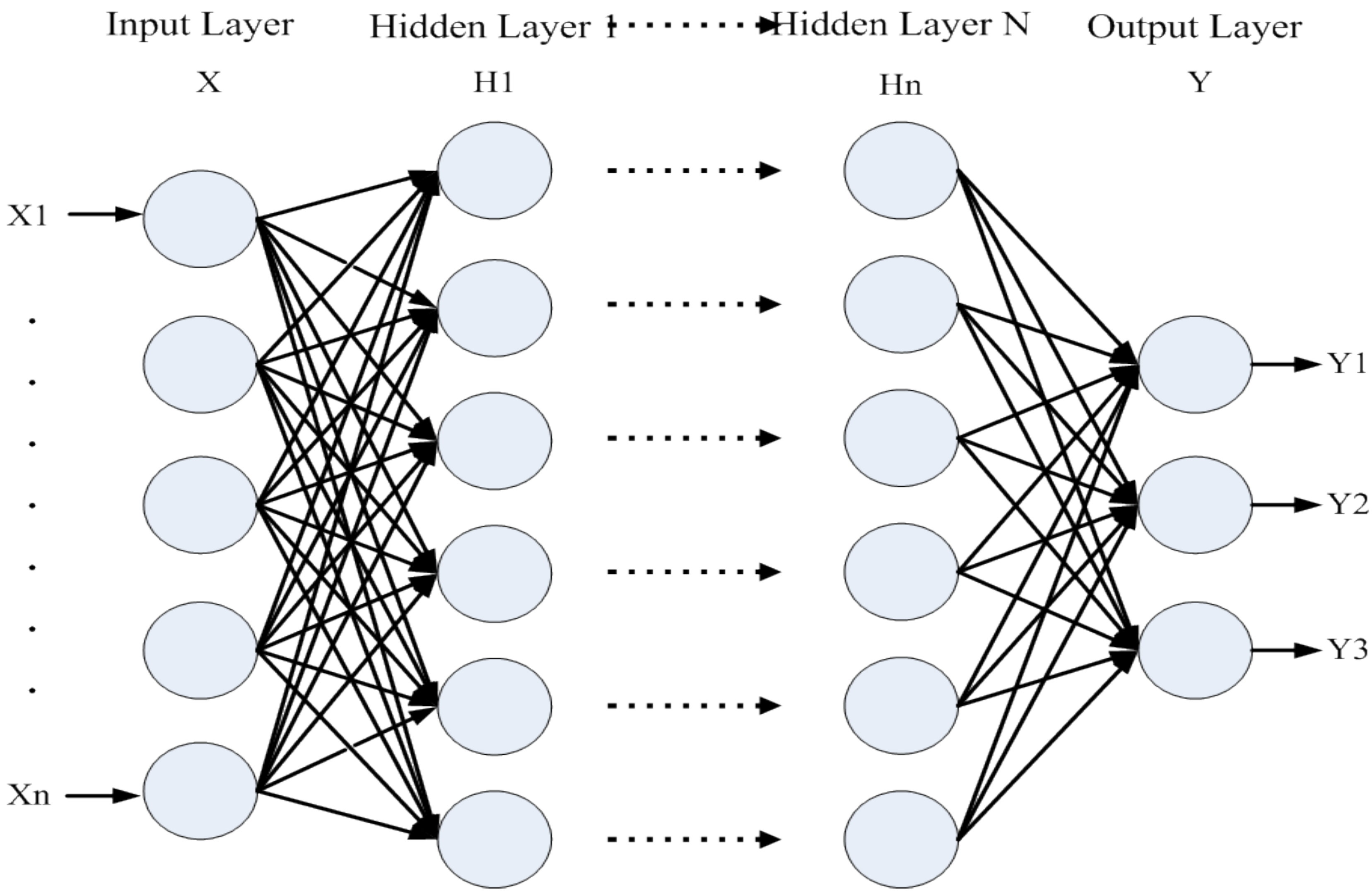
Hidden Layers
(H)

Output Layer
(Y)

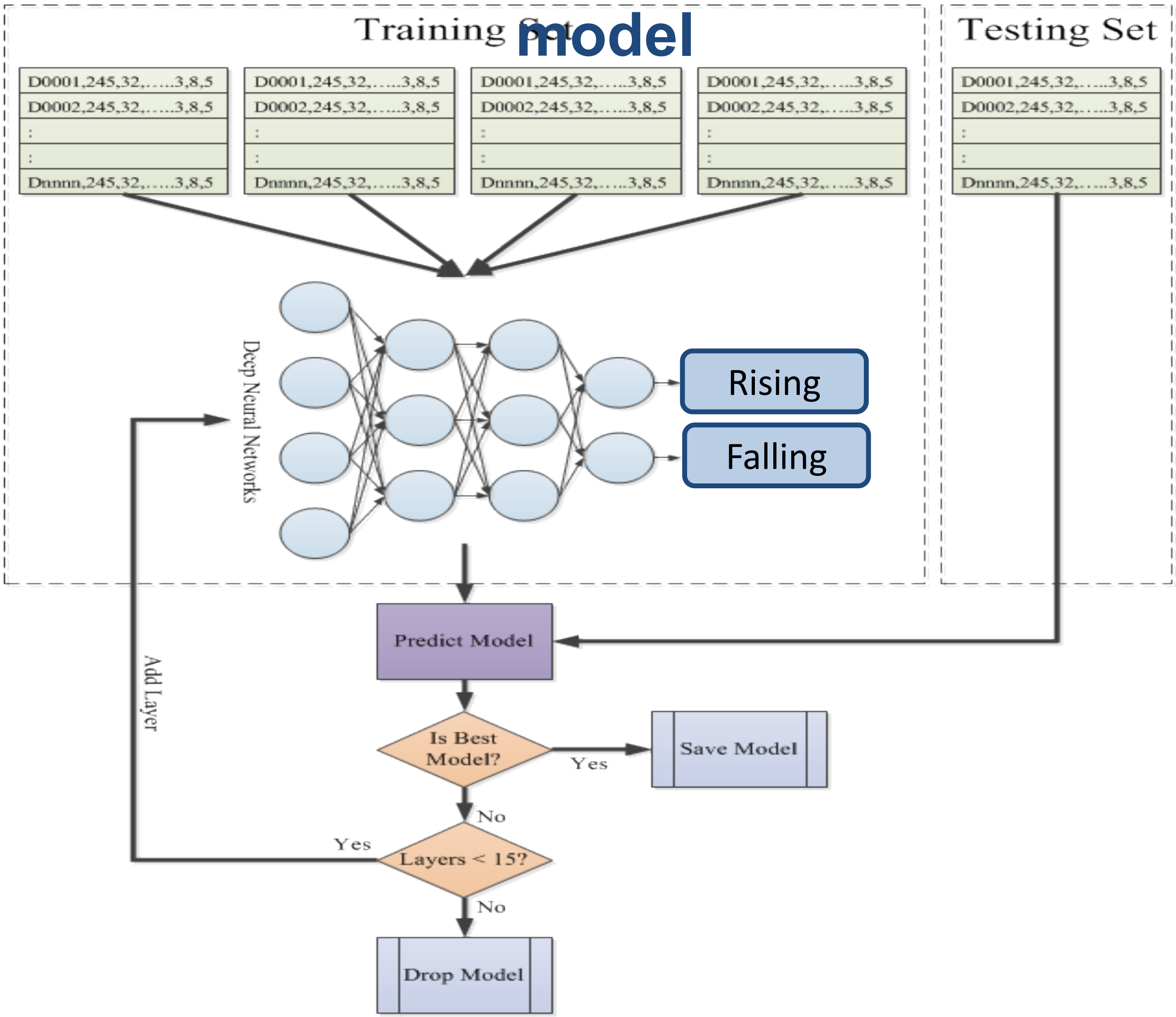
Deep Neural Networks
Deep Learning



Deep Neural Networks



Construction flow chart of deep learning prediction model



Comparison of editorial team and contents of news providers

News providers	Platform	Content	Editorial team
NowsNews	Electronic media	Comprehensive	Owned
AppleDaily	Electronic media/newspaper	Comprehensive	Owned
LTN	Electronic media/newspaper	Comprehensive	Owned
MoneyDJ	Electronic media	Finance	Owned

Developing Finance Sentiment Dictionary

- NTUSD
- HowNet-VSA
- NTUFSD
- iMFinanceSD

Sample iMFinanceSD opinion words

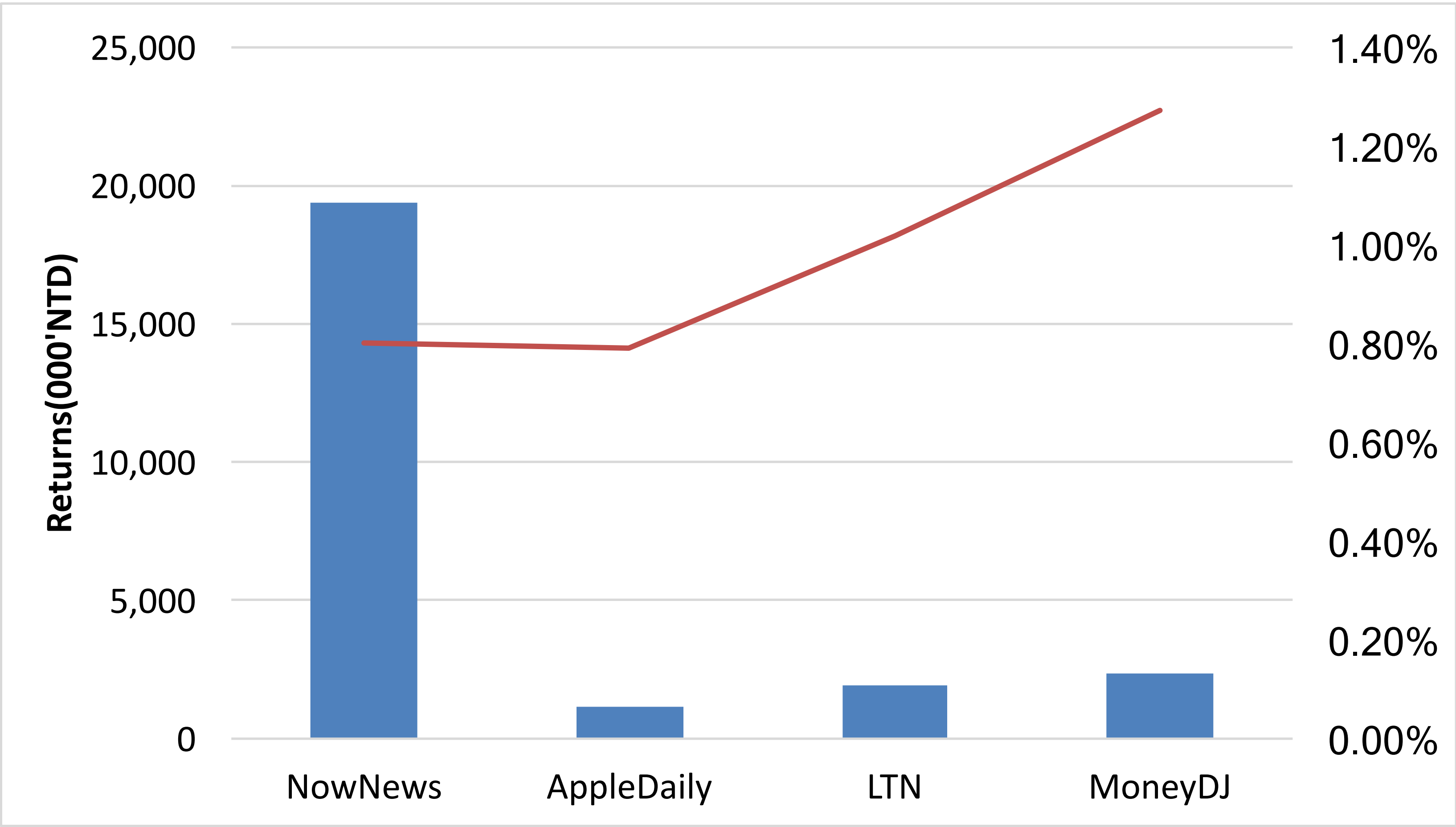
Positive		Negative	
No	Word	No	Word
1	衝破 (break through)	1	震盪 (shock)
2	穩住 (stabilize)	2	低於 (less then)
3	挹注 (inject)	3	放緩 (slowdown)
4	歷史新高 (Historical high)	4	走緩 (slowly)
5	不遑多讓 (be no slouch)	5	大幅砍殺 (Stabbed sharply)

Experimental Results and Discussion

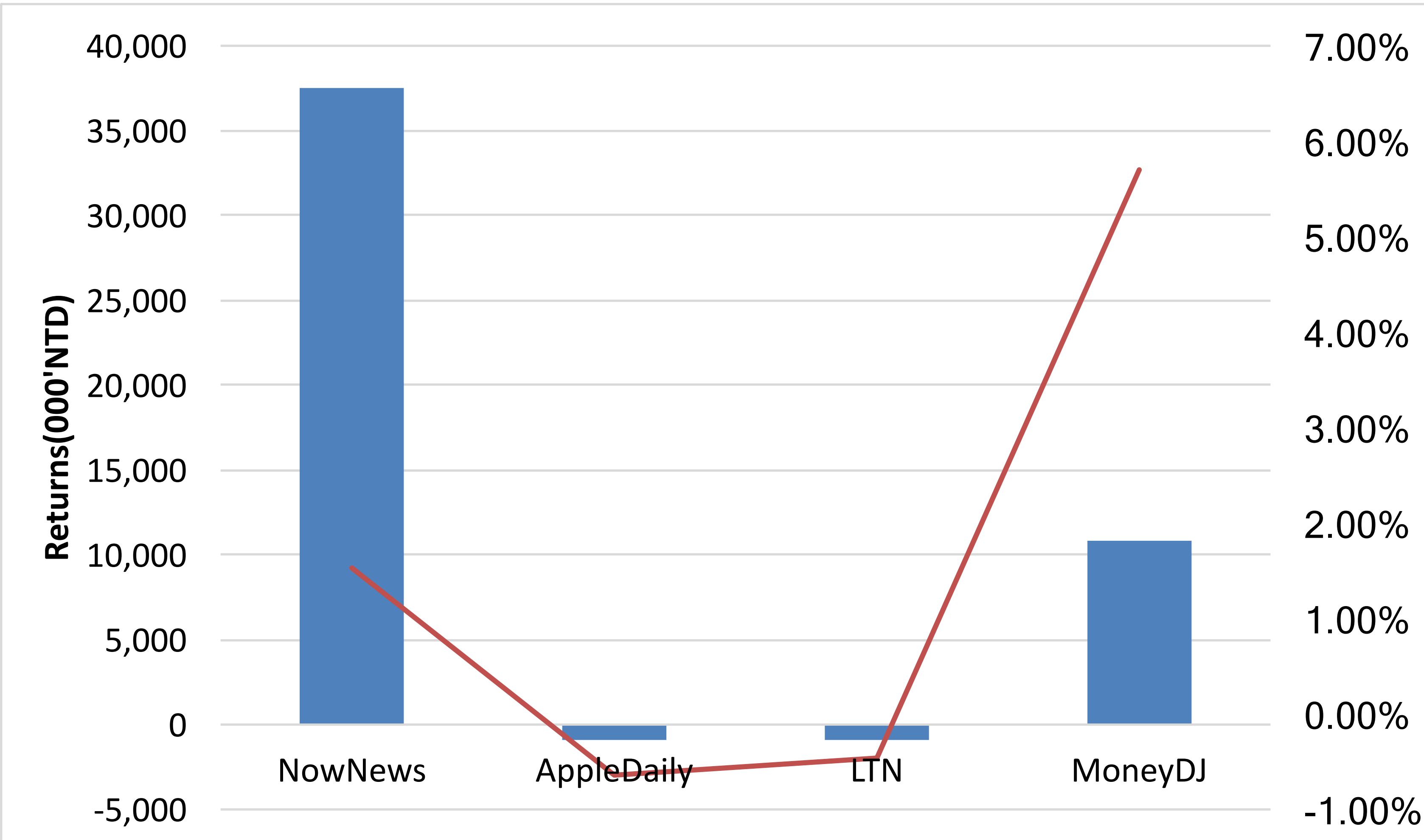
Total news of each news provider for Financial Sentiment Analysis (Text Data)

Provider	NowNews	AppleDaily	LTN	MoneyDJ
Total	5,499	456	1,147	1,370

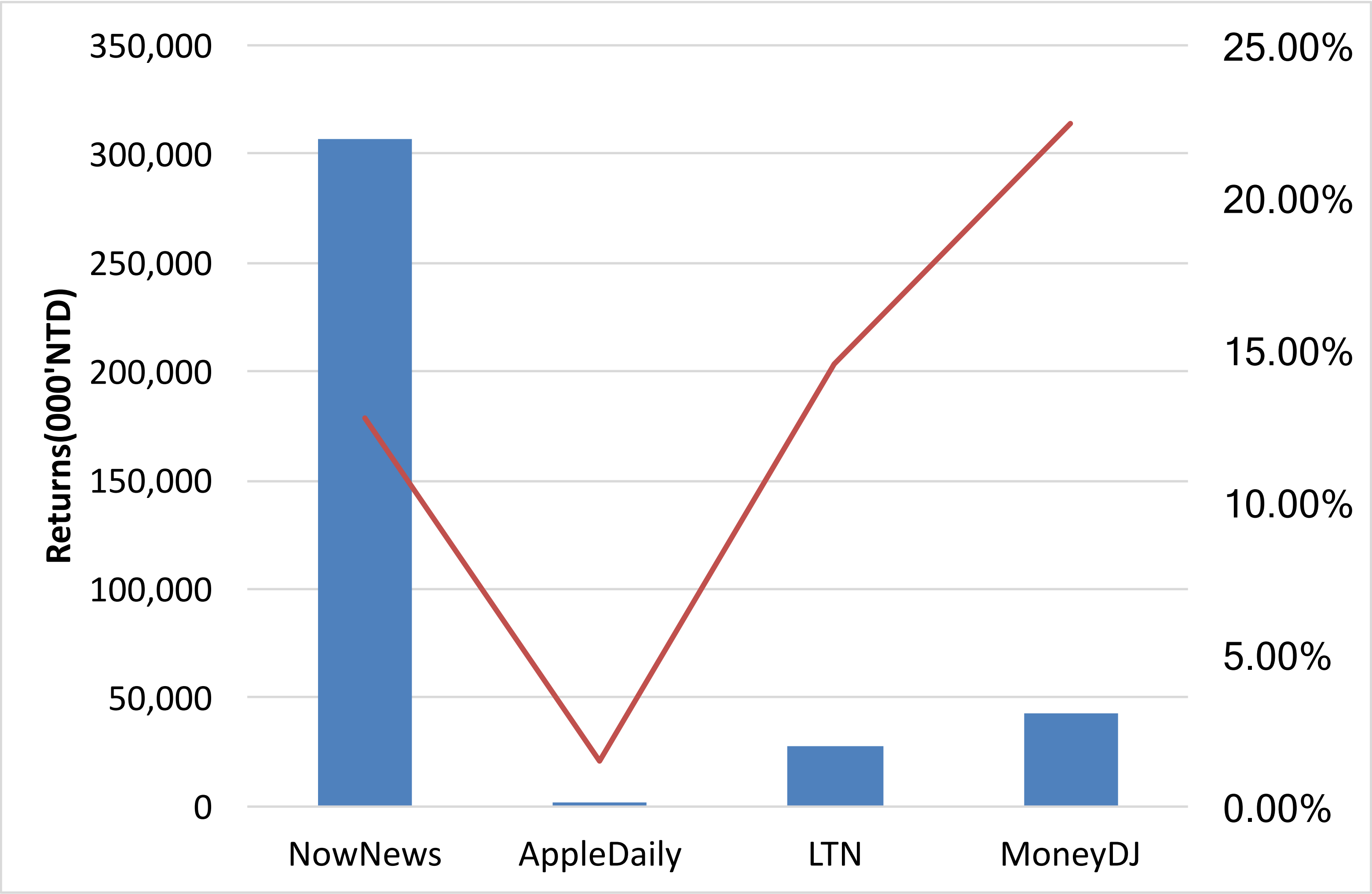
ROI of 5 days trading with Deep Learning predicting stock price trend



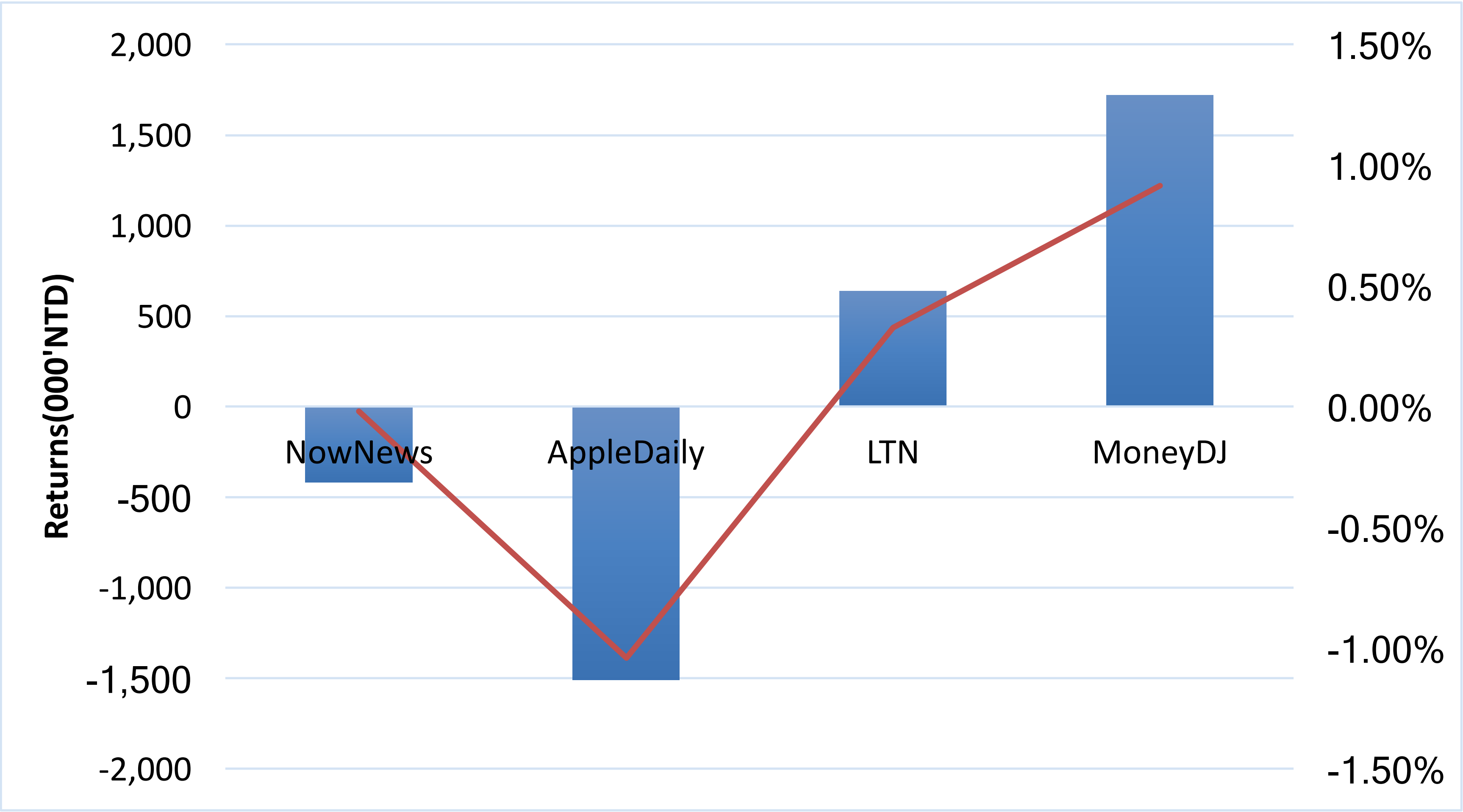
ROI of 20 days trading with Deep Learning predicting stock price trend



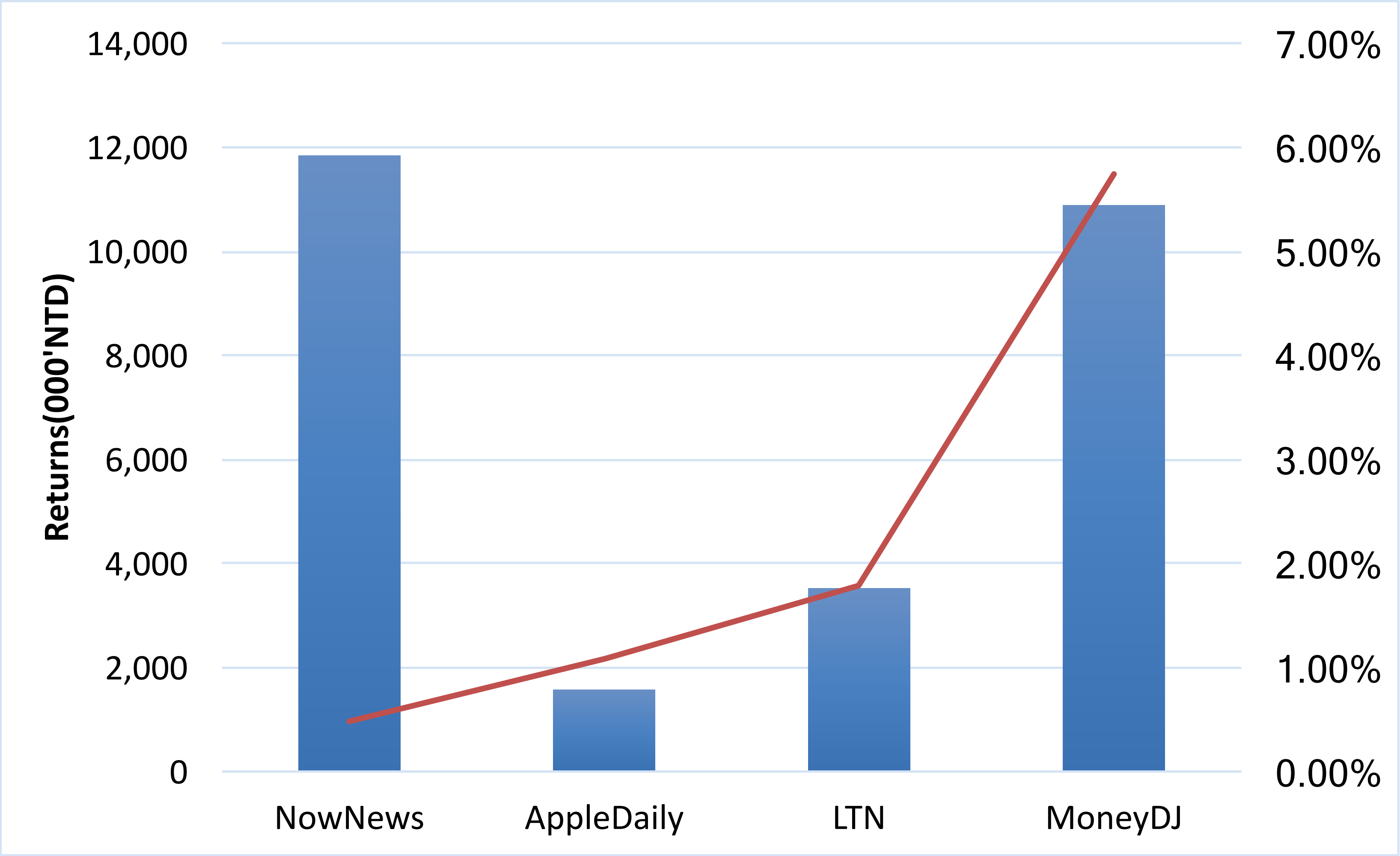
ROI of 60 days trading with Deep Learning predicting stock price trend



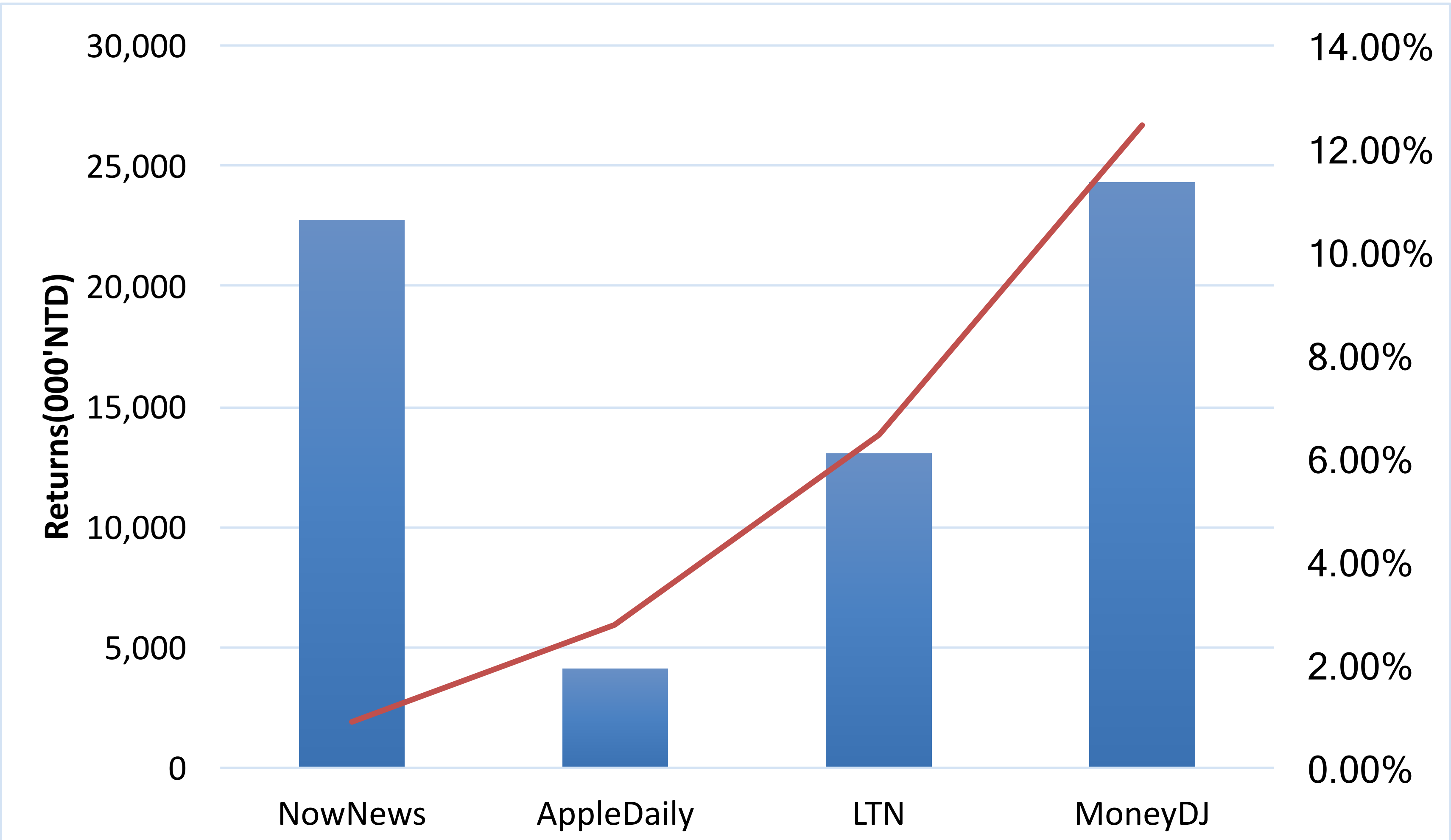
ROI of 5 days with Lexicon-based Trading



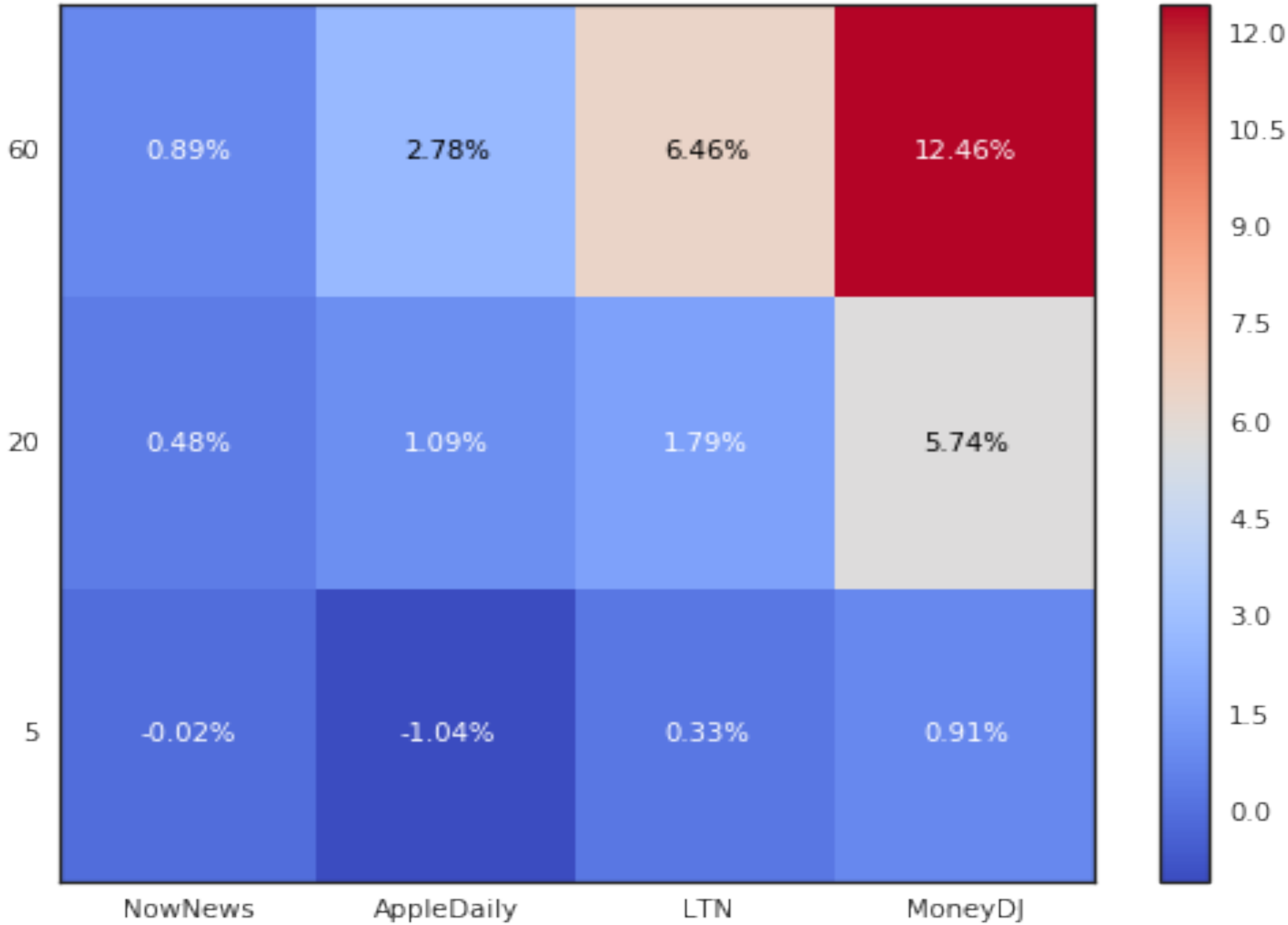
ROI of 20 days with Lexicon-based Trading



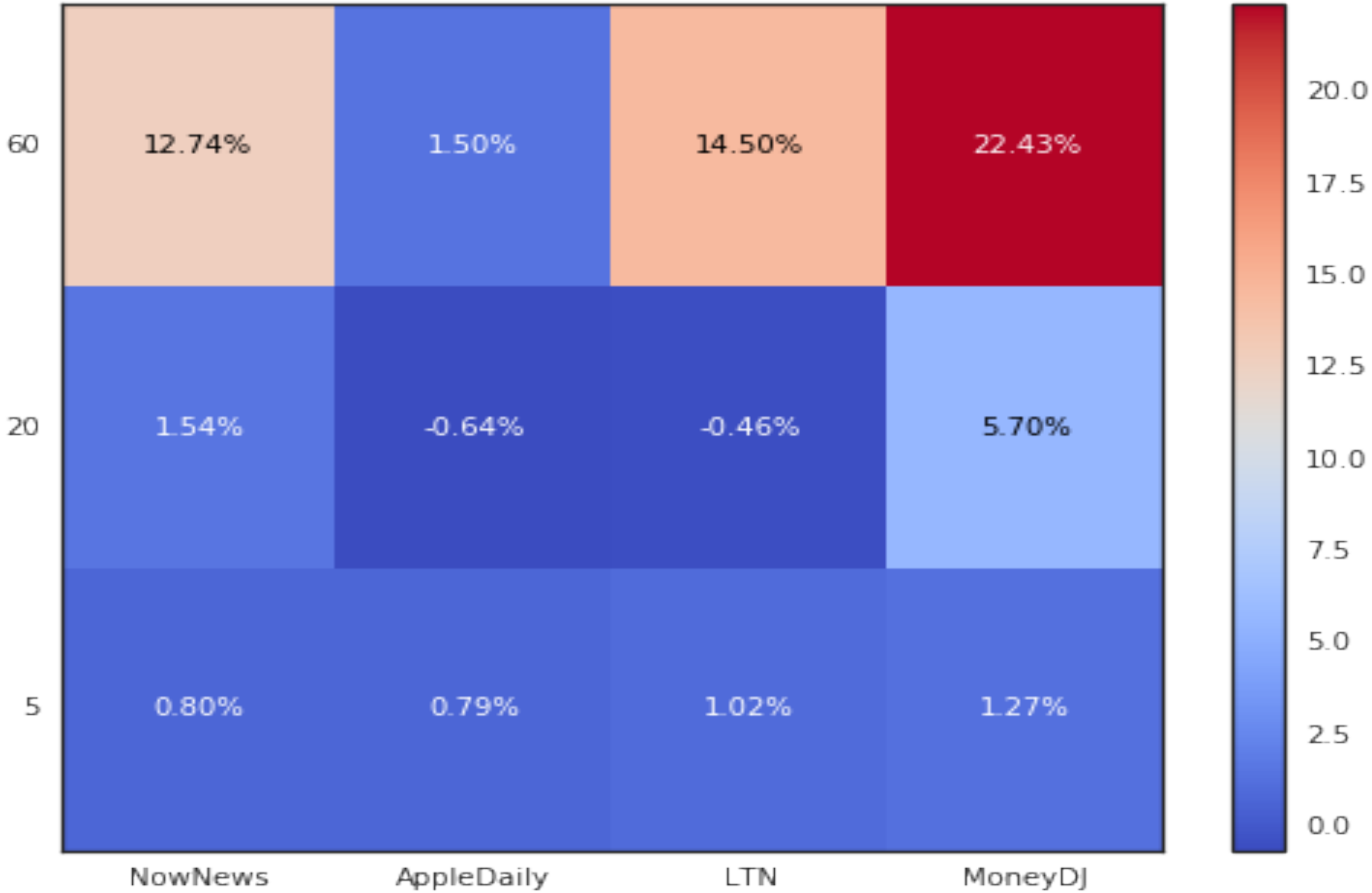
ROI of 60 days with Lexicon-based Trading



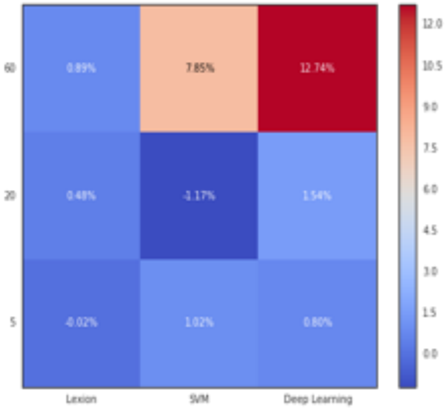
ROI Heatmap of Lexicon-base Trading



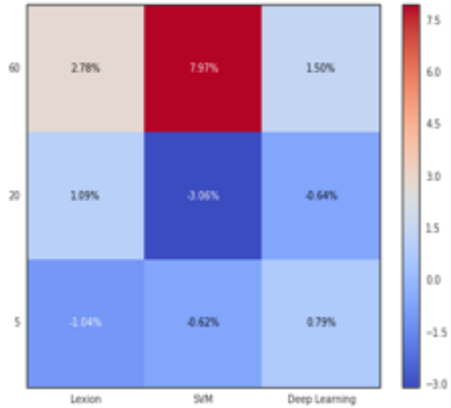
ROI Heatmap of Trading with Deep Learning Approach



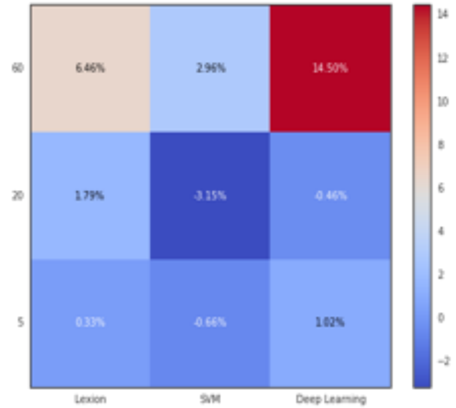
Comparison of ROI Heatmap from Various Finance News Media



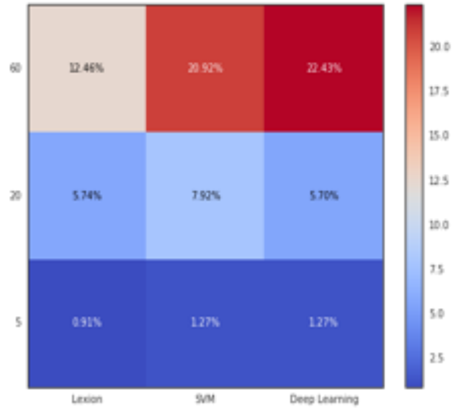
NowNews



AppleDaily



LTN



MoneyDJ

Conclusions

Findings

- We proposed analytical methods with deep learning in financial news sources on the stock price trend forecasts.
- The results showed that the source of financial news media for the exclusive domain of Finance and Economics, revealed its investment information representing a reference value.

Contributions

- The different news media release financial news different from its reference value level messages containing investment, as investors choose finance message referenced sources.
- The prediction accuracy will be improved via a prediction model of the deep learning.

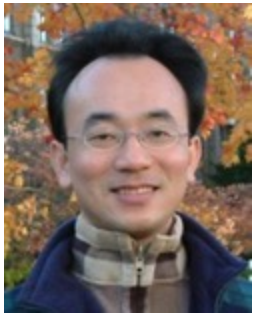
Managerial Implications

- Different news media with their own characteristics and specializations.
- The values of the financial information may be different due to the following reasons:
  - Company's business principles.
  - Edition team's specializations and their knowledge of industry.
  - Journalist's habits and preferences in wording.
  - Sensitivity of financial market trends of the Media.

Q&A

Tamkang University

Deep Learning for Financial Sentiment Analysis on Finance News Providers



Min-Yuh Day

Chia-Chou Lee

Department of Information Management

Tamkang University, Taiwan

myday@mail.tku.edu.tw

The 7th International Workshop on Mining and Analyzing Social Networks for Decision Support (MSNDS 2016)

MSNDS 2016, IEEE/ACM ASONAM 2016, San Francisco, CA, USA, August 18-21, 2016

Summary

- Big Data Analytics
- FinTech
- Financial Sentiment Analysis

Q & A

Tamkang University
淡江大學

Big Data Analytics for Financial Sentiment Analysis in FinTech

(金融科技財務大數據情感分析)

Time: 2016/11/15 (Tue) (10:25-12:10)
Place: 朝陽科技大學資訊工程研究所 <教室： E520 演講廳>
Host: 吳世弘 教授 (Professor Shih-Hung Wu)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail. tku.edu.tw/myday/