The Future of FinTech

Sanjiv R. Das
Santa Clara University

NFIC @Stanford
Palo Alto
May 2017
Using **theory** to develop models to apply big data.

Question/problems are **primary**, data is secondary.

**Simplicity, transparency** of models fosters implementability.

**Dimension reduction** for sufficient statistics.

Analytics per se is **multidisciplinary**.

**Disparate** data is the norm.
Benefits of Analytics for Large Banks

- Monitoring corporate buzz.
- Analyzing data to detect, analyze, and understand the more profitable customers or products.
- Targeting new clients.
- Customer retention.
- Lending activity (automated)
- Market prediction and trading.
- Risk management.
- Automated financial analysts.
- Financial forensics to prevent rogue employees.
- Credit cards: optimizing use, marketing offers.
- Fraud detection.
- Detecting market manipulation.
- Social network analysis of clients.
- Measuring institutional risk from systemic risk.
Machine Learning is Effective

Source: Eurekahedge
FinTech Landscape

- 1,400 FinTech companies with $33 BN in funding.
- Losses from credit card fraud are $31 BN per year.
- FinTech adoption rates:

```
<table>
<thead>
<tr>
<th>Country</th>
<th>Adoption Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td></td>
</tr>
</tbody>
</table>
```

“Using Big Data to Detect Financial Fraud Aided by FinTech Methods”
- S. Srinivasan, Texas Southern U.
The high cost of financial intermediation is being dis-intermediated by data-driven technologies.

Philippon (2016)
FinTech 1. Network Models

1. A growing literature in this area. Espinosa-Vega (2010); Espinosa-Vega and Sola (2010); Billio, Getmansky, Lo, and Pelizzon (2012); Merton, Billio, Getmansky, Gray, Lo, and Pelizzon (2013); and Das (2016).

2. Example: the Midas Project.¹

- Focus on financial companies that are the domain for systemic risk (SIFIs).
- Extract information from unstructured text (filings).
- Information can be analyzed at the institutional level or aggregated system-wide.
- Applications: Systemic risk metrics; governance.
- Technology: information extraction (IE), entity resolution, mapping and fusion, scalable Hadoop architecture.

## Systemically Important Financial Institutions (SIFIs)

<table>
<thead>
<tr>
<th>Year</th>
<th>#Colending banks</th>
<th>#Coloans</th>
<th>Colending pairs</th>
<th>$R = E(d^2)/E(d)$</th>
<th>Diam.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>241</td>
<td>75</td>
<td>10997</td>
<td>137.91</td>
<td>5</td>
</tr>
<tr>
<td>2006</td>
<td>171</td>
<td>95</td>
<td>4420</td>
<td>172.45</td>
<td>5</td>
</tr>
<tr>
<td>2007</td>
<td>85</td>
<td>49</td>
<td>1793</td>
<td>73.62</td>
<td>4</td>
</tr>
<tr>
<td>2008</td>
<td>69</td>
<td>84</td>
<td>681</td>
<td>68.14</td>
<td>4</td>
</tr>
<tr>
<td>2009</td>
<td>69</td>
<td>42</td>
<td>598</td>
<td>35.35</td>
<td>4</td>
</tr>
</tbody>
</table>

### Node # | Financial Institution | Normalized Centrality
--- | ---------------------- | ------------------
143 | J P Morgan Chase & Co. | 1.000 |
29  | Bank of America Corp.  | 0.926 |
47  | Citigroup Inc.         | 0.639 |
85  | Deutsche Bank Ag New York Branch | 0.636 |
225 | Wachovia Bank NA       | 0.617 |
235 | The Bank of New York   | 0.573 |
134 | Hsbc Bank USA          | 0.530 |
39  | Barclays Bank Plc      | 0.530 |
152 | Keycorp                | 0.524 |
241 | The Royal Bank of Scotland Plc | 0.523 |
6   | Abn Amro Bank N.V.     | 0.448 |
173 | Merrill Lynch Bank USA | 0.374 |
198 | PNC Financial Services Group Inc | 0.372 |
180 | Morgan Stanley         | 0.362 |
42  | Bnp Paribas            | 0.337 |
205 | Royal Bank of Canada   | 0.289 |
236 | The Bank of Nova Scotia| 0.289 |
218 | U.S. Bank NA           | 0.284 |
50  | Calyon New York Branch | 0.273 |
158 | Lehman Brothers Bank Fsb| 0.270 |
213 | Sumitomo Mitsui Banking| 0.236 |
214 | Suntrust Banks Inc     | 0.232 |
221 | UBS Loan Finance Llc   | 0.221 |
211 | State Street Corp      | 0.210 |
228 | Wells Fargo Bank NA    | 0.198 |
Model Data (standard Merton model inputs) for each firm:

- Equity price = \( s = \{s_1, s_2, \ldots, s_n\} \)
- Equity volatility = \( \sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\} \)
- Number of shares = \( m = \{m_1, m_2, \ldots, m_n\} \)
- Risk free rate = \( r \)

Model Variables (all derived from the Merton model):

- \( n \) = number of banks in the system
- \( a = n \)-vector with components \( a_i \) that represent the assets in bank \( i \) (derived from \( s, \sigma, m, r \)).
- \( \lambda = n \)-vector with components \( \lambda_i \) that represent the average yearly chance of bank \( i \) defaulting (from \( s, \sigma, r \)).
- \( E = n \times n \) matrix with components \( E_{ij} \) that represent the probability that if bank \( j \) defaults, it will cause bank \( i \) to default (from \( s, \sigma, r \)).
Define \( \mathbf{c} \) to be an \( n \)-vector with components \( c_i \) that represent bank \( i \)'s credit risk. More specifically, we define

\[
\mathbf{c} = \mathbf{a} \odot \mathbf{\lambda},
\]

where \( \odot \) represents component multiplication; that is, \( c_i = a_i \lambda_i \).

The aggregate systemic risk created by the \( n \) banks in our system is

\[
R = \frac{\sqrt{\mathbf{c}^\top \mathbf{E} \mathbf{c}}}{\mathbf{1}^\top \mathbf{a}},
\]

where \( \mathbf{1} \) is an \( n \)-vector of ones, so the denominator \( \mathbf{1}^\top \mathbf{a} = \sum_{i=1}^{n} a_i \) represents the total assets in the \( n \) banks.

\( r \) is linear homogenous in \( \mathbf{\lambda} \).
Property 1: All other things being equal, $S$ should be minimized by dividing risk equally among the $n$ financial institutions, and maximized by putting all the risk into one institution.

Property 2: $S$ should increase as the financial institutions become more entwined.

Property 3: If all the assets, $a_i$, are multiplied by a common factor, $\alpha > 0$, it should have no effect on $S$.

Property 4: Substanceless partitioning of a bank into two banks should have no effect on $S$. 
Computational Properties

\[ R = \frac{\sqrt{c^T Ec}}{1^T a}, \quad c = a \odot \lambda \]

- \( R \) is linear homogeneous in \( \lambda \): Let \( \alpha \) be any scalar constant. If we replace \( \lambda \) with \( \alpha \lambda \), it immediately follows that \( c \) is replaced by \( \alpha c \), and, by our equation for \( R \), we see that \( R \) is replaced by \( \alpha R \).

- Sensitivity of \( R \) to changes in \( \lambda \): Differentiating our equation for \( R \) with respect to \( \lambda \)

\[ \frac{\partial R}{\partial \lambda} = \frac{1}{2} \frac{a \odot [(E + E^T)c]}{1^T a \sqrt{c^T Ec}} \]

whose components represent the sensitivity of \( R \) to changes in each bank’s value of \( \lambda \). This is the basis of Risk Decomposition, equal to \( (\frac{\partial R}{\partial \lambda} \cdot \lambda) \), a vector containing each bank’s contribution to \( R \).
## Top SIFIs Over Time, 2006-2015

<table>
<thead>
<tr>
<th>Date</th>
<th>SIFI1</th>
<th>SIFI2</th>
<th>SIFI3</th>
<th>SIFI4</th>
<th>SIFI5</th>
<th>SIFI6</th>
<th>SIFI7</th>
<th>SIFI8</th>
</tr>
</thead>
<tbody>
<tr>
<td>20051230</td>
<td>ABNYY</td>
<td>FNMA</td>
<td>FMCC</td>
<td>BCS</td>
<td>SAN</td>
<td>CS</td>
<td>LEHMQ</td>
<td>GS</td>
</tr>
<tr>
<td>20060630</td>
<td>FNMA</td>
<td>LEHMQ</td>
<td>DB</td>
<td>GS</td>
<td>SAN</td>
<td>C</td>
<td>JPM</td>
<td></td>
</tr>
<tr>
<td>20061229</td>
<td>LEHMQ</td>
<td>BSC.1</td>
<td>ABNYY</td>
<td>SAN</td>
<td>BBVA</td>
<td>MS</td>
<td>C</td>
<td>FNMA</td>
</tr>
<tr>
<td>20070629</td>
<td>MTU</td>
<td>LEHMQ</td>
<td>MS</td>
<td>GS</td>
<td>ABNYY</td>
<td>FNMA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20071231</td>
<td>ABNYY</td>
<td>SAN</td>
<td>FNMA</td>
<td>FMCC</td>
<td>MTU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20080630</td>
<td>MTU</td>
<td>FNMA</td>
<td>FMCC</td>
<td>DB</td>
<td>C</td>
<td>MS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20081231</td>
<td>MTU</td>
<td>FMCC</td>
<td>FNMA</td>
<td>C</td>
<td>BAC</td>
<td>DB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20090630</td>
<td>MFG</td>
<td>FNMA</td>
<td>MTU</td>
<td>C</td>
<td>BAC</td>
<td>DB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20091231</td>
<td>MFG</td>
<td>MTU</td>
<td>NMR</td>
<td>FNMA</td>
<td>FMCC</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20100630</td>
<td>MTU</td>
<td>MFG</td>
<td>NMR</td>
<td>CS</td>
<td>BBVA</td>
<td>DB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20101231</td>
<td>MFG</td>
<td>MTU</td>
<td>BBVA</td>
<td>SAN</td>
<td>CS</td>
<td>DB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20110630</td>
<td>MFG</td>
<td>MTU</td>
<td>MS</td>
<td>BBVA</td>
<td>SAN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20111230</td>
<td>BAC</td>
<td>C</td>
<td>MS</td>
<td>DB</td>
<td>BBVA</td>
<td>WFC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20120629</td>
<td>BAC</td>
<td>MS</td>
<td>C</td>
<td>JPM</td>
<td>BBVA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20121231</td>
<td>BBVA</td>
<td>DB</td>
<td>BAC</td>
<td>C</td>
<td>MS</td>
<td>SAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20130628</td>
<td>LYG</td>
<td>MFG</td>
<td>SMFG</td>
<td>RBS</td>
<td>MS</td>
<td>BBVA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20131231</td>
<td>LYG</td>
<td>SMFG</td>
<td>MFG</td>
<td>SAN</td>
<td>BCS</td>
<td>HSBC</td>
<td>JPM</td>
<td>DB</td>
</tr>
<tr>
<td>20140630</td>
<td>LYG</td>
<td>BCS</td>
<td>SMFG</td>
<td>BBVA</td>
<td>BAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20141231</td>
<td>RBS</td>
<td>LYG</td>
<td>MFG</td>
<td>MTU</td>
<td>SMFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20150630</td>
<td>RBS</td>
<td>LYG</td>
<td>SAN</td>
<td>MTU</td>
<td>SMFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20151231</td>
<td>RBS</td>
<td>SAN</td>
<td>BCS</td>
<td>DB</td>
<td>MS</td>
<td>BAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Symbol 1</td>
<td>Symbol 2</td>
<td>Symbol 3</td>
<td>Symbol 4</td>
<td>Symbol 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20051230</td>
<td>FMCC:ABNYY</td>
<td>ABNYY:FMCC</td>
<td>ABNYY:BCS</td>
<td>ABNYY:JPM</td>
<td>ABNYY:CS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20060630</td>
<td>FNMA:DB</td>
<td>DB:FNMA</td>
<td>LEHMQ:GS</td>
<td>GS:LEHMQ</td>
<td>FNMA:C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20061229</td>
<td>BSC.1:LEHMQ</td>
<td>LEHMQ:BSC.1</td>
<td>LEHMQ:MS</td>
<td>DB:C</td>
<td>DB:ABNYY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20070629</td>
<td>LEHMQ:MTU</td>
<td>MS:MTU</td>
<td>MTU:MS</td>
<td>LEHMQ:GS</td>
<td>MTU:LEHMQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20071231</td>
<td>SAN:ABNYY</td>
<td>ABNYY:SAN</td>
<td>FMCC:ABNYY</td>
<td>FNMA:ABNYY</td>
<td>FNMA:SAN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20080630</td>
<td>FNMA:MTU</td>
<td>FMCC:MTU</td>
<td>MTU:FNMA</td>
<td>MTU:FMCC</td>
<td>FNMA:FMCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20081231</td>
<td>FMCC:MTU</td>
<td>FNMA:MTU</td>
<td>C:MTU</td>
<td>MTU:DB</td>
<td>FMCC:DB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20090630</td>
<td>MFG:MTU</td>
<td>FNMA:MTU</td>
<td>FNMA:MFG</td>
<td>MTU:MFG</td>
<td>C:MTU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20091231</td>
<td>MFG:MTU</td>
<td>NMR:MTU</td>
<td>FNMA:MTU</td>
<td>FMCC:MTU</td>
<td>MTU:MFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20100630</td>
<td>MFG:MTU</td>
<td>MTU:MFG</td>
<td>NMR:MTU</td>
<td>NMR:MFG</td>
<td>BBVA:MTU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20101231</td>
<td>MFG:MTU</td>
<td>MTU:MFG</td>
<td>BBVA:MTU</td>
<td>BBVA:MFG</td>
<td>SAN:MTU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20110630</td>
<td>MFG:MTU</td>
<td>MTU:MFG</td>
<td>MS:MTU</td>
<td>MS:MFG</td>
<td>BBVA:MTU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20111230</td>
<td>C:BAC</td>
<td>BAC:C</td>
<td>MS:BAC</td>
<td>BAC:DB</td>
<td>MS:DB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20120629</td>
<td>MS:BAC</td>
<td>C:BAC</td>
<td>BAC:C</td>
<td>BAC:JPM</td>
<td>JPM:BAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20121231</td>
<td>BBVA:BAC</td>
<td>C:DB</td>
<td>MS:BAC</td>
<td>DB:C</td>
<td>BAC:BBVA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20130628</td>
<td>LYG:MFG</td>
<td>SMFG:MFG</td>
<td>MFG:SMFG</td>
<td>LYG:SMFG</td>
<td>MFG:LYG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20131231</td>
<td>LYG:SMFG</td>
<td>SMFG:LYG</td>
<td>SMFG:MFG</td>
<td>LYG:MFG</td>
<td>MFG:SMFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20140630</td>
<td>LYG:BCS</td>
<td>SMFG:BCS</td>
<td>LYG:SMFG</td>
<td>BCS:SMFG</td>
<td>BCS:LYG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20141231</td>
<td>RBS:MTU</td>
<td>LYG:MTU</td>
<td>MFG:MTU</td>
<td>SMFG:MTU</td>
<td>RBS:MFG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20150630</td>
<td>LYG:RBS</td>
<td>SAN:RBS</td>
<td>RBS:SAN</td>
<td>LYG:SAN</td>
<td>RBS:LYG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20151231</td>
<td>RBS:SAN</td>
<td>SAN:RBS</td>
<td>SAN:BCS</td>
<td>RBS:BCS</td>
<td>BCS:SAN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Systemic Risk in Indian Banks

Fragility

2.91

Systemic Risk Score

15.75

Risk Decomposition
FinTech 2. Consumer Finance

- Wei, Yildirim, den Bulte, and Dellarocas (2015), application using social media interactions.
- Big data helps eliminate bias from small data, see Choudhry, Das, and Hartman-Glaser (2016), ills are outlined in detail in O’Neill (2016).
- Robo-Advising is emerging as a huge force for disintermediation.
- In general, there are four issues that Fintech is resolving in consumer finance:
  - Low returns, makes retirement targets hard to achieve.
  - Longevity risk.
  - High volatility.
  - High cost providers.
FinTech 3. Nowcasting

- Recent literature: Evans (2005); Giannone, Reichlin, and Small (2008); and Babura, Giannone, Modugno, and Reichlin (2013).
- GDPNow, a model from the Atlanta Fed.

Root Mean Square Forecast Error of GDP Growth (SAAR) For GDPNow Model: 2000:Q1 – 2013:Q4
BILLIQ: An Index-Based Measure of Illiquidity.
Uses option pricing theory to derive a measure for the cost of immediacy:

\[
BILLIQ = -10,000 \times \ln \left( \frac{NAV}{NAV + |ETF - NAV|} \right)
\]

The paper that derives this measure of illiquidity is:

Text analytics: Das (2014); Jegadeesh and Wu (2013); Loughran and McDonald (2014). Topic analysis, Blei, Ng, and Jordan (2003). Opens up new areas of risk.

- Number of risk words predicts earnings next quarter to be lower.
- Lower *readability* predicts lower next quarter performance.
- Larger annual report MD&A predicts lower next quarter performance.
- Size of filing to SEC server predicts worse performance.
Title: “Zero-Revelation Linguistic Regulation: Detecting Risk Through Corporate Emails and News”

- Financials are often delayed indicators of corporate quality.
- Internal discussion may be used as an early warning system for upcoming corporate malaise.
- Emails have the potential to predict such events.
- Software can analyze vast quantities of textual data not amenable to human processing.
- Corporate senior management may also use these analyses to better predict and manage impending crisis for their firms.
- The approach requires zero revelation of emails.
Enron: Email Length

Weekly Time Series Plot of Average Length per Email Message

Average Number of Characters

Weekly Time Series Plot of Average Length per Email Message

Average Number of Characters

Sanjiv R. Das  Future of FinTech  Stanford NFIC 2017
Enron: Sentiment and Returns

![Graph showing sentiment and stock returns over time](image-url)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate (t-statistic)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>MA Net Sentiment</td>
<td>XXX***</td>
<td>0.575</td>
</tr>
<tr>
<td></td>
<td>(XXX)</td>
<td>(0.63)</td>
</tr>
<tr>
<td>MA Email Length</td>
<td>0.584***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.97)</td>
<td></td>
</tr>
<tr>
<td>MA Total Emails</td>
<td></td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.10)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.406*</td>
<td>-0.671***</td>
</tr>
<tr>
<td></td>
<td>(-1.93)</td>
<td>(-3.08)</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>XXX</td>
<td>0.09</td>
</tr>
</tbody>
</table>
| Number of observations   | 88                               | 88    | 88    | 88
Enron: WordPlay

Select Word

credit

Weekly Time Series Plot of Word Frequency

Black: words/#emails, Red: sentiment

Word over time  Word_vs_Sentiment  Word_vs_Return  Word_Sentiment_Correlation  Word_Return_Correlation
Enron: Topic Analysis

![Bar chart showing topic proportions over months](image_url)
Conversations across India and around RBI

- Conversations across India on RBI, its people and the monetary policy
- Governor features in many conversations across both rural and urban areas
- Some conversations specifically around monetary policy

- Bubbles show split of conversations around Deputy RBI Governor, Monetary Policy, Raghuram Rajan, RBI and RBI Governor.
- Based on count of unique conversations
- Date Range: 1st – 14th April, 2015
### Top Topics along with RBI

<table>
<thead>
<tr>
<th>Topic</th>
<th>Sentiment Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>Negative 120, Neutral 300, Positive 80</td>
</tr>
<tr>
<td>Repo rate</td>
<td>Negative 300, Neutral 200, Positive 100</td>
</tr>
<tr>
<td>March</td>
<td>Negative 150, Neutral 200, Positive 50</td>
</tr>
<tr>
<td>Deputy RBI Governor</td>
<td>Negative 100, Neutral 150, Positive 50</td>
</tr>
<tr>
<td>Rate</td>
<td>Negative 100, Neutral 200, Positive 100</td>
</tr>
<tr>
<td>RBI Governor</td>
<td>Negative 20, Neutral 120, Positive 50</td>
</tr>
<tr>
<td>Jaitley</td>
<td>Negative 10, Neutral 100, Positive 50</td>
</tr>
<tr>
<td>Monetary Policy</td>
<td>Negative 20, Neutral 100, Positive 30</td>
</tr>
<tr>
<td>April</td>
<td>Negative 20, Neutral 100, Positive 30</td>
</tr>
<tr>
<td>Governor</td>
<td>Negative 20, Neutral 100, Positive 30</td>
</tr>
<tr>
<td>Rate cut</td>
<td>Negative 20, Neutral 100, Positive 30</td>
</tr>
<tr>
<td>RBI</td>
<td>Negative 20, Neutral 100, Positive 30</td>
</tr>
</tbody>
</table>

**Sentiment Categories**
- **Negative**
- **Neutral**
- **Positive**

- Repo rate evokes negative sentiment as people don’t expect it to be changed.
- Repo rate, rate cut and monetary policy are discussed frequently with RBI.

---

**Text excerpts**
- "@NDTVProfit: RBI unlikely to change repo rate at policy review smiln" by @NDTVProfit
- "Digging India’s RBI Out of Morass of Debt" by @NDTVProfit
- "Financial stability is like Pornography. You can’t define it but when you see it you know it" - D Subbarao (RBI Governor)
- "I was disappointed by the fiscal relaxation." - Ex-RBI Governor on India’s budget and growth:
- "Rajan is perfect, he explains complex economic," PM Modi on RBI governor.
- "RBI Conference" shows up as trending topic in India at rank 18.
Cybersecurity is a massive application area.

Three sources of risk:
- State actors.
- Organized crime.
- Internal agents.

The Dark Web.

Interesting books
- Kingpin
- Fatal System Error
ATTACK ORIGINS BY GEOGRAPHY

Total number of attacks detected by geography of origin.

COUNTRY RANKING

Top 5
Top 6 - 10
Top 11 - 20
Top 21 - 50
Top 31 - 50
Top 50+
Financial fraud allows perpetrator to be removed from the scene of the crime.

Therefore, logging all financial activity to enable traceback is critical.

Limited defense at the authentication stage if there is a data breach.

Widespread use of machine learning.

Social media based; highly consumer-centric. Device usage, Email use, customer location at time of transaction.

Adaptive behavioral analytics, e.g., Bionym, EyeVerify, BioCatch.
FinTech 7. Payment Systems

- Bypassing the banks, e.g., Apple pay, Samsung pay, Google pay, Venmo, Dwolla.
- Demonetization in India leads to the rise of PayTm.
- Disintermediation of the payday lending business, e.g. PayActiv.
TradeWorx (http://www.tradeworx.com/) and Automated Trading Desk (ATD, bought by Citibank for $680M in 2007) were pioneers in the field. Algorithmic trading, 50% of executed trades in the equity markets, down from around 2/3 of stock trades in late 2000s, profits from algorithmic trading are under competitive pressure, and regulatory oversight.

1. Since 2013, 2/3 of the top 30 cited papers on HFTs show positive market effects.
2. Automated firms reduced trading costs, and contrary to popular opinion, improved market depth and stability.
3. Much of the research is possible because this sort of FinTech data has become available.
4. Work by Hendershott and Riordan (HFTs stabilize markets); Hasbrouk and Saar (HFTs improve market quality, reduce bid-ask spreads); Menkveld (HFTs reduce trading costs).
A decentralized record, with copies of the blockchain being maintained by several entities, with (hopefully) comprehensive security and consensus updates.

- Acronym DIST (a file that is Distributed, Immutable, Secure, and Trusted).
- Banks are experimenting with blockchains for automated settlement, and have formed consortia such as R3 (https://r3cev.com/).
- USC (Utility Settlement Coin) from UBS and three other major banks, as well as SETL coin from Goldman Sachs.
Bridgewater Associates: World’s largest hedge fund has a project to automate decision-making to save time and eliminate human emotion volatility.

Goldman Sachs: Two out of the 600 equity traders left. Found that four traders can be replaced by one computer engineer.

Transactions: By 2020 at least five percent of all economic transactions will be handled by autonomous software. AI will process payment functions and learn from customer behaviours, through Intelligent Payment Management (IPM).

Savings: AI will help consumers make daily financial decisions and monitor spending. New Personal Financial Management apps use *contextual awareness*, which measures spending habits and online footprints to create personalised advice. Combining pooled financial data with end-user control to offer tailor-made services is a classic AI solution.
AI in Finance

- Cross-selling: Categorization-as-a-Service (Caas), for understanding customer transactions for cross selling.
- Fraud prevention: Mining user data to detect abnormal behavior, anomalies, and unusual transactions.
- Mizuho Financial Group sent Pepper, its humanoid robot into its Tokyo branch to handle customer inquiries. Partnering with IBM to enable Pepper to understand human emotions, and build interaction into apps.
- RBS is trialing Luvo AI, a customer service assistant to interact with staff and customers.
- AXA (insurer) has an app-based bot called Xtra, it engages in bespoke conversations with customers about healthy living.
- AI is used in peer2peer lending.
At JPMorgan Chase & Co., a learning machine is parsing financial deals that once kept legal teams busy for thousands of hours.

The program, called COIN, for Contract Intelligence, does the mind-numbing job of interpreting commercial-loan agreements that, until the project went online in June, consumed 360,000 hours of work each year by lawyers and loan officers. The software reviews documents in seconds, is less error-prone and never asks for vacation.
Blackrock: replacing human stock pickers with machine algorithms.


Numerai: Hedge fund makes trades by aggregating trading algorithms submitted by anonymous contributors, prizes are awarded in BTC.

Emma: Evolved a hedge fund using a software that writes news articles.

**Challenges:**
- Very little data about the track record of these hedge funds, as the business remains secretive.
- Investor reluctance to turn over money completely to a machine.
Deep Learning Beats the Market

- Use all stocks in the S&P500 from 1963 till today
- 5 hidden layers; 512 nodes each; ReLU with Dropout
- Training 80%, validation 20%
- Accuracy of prediction of the market direction next day = 57%
- The statistical significance is a t-stat = 7.54
Pitfalls to Avoid in FinTech

- GIGO: Garbage in, garbage out.
- IO (Information Overload): Collecting too much data and not using it correctly.
- BiNB (Bigger is Not Better): Big data leads to bigger errors if misused.
- CCC: Confusing correlation with causality.
- $$\$: May involve expensive infrastructure.
- TiP (Trust is Paramount): Privacy issues.
- CS (Customer Satisfaction): Excessive misdirected automation leading to poor client service.