FinTech Platforms* Data, and AI**
In the post-COVID world

Vasant Dhar
Professor, Stern School of Business & Center for Data Science, NYU

@vasantdhar

* Dhar, V., and Stein, R., FinTech Platforms & Strategy, Communications of the ACM, October, 2017
**Dhar, V., Data Science and Prediction, Communications of the ACM, December 2013.
**Dhar, V,. When to Trust Robots with Decisions and When Not To, Harvard Business Review, May 2016
**Dhar, V., and Yu, H., On the Stability of Machine Learning Models: Measuring Model and Outcome Variance, Journal of Investment Management, Summer issue, 2020

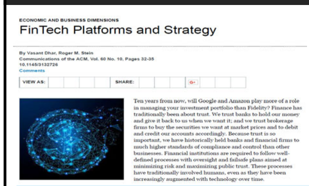
The History of Technology in Financial Services

- Processes, such as matching, execution, settlement, and payment were based on human systems prior to computing
- Computers gradually reduced friction and human touch, which is still happening
- Platforms emerged
- COVID has accelerated the move to virtualization and hence FinTech

Who Am I?

- Professor Stern/CDS, Director of the PhD program in Data Science, Center for Data Science at NYU
- "Pracademic," have grown up with three things
  - Artificial Intelligence
  - Data
  - Machine Learning
- Set up the first machine-learning-based hedge fund in mid-90s
- Mostly worked in the financial arena, but also sports, business, healthcare, education and social
- Advisor to government and industry

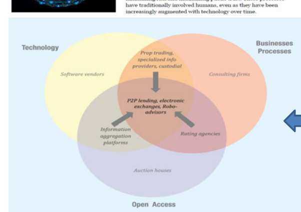
Current Drivers: FinTech Platforms and Data/AI


ECONOMIC AND BUSINESS DIMENSIONS
FinTech Platforms and Strategy

By Vasant Dhar, Roger M. Stein
Communications of the ACM, Vol. 60 No. 10, Pages 32-35
10.1145/3132726
Comments

VIEW AS: SHARE:

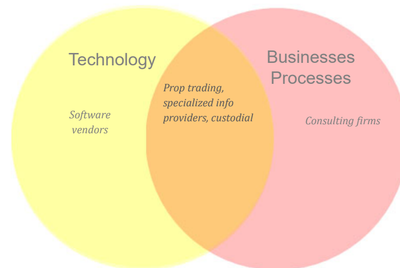
Ten years from now, will Google and Amazon play more of a role in managing your investment portfolio than Fidelity? Finance has traditionally been about trust. We trust banks to hold our money and give it back to us when we want it; and we trust brokerage firms to buy the securities we want at market prices and to debit and credit our accounts accordingly. Because trust is so important, we have historically held banks and financial firms to much higher standards of compliance and control than other businesses. Financial institutions are required to follow well-defined processes with oversight and failsafe plans aimed at minimizing risk and maximizing public trust. These processes have traditionally involved humans, even as they have been increasingly augmented with technology over time.

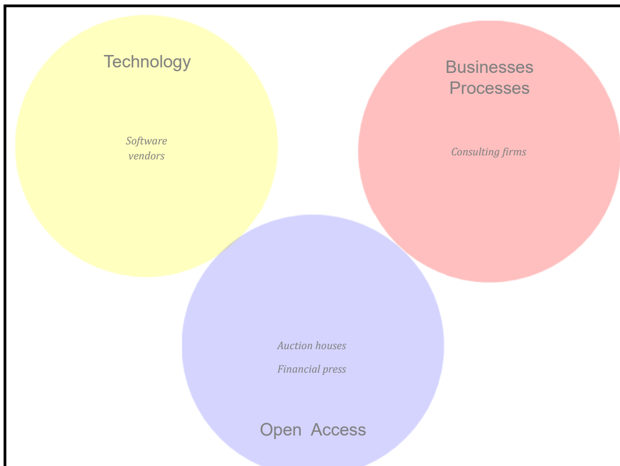


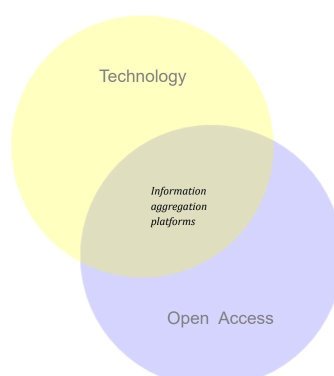
Data and Machine Learning

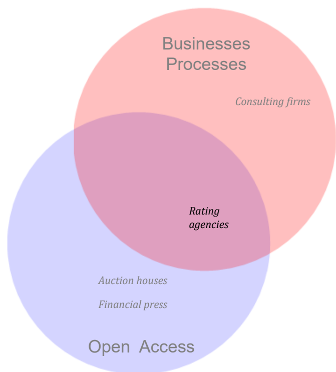
x1	x2	x3	x4	x5	x6	x7	x8	x9	Y
1	0	1	1	0	0	0	1	1	1
0	0	0	0	1	0	0	1	0	0
1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	0	0
1	0	1	0	0	0	1	0	0	1
0	1	0	0	0	0	0	1	0	0
1	0	1	0	0	0	0	1	0	0

FinTech Platforms





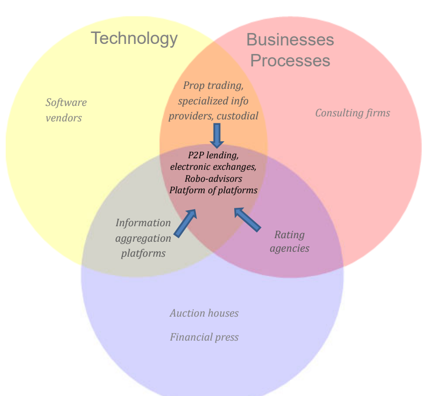




How Platforms are Developing

- Decision Automation
  - “Data pipelines:” use data to automate processes → complete platforms
  - Where do humans add value?
- Virtualization (a la Isaac Asimov “The Naked Sun”)
  - Acceleration post-COVID
  - Cloud-based everything
- Component standandization
  - Business processes (eventually Blockchain will replace Title, Depositories, etc, which are the backbone of trust)
  - RegTech replacing bespoke home-grown processes

COMPLETE PLATFORMS



Platforms Trends and Risks

- Winner take all outcomes
  - A few big winners
  - Many powerless providers
- Invasion of privacy
  - Caused by lack of choice
  - Changing expectations
- Big-Tech have become rapidly become the “digital utilities” that EVERYONE relies on
  - Have you tried to live without your cloud, maps, search and retail apps?!*

*https://www.nytimes.com/2020/07/31/technology/blocking-the-tech-giants.html

Platform Dependencies

- Avoid Amazon?
  - Can't use NETFLIX! Its direct competitor!
  - Can't use AWS Cloud
- Google?
  - Slow Internet!
  - No cloud → no dropbox etc.
  - No Uber/Lyft, or anything that uses Maps!
- Apple?
  - No iphone, App Store etc
- Facebook?
  - No Instagram

Global Platform Trends and Regulation*

- China
  - Efficiency/control driven: state owns/controls data
- United States
  - Advertising driven; low individual protection or control of data; but this might change
- Europe
  - Protection driven; higher individual protection
- India
  - Inclusion driven: high individual protection and control of data
  - Public "data trusts"

https://www.washingtonpost.com/news/monkey-cage/wp/2018/09/25/who-controls-your-data-india-may-pass-a-law-ensuring-that-you-do/

And Who Has the AI Infrastructure?

- Amazon (Alexa, etc)
- Google (BERT, etc)
- Apple (SIRI, etc)
- Microsoft (Azure, etc)
- Facebook (VR)
- NVIDIA (deep learning etc)

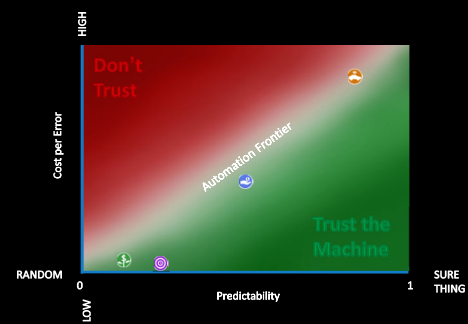
Will these platforms become the FinTech platforms of the future?

Human & Machine Intelligence

When Do We Trust Autonomous Machine-Based Learning Systems?

- Automated decision making via autonomous learning systems is at the heart of FinTech platforms
- When do we let machines make decisions for us?

The Automation Frontier: When We Trust Machines With Decision Making

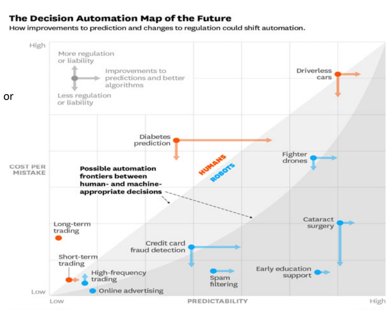


The Automation Frontier: When We Trust Machines With Decision Making

These costs are from data or human experts

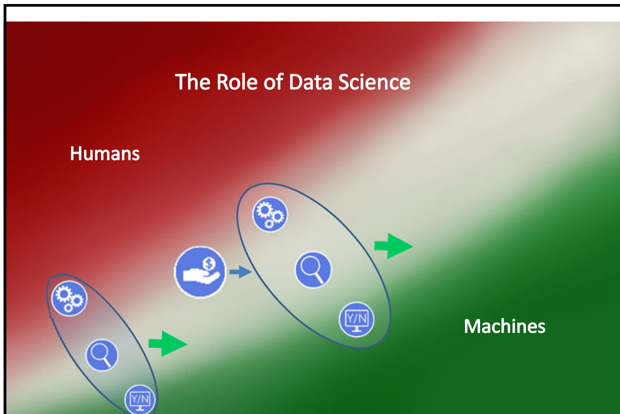
↓

i.e. C(P+,A-)&C(P-,A+)



↑

Can be estimated from data (i.e. Bayes rate)

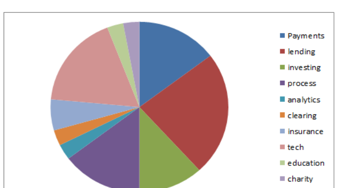


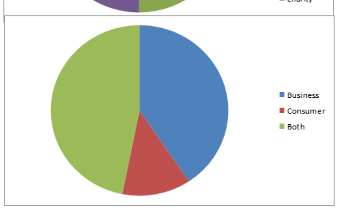
Artificial Intelligence in FinTech

- Lending (credit)
- Investing
- Risk Management
- Payments
- Processes (RegTech, etc)
- Insurance
- Analytics

Lending

A VC FinTech Portfolio





Credit Use Cases

	Use-Case	Traditional
Credit	1 **Underwriting**	• Manual processes and ad-hoc desktop tools assess risk • Decisions are driven by policies rather than data
	2 **Credit line management and early warning systems**	• Individual excel-based tools are used to determine the line for each account • Relationship managers check-in quarterly to monitor risk
	3 **Collections**	• Accounts are prioritized to collectors in a first-in-first-out sequence • Aggregate metrics are tracked using monthly reports

Source: McKinsey

Credit Use Cases

Credit

Use-Case	Traditional	Advanced
1 Underwriting	• Manual processes and ad-hoc desktop tools assess risk • Decisions are driven by policies rather than data	• Statistical models are used in all credit decisions • Traditional data sources under-pin models with less than 15 inputs
2 Credit line management and early warning systems	• Individual excel-based tools are used to determine the line for each account • Relationship managers check-in quarterly to monitor risk	• Standard statistical models determine the credit limit using traditional data sources • Managers track risk and manually update client data on an ad-hoc basis
3 Collections	• Accounts are prioritized to collectors in a first-in-first-out sequence • Aggregate metrics are tracked using monthly reports	• A risk-model is used to focus effort on the highest risk accounts • More experienced teams manage higher risk accounts

Source: McKinsey

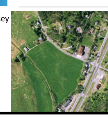
Explanation

Credit Use Cases

Credit

Use-Case	Traditional	Advanced	Leading-edge
1 Underwriting	• Manual processes and ad-hoc desktop tools assess risk • Decisions are driven by policies rather than data	• Statistical models are used in all credit decisions • Traditional data sources under-pin models with less than 15 inputs	• **Deep-learning algorithms** process vast amounts of data to assist decision-making • **Emerging sources of data** (e.g., images, language) improve accuracy
2 Credit line management and early warning systems	• Individual excel-based tools are used to determine the line for each account • Relationship managers check-in quarterly to monitor risk	• Standard statistical models determine the credit limit using traditional data sources • Managers track risk and manually update client data on an ad-hoc basis	• **Machine learning algorithms** continuously assess the optimal line • **Accurate and timely data** on clients (e.g., credit score) and local markets enable point-in-time estimates of risk
3 Collections	• Accounts are prioritized to collectors in a first-in-first-out sequence • Aggregate metrics are tracked using monthly reports	• A risk-model is used to focus effort on the highest risk accounts • More experienced teams manage higher risk accounts	• A **machine learning engine** helps collectors make the **right offer**, at the **right time** through the **right channel** • Granular tracking of results to inform daily huddles

Source: McKinsey



Input feature

Input feature

Credit Use Cases

Continuous A/B Testing

Credit

Use-Case	Traditional	Advanced	Leading-edge
1 Underwriting	• Manual processes and ad-hoc desktop tools assess risk • Decisions are driven by policies rather than data	• Statistical models are used in all credit decisions • Traditional data sources under-pin models with less than 15 inputs	• **Deep-learning algorithms** process vast amounts of data to assist decision-making • **Emerging sources of data** (e.g., images, language) improve accuracy
2 Credit line management and early warning systems	• Individual excel-based tools are used to determine the line for each account • Relationship managers check-in quarterly to monitor risk	• Standard statistical models determine the credit limit using traditional data sources • Managers track risk and manually update client data on an ad-hoc basis	• **Machine learning algorithms** continuously assess the optimal line • **Accurate and timely data** on clients (e.g., credit score) and local markets enable point-in-time estimates of risk
3 Collections	• Accounts are prioritized to collectors in a first-in-first-out sequence • Aggregate metrics are tracked using monthly reports	• A risk-model is used to focus effort on the highest risk accounts • More experienced teams manage higher risk accounts	• A **machine learning engine** helps collectors make the **right offer**, at the **right time** through the **right channel** • Granular tracking of results to inform daily huddles

Source: McKinsey

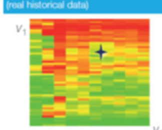
Explanation

Prediction

Why/How are ML Models Better?

+ Our client ■ High risk ■ Medium risk ■ Low risk

Actual phenomenon
(real historical data)

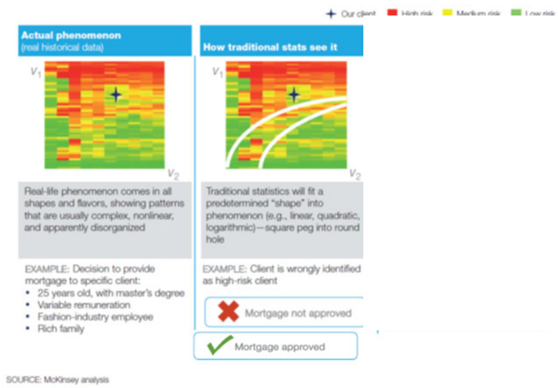


Real-life phenomenon comes in all shapes and flavors, showing patterns that are usually complex, nonlinear, and apparently disorganized

EXAMPLE: Decision to provide mortgage to specific client:
- 25 years old, with master's degree
- Variable remuneration
- Fashion-industry employee
- Rich family

SOURCE: McKinsey analysis

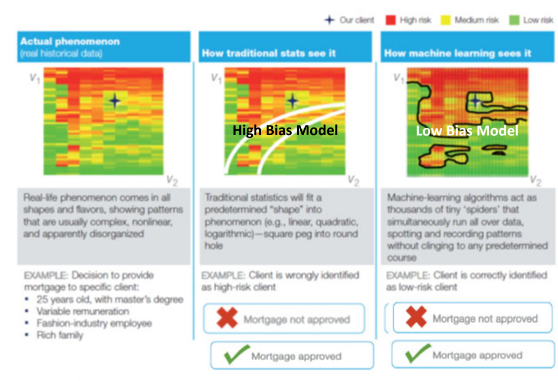
Why/How are ML Models Better?



Real-Time Credit Line Management: Instant Decisions

- Offer line increase
- Do not change credit line
- Monitor and reduce the line
- Attempt to reactivate the card
- Strategize to boost card usage

Why/How are ML Models Better?



Collections

- Right offer at the right time through the right channel
- Timing is critical

REAL ESTATE

SOME "BIG QUESTIONS" IN REAL ESTATE

What is the (predictive) relationship between **real estate lending activity** and **economic activity** by sector and geographic area?

- Listings
- Sales
- Foreclosures
- Store activity
- Taxes, gross income trends, etc

How is **economic activity** and **real estate lending activity** in a sector or geographic area related to **credit risk**? To sectors in **capital markets**?

Where are the biggest **market inefficiencies**?

Real Estate Market Efficiency

- Highly inefficient
  - Illiquid
  - Fragmented
- Huge amounts of untapped data
  - Requires some sweat of the brow to clean and integrate
- Lots of scope in SMBs
  - Banks are too slow and unable to assess risk properly

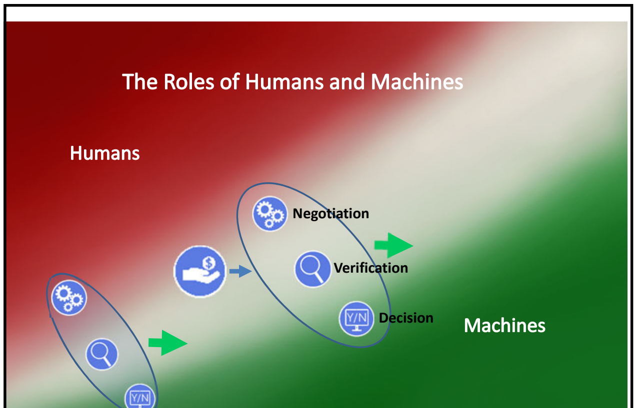
Capital Markets

The Trading/Investing Landscape

	EQUITIES	FX	COMMODITIES	CREDIT
Sell Side	Equities trading, which shifted decades ago to electronic platforms, is one of the first testing grounds for **using artificial intelligence to execute orders**	Firms are tapping big data and machine learning to anticipate client demand and price swings. Software also is helping to design and **manage banks' inventory of more complex rate swaps and currency derivatives**	From liquid contracts tied to assets like gold and oil to physical commodities, the diverse world of commodities doesn't always lend itself to automation. Banks are cataloging trader and salesperson conversations to create **profiles of clients to help better anticipate their desires**	The art of dealing in bonds and more bespoke types of credit has proven more challenging for computers than their much-faster takeover of stock exchanges. Infrequent or opaque trading requires humans to negotiate prices, and banks must carefully juggle holdings to minimize the burden on balance sheets.
Buy Side	Hedge funds and asset managers are using predictive analytics for tasks such as timing stock purchases and assessing risk based on market liquidity. **Computers are also digesting vast data sets -- everything from car registrations to oil-drilling concessions -- to help predict how stocks will perform.**		Global Macro: **Firms are trying to build economists**. They're toying with natural-language processing to sift central bank commentary for clues on future monetary policy. They're also experimenting with algorithms that scour far-flung data, like oil-tanker shipments from the Middle East or satellite images of Chinese industrial sites, to forecast growth.	Some funds are **teaching computers to scan and understand a much larger universe of bond covenants, legal documents and court rulings.** Fully automating analysis of contract and illiquid assets underpinning securities in opaque markets remains a challenge, for now

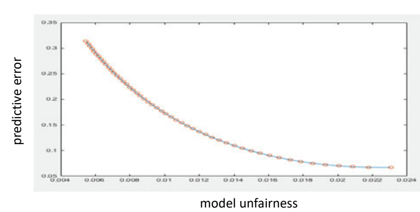
Source: https://www.bloomberg.com/graphics/2017-wall-street-robots/

Where are we Headed?
The Future of FinTech



Increasing Regulation of Platforms

- Domestic and International
  - Principles-based compliance
  - Automated/remote compliance
  - ZeroBias/Fairness demonstration



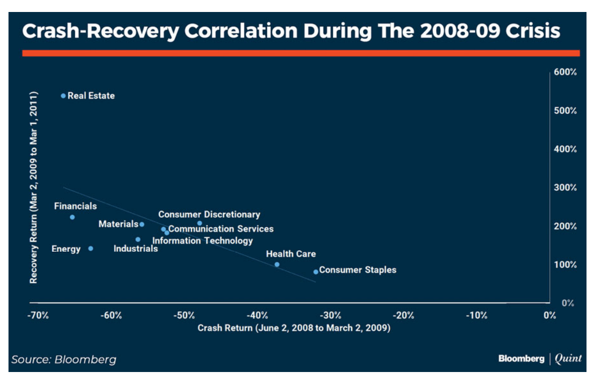
Changing Customer Exectations and GenZ

- Automated Instant Decisions (a la Kabbage)
  - Fraud
  - Credit
  - Authorization
- Personalization
  - Smooth robotic interface
  - Customized everything
- Risk evaluation machines ubiquitous!

Closing Thoughts on Finance as an Industry

Data Assets are Central

- Creation of “alternative databases”
  - i.e. all real estate transactions are recorded, but it requires some “sweat of the brow” to create **clean** and **integrate noisy/incomplete historical data**
    - Naming mismatches
    - Errors in translation of physical to electronic records
    - New players like SnowFlake are addressing this space
- Big-Tech making major inroads into these spaces via **new kinds of data** and better **machine learning** methods
- Will they remain “tools providers” or eat the lunch of existing players?



Thank you!