Digital Readiness Series

Data analytics

Data for the upcoming world: Horizon scanning

By
Dr. Carlo Lipizzi
August 13, 2020

www.sercuarc.org
Data for the upcoming world: Horizon scanning

• Future cannot be predicted, but in science there is a high level of consistency over time. Data Science today is a steppingstone for an even more informed and complex way of living and doing business, with a continuous integration of sources and media, creating semantic synergies, pushing the boundaries of convenience, value and privacy.

• In this seminar, we scan the major trends in Data Science, starting from the current emerging trends, extrapolating scenarios and presenting live examples of emerging applications
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
Faster pacing

- The driving force for the acceleration is communication/data sharing, that is making the world “smaller” for the good and the bad.

- The rate of adoption of new technologies with high impact on the society is increasing dramatically.

![Diagram of technology adoption by households in the United States](image1)

![Diagram of the evolution of communication](image2)
Data growth: what we share

- The digital tools we are using every day are creating data from everything we do at an unprecedented rate: every day, 2.5 quintillion (10^18) bytes of data are created and 90% of the data in the world today was created within the past two years.

- The increasing rate of change, the highly integrated world and the growing massive amount of data we generate and share are the driving forces for the upcoming changes in society and business.
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
Mediated Learning Experience refers to the way in which stimuli experienced in the environment are transformed by a mediating agent, usually a parent, teacher, sibling, or other decisive element in the life of the learner.
The “Always-on Internet” society

Focused on

Sharing
Cooperation
Collective Actions
Reduced Privacy
Source uncertainty
Information overload
“Living in the cloud”
“Living in the cloud” effects: the overstimulation

- The average attention span in 2015 - 8 seconds
- The average attention span in 2000 - 12 seconds
- The average attention span of a goldfish - 9 seconds
- Percent of teens who forget major details of close friends and relatives - 25 %
- Percent of people who forget their own birthdays from time to time - 7 %
- Average number of times per hour an office worker checks their email inbox - 30
- Average length watched of a single internet video - 2.7 minutes
A snapshot of the generations

Among Millennials, women outpacing men in college completion
% of 25- to 37-year-olds who have completed at least a bachelor’s degree

<table>
<thead>
<tr>
<th>Generation</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Millennial (2018)</td>
<td>43%</td>
<td>36%</td>
</tr>
<tr>
<td>Gen X (2001)</td>
<td>28%</td>
<td>28%</td>
</tr>
<tr>
<td>Late Boomer (1989)</td>
<td>24%</td>
<td>26%</td>
</tr>
<tr>
<td>Early Boomer (1982)</td>
<td>21%</td>
<td>27%</td>
</tr>
<tr>
<td>Silent (1968)</td>
<td>11%</td>
<td>19%</td>
</tr>
</tbody>
</table>


PEW RESEARCH CENTER

Millennials became the largest generation in the labor force in 2016
U.S. labor force, in millions

Note: Labor force includes those ages 16 and older who are working or looking for work. Annual averages shown. Source: Pew Research Center analysis of monthly 1994-2018 Current Population Survey (IPUMS).

PEW RESEARCH CENTER

Millennials lead on some technology adoption measures, but Boomers and Gen Xers are also heavy adopters
% of U.S. adults in each generation who say they ...

<table>
<thead>
<tr>
<th>Technology</th>
<th>Millennial</th>
<th>Gen X</th>
<th>Silent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own a smartphone</td>
<td>92%</td>
<td>90%</td>
<td>60%</td>
</tr>
<tr>
<td>Own a tablet computer</td>
<td>68%</td>
<td>62%</td>
<td>40%</td>
</tr>
<tr>
<td>Use social media</td>
<td>82%</td>
<td>76%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Note: Those who did not give an answer are not shown. Source: Survey conducted Jan. 8 - Feb. 7, 2019.

PEW RESEARCH CENTER
A snapshot of the generations – the take

- Using more technology in everyday activities
- More “sophisticated” use of devices and data
- More data will be generated
- Pushing upward the automation
- More “embedded analytics”
Social Issues

Majority of Americans say gender equality is very important, whether they identify as feminists or not

% of ___ who say ___

- Non-feminists
- Feminists
- All adults

- It is very important for women to have equal rights with men
- Favor the ERA being added to the U.S. Constitution
- People NOT seeing discrimination against women where it really DOES exist is the bigger problem*
- Our country has not gone far enough when it comes to giving women equal rights with men

*Compared with people seeing discrimination against women where it really does NOT exist.

Notes: Among those enrolled in elementary and secondary public schools. Data for 2010 is projected. White, Black and Asian are single race, non-Hispanic. Hispanics are of any race. Asian includes Pacific Islander. Multiracial and other single-race groups not shown.


The #MeToo hashtag has been used roughly 19 million times on Twitter in the past year, and usage often surges around news events

Number of Twitter posts mentioning the #MeToo hashtag, Oct. 15, 2017-Sept. 30, 2018

Source: Pew Research Center analysis of publicly available tweets using Division Hexagon.

PEW RESEARCH CENTER
Social Issues – the take

- More diversified cultural predominant elements
- Stronger formal ethical principles
- More polarizing conditions
- More need for customizations/individual-level targeting
Technology and the Society

Some groups have reached near-saturation levels for adoption of basic technologies
% of U.S. adults who say they own or use this technology

<table>
<thead>
<tr>
<th>Technology</th>
<th>Ages 18-49</th>
<th>College grad+</th>
<th>Household income $75K+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellphone</td>
<td>99</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>Internet</td>
<td>97</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>Smartphone</td>
<td>91</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>Social media</td>
<td>82</td>
<td>79</td>
<td>77</td>
</tr>
<tr>
<td>Desktop/laptop/</td>
<td>77</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>computer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tablet</td>
<td>58</td>
<td>66</td>
<td>72</td>
</tr>
</tbody>
</table>

Source: Survey conducted Jan. 3-10, 2018.
PEW Research Center

Total hours worked in Europe and United States, 2016 vs 2030 estimate, billion

<table>
<thead>
<tr>
<th>Physical and manual skills</th>
<th>203</th>
<th>2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic cognitive skills</td>
<td>115</td>
<td>151</td>
</tr>
<tr>
<td>Higher cognitive skills</td>
<td>140</td>
<td>148</td>
</tr>
<tr>
<td>Social and emotional skills</td>
<td>119</td>
<td>113</td>
</tr>
<tr>
<td>Technological skills</td>
<td>73</td>
<td>55</td>
</tr>
</tbody>
</table>

Change in hours spent by 2030, %

-14
-15

Source: McKinsey Global Institute Workforce Skills Model; McKinsey Global Institute analysis

Note: Respondents who did not give an answer or gave other responses are not shown.
Source: Survey of U.S. adults conducted Jan. 3-10, 2018. Trend data from previous Center surveys.
"Mobile Technology and Home Broadband 2019"
PEW Research Center
Technology and the Society – the take

- More mobile access to the net/info, but Covid-19 may change this for good
- More control over technologies
- Plateau effect for leading technology -> need to analytics to find niches
- Continuous shift toward STEM/high added value competencies -> potential increase of the social divide on top jobs; potential increase of “easy” user interfaces for complex systems
- The gap in skill demand and availability will drive up the use of AI-like solutions and global re-skill (when applicable)
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
The Business environment

Defying the average is more important than ever

How to play is much more important than where to play

If you want to see your data again...

- Number of ransomware attacks per minute
The Business environment – the take

- Security and privacy are soaring in importance: 38% of corporates say improving security and securing the enterprise is a top improvement priority, on par with driving innovation (39%) and improving customer experience (40%)

- Businesses are pivoting around data. If cloud is the digital foundation, then data are the building blocks of the new world technology. Only 4% of corporates claim to have a "highly sophisticated" approach to leveraging data

- “How you play is more relevant than where you play” -> data analytics play a critical role

- There will be an increasing demand for AI/analytics driven solutions to contain cybercrime and vetting information sources
Impact on Finance

Vision of the future
United States, public equity assets
Latest available, % of total public equities (worth $31tn)

Managed funds
Automated 35.1%
Managed funds
Human 24.3%
Other owners 40.6%

Sources: Russell 3000; Federal Reserve; Bloomberg; Morningstar; ETF.com; HFR, Prequin; JPMorgan Chase
*Estimate
#Government, insurance, foreigners

Tech That Will Change Wealth Management

- Mobile apps: 40%
- Robo advisors: 37%
- Digital platforms (e.g., client portals): 36%
- Behavioral finance software: 35%
- Automated investment advice: 27%
- Real-time financial activity reporting: 22%
- Predictive analytics/big data: 21%
- Advanced risk-profiling software: 19%
- Artificial intelligence/chatbots: 13%
- Virtual Reality/3-D conferencing: 10%

Source: Financial Planning Tech Survey, 2017

Projectured Reductions
Banking job cuts in Europe are expected to taper going forward

Peak 2015 2025
3.5m 2.5m 1.5m 0.5m

Source: ECB, Citi GPS

Norway's bank customers preferred to use digital platforms when applying for products and services
Respondents who use branches or digital platforms.

- Branch
- Digital (online or mobile)

- Checking account: 54% of 80%
- Savings account: 16% of 75%
- Debit card: 21% of 68%
- Credit card: 18% of 71%
- Personal loan: 42% of 45%
- Wealth management: 40% of 40%

Note: Percentages do not total 100 percent because the data for "contact centers" is not included.
Source: Deloitte Center for Financial Services analysts.
Impact on Military

AI warfare market

Market Summary
CAGR 40.25%

Source: Mordor Intelligence

Artificial Intelligence in Modern Warfare Market - Growth Rate by Region (2019 - 2024)

Regional Growth Rates
- High
- Mid
- Low
Source: Mordor Intelligence

Unmanned Sea Systems Market: Revenue (%), by Capability, Global, 2019

Source: Mordor Intelligence

AI TALENT CONCENTRATION

China has the second-largest number of researchers who have published AI papers or been issued patents in the past decade. But the proportion of those considered to be in the top 10% of the field is smaller than in other AI-leading nations.
Impact on Wearable - Healthcare

Source: globalmarketinginsights.com

Source: Intersog
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
Technology trends

Source: Gartner
Technology trends – the take

- Technology innovation is data driven: increasing the complexity of the systems requires advanced analytics for management and optimization

- Human Augmentation - Enhancing Cognitive Abilities – Hybrid Systems: combining the speed and pattern-matching capabilities of algorithms with the higher-order reasoning and imagination of humans

- Digital ecosystems: leveraging on an interdependent group of actors (enterprises, people and things) sharing digital platforms to achieve a mutually beneficial purpose

- Sensing and Mobility: combining sensor technologies with AI, systems get a better understanding of the world around them, enabling mobility and manipulation of objects
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
Top data trends happening now

Key contributing factors already happening

- *Decline of the Dashboard* - Data stories will be the most widespread way of consuming analytics, and stories will be automatically generated using augmented analytics techniques
- *X Analytics* (e.g., text analytics, video analytics, audio analytics, etc.) - AI-supported content analytics for video, audio, vibration, text, emotion will trigger major innovations and transformations
- *Augmented Data Management*: Metadata Is “the New Black” - Organizations will utilize active metadata, machine learning and data fabrics to dynamically connect, optimizing and automating most of the data management processes
- *Data Marketplaces and Exchanges* - Large organizations will be either sellers or buyers of data via formal online data marketplaces
- *Relationships Form the Foundation of Data and Analytics Value* - Graph technologies will facilitate rapid contextualization for decision making

*Source: processed Gartner info*
Data & NLP trends

Top 10 Data and Analytics Trends That Will Change Your Business

- Smarter, faster, more responsible AI
- Decline of the dashboard
- Decision intelligence
- X analytics

- Augmented data management: Metadata is the new black
- Cloud is a given
- Data and analytics worlds collide

- Data marketplaces/exchanges
- Practical blockchain (for data and analytics)
- Relationships form the foundation of data and analytics value

Source: Gartner

gartner.com/SmarterWithGartner

© 2019 Gartner, Inc. All rights reserved. FR_78/030
Data/NLP trends – the take

- Augmented Analytics: autonomous or semiautonomous examination of data reducing the data preparation phase, based on AI/ML
- Real-time intelligence, combining analytics with back-end systems able to capture "momentary markets"
- Blockchain - enabling data analytics applications
- Data Fabric to enable seamless data sharing across a distributed network
- NLP technology: end of fully supervised systems, rise of semi supervised, able to leverage on domain-specific knowledge
- NLP technology: data/text convergence via computational representations of knowledge
- NLP technology: semantic data preparation; bias and context detection
- NLP applications: automatic content generation; smart chatbots; primary UI for analytics; semantic search; semantic real time translators
• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
The “crystal ball”

Future cannot be predicted, but 1. can be estimated; 2. rarely is not heavily rooted in the past

- We take several million documents from the past 10 years
- Transform them into semantic numerical forms (“embeddings”)
- Interpolate the evolution of the elements that are more coherent in time
- The interpolated element is the “predicted” value
- We are experimenting on technology evolution
The “crystal ball” – the network approach

- We create a network of words ("semantic network") per each time slot, “learn” how they morph over time, predict the next morphed network.

- We use a computational and visual analysis of changes in technology related semantic elements over the years.

- In particular, we use Graph Neural Network and Recurrent Neural Network to evaluate and predict the morphing.
The “crystal ball” – the linear algebra approach

- We create one computational representation of the text (“embeddings”) per each time slot, consolidate the embeddings keeping their time stamps, predict the next morphed embedding and provide a graphical representation.

Example of embeddings: a matrix with the unique “words” as 1st column, each word with numbers defining it in the given text.
The “crystal ball” – where we are

• The system has been developed as part of a project sponsored by the U.S. Department of Defense through the Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) under Contract [HQ0034-19-D-0003, TO#0150]

• At the current stage, the prototype can predict
  • The “neighborhoods” where a technology can be in the future
  • The elements/keywords that will characterize the predicted technology

• We are testing the system “predicting” the present, based on the past. Results are improving, fine tuning the different steps

• One of the critical element is the way the “embeddings” are created. We currently use a leading vectorization method (Word2Vec), but we are evaluating a method based on non-Euclidian spaces/“quantum spaces”
The datatification
How Society is changing
How Business environment is changing
Technology trends
Data & NLP trends
Building the crystal ball
Sample Applications
Data Science Trilogy: key take-aways
### Samples of application

**Pouria Babvey** is a PhD candidate in School of Systems and Enterprises in Stevens Institute of Technology. He holds degrees in Electrical Engineering from University of Sharif (MSc) and the University of Tehran (BSc). His research efforts focus on the development of artificial intelligence architectures for natural language processing and social network analysis. During the time at Stevens he has been participating in projects for UNICEF, Wiley, and NSF on scholar networks, social media conversational patterns, evolutionary trends in the tech-products, and human-bot collaboration. He also has been teaching Computer Algorithms for several years and published some books.

**Fernanda Capela** is a PhD candidate in School of Systems and Enterprises in Stevens Institute of Technology. She holds a degree in Industrial Engineering with a minor in Financial Management from the Federal University of Rio de Janeiro (UFRJ, Brazil). She is a data scientist and her research explores the impacts of big and crowdsourced online data in the modern society. Her areas of interest are Data Analytics, Data Mining, Text Mining, Emotion Recognition, Sentiment and Opinion Analysis, Machine Learning, Social Media and Social Network Analysis, Data Visualization, Data Story-telling, and Computational Social Sciences.
Agenda

• The datatification
• How Society is changing
• How Business environment is changing
• Technology trends
• Data & NLP trends
• Building the crystal ball
• Sample Applications
• Data Science Trilogy: key take-aways
Data Science Trilogy: key take-aways

• We are in a digital transformation process that is creating a new kind of economy based on the “datafication” of virtually any aspect of human social, political and economic activity as a result of the information generated by the digitally connected individuals, companies, institutions and machines
• Data Science projects are processes, not just an application of algorithms and tools
• You need to know and interact with your data to get meanings out of it
• Data Science is part science, part craftsmanship. The science is getting better, but the craftsmanship will stay for a while
• Intelligent systems are formal representations of someone’s intelligence: focus first on the intelligence you are bringing to the process, not on the technology
• Data Science develops models of the reality based on the data you can collect: data has to be representative of that reality. You need to know the reality and be sure it is properly represented by the data you have
• Don’t disregard results if they look different from what you expect: check the model first but then have an open mind on what the data is telling you
Thank you!

Dr. Carlo Lipizzi
clipizzi@stevens.edu
Content-aware Galaxies: Digital Fingerprints of Discussions on Social Media
Donald J. Trump 📢 @realDonaldTrump · Mar 21
HYDROXYCHLOROQUINE & AZITHROMYCIN, taken together, have a real chance to be one of the biggest game changers in the history of medicine. The FDA has moved mountains - Thank You! Hopefully they will BOTH (H works better with A, International Journal of Antimicrobial Agents).....

Donald J. Trump 📢 @realDonaldTrump · Aug 13, 2019
As usual, China said they were going to be buying “big” from our great American Farmers. So far they have not done what they said. Maybe this will be different!

Elizabeth Warren 📢 @ewarren · Jan 31, 2019
Our children & grandchildren should grow up in a world where they can breathe the air & drink the water - and go outside without risking their lives in extreme temperatures. It’s time to protect our planet & pass a Green New Deal. #PolarVortex2019
Framework

Data Collection
1) Collecting the core using keywords
2) Extending the dataset to get the replies

Reconstructing the network between tweets

Text Analytics
- Unsupervised topic extraction using an attention-based auto-encoder
- Unsupervised sentiment detection using a rule-based model
- Unsupervised coherency measurement between tweet pairs using BERT
- Supervised stance classification of tweets using supervised learning

Content-aware Galaxies
- topic-aware
- sentiment-aware
- coherency-aware
- stance-aware
Network visualization

-criticize
-support
-public policy
-election
-foreign policy
-legislation
-share news
-general arguments
Topic-aware Galaxies
Word2Vec is used to convert words to vectors.

“He is subsidizing farmers with tariff money from China, not borrowed money from China."
Attention Mechanism

by *ent423*, *ent261* correspondant updated 9:49 pm et, Thu March 19, 2015 (*ent261*) a *ent114* was killed in a parachute accident in *ent45*, *ent85*, near *ent312*, a *ent119* official told *ent261* on Wednesday, he was identified Thursday as special warfare operator 3rd class *ent23*, 29, of *ent187*, *ent265*. *ent23* distinguished himself consistently throughout his career, he was the epitome of the quiet professional in all facets of his life, and he leaves an inspiring legacy of natural tenacity and focused

... 

*ent119* identifies deceased sailor as **X**, who leaves behind a wife

by *ent270*, *ent223* updated 9:35 am et, Mon March 2, 2015 (*ent223*) *ent63* went familial for fall at its fashion show in *ent231* on Sunday, dedicating its collection to "mamma" with nary a pair of "mom jeans" in sight. *ent164* and *ent21*, who are behind the *ent196* brand, sent models down the runway in decidedly feminine dresses and skirts adorned with roses, lace and even embroidered doodles by the designers' own nieces and nephews. Many of the looks featured saccharine needlework phrases like "I love you", ...

**X** dedicated their fall fashion show to moms
Model Architecture

Tariffs are bringing companies back from China to the USA.

Word2vec

Words’ semantic space

Attention

Twitter Embedding

<table>
<thead>
<tr>
<th></th>
<th>tariff</th>
<th>bringing</th>
<th>companies</th>
<th>back</th>
<th>China</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.04</td>
<td>0.3</td>
<td>0.01</td>
<td>0.2</td>
<td>0.15</td>
</tr>
</tbody>
</table>

\[ \vec{V}_{Tweet} = 0.5 \times \vec{V}_{tariff} + 0.04 \times \vec{V}_{bringing} + \ldots \]

Dimension reduction

Topic Weights

- Topic 1: 0.05
- Topic 2: 0.05
- Topic 3: 0.1
- Topic 4: 0.1
- Topic 5: 0.7
Case Study

President Trump motivated Tweets: ~ 30 million tweets Jan 2019-Sep 2019
<table>
<thead>
<tr>
<th>Topic Label</th>
<th>Sub-topic Label</th>
<th>Top words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support President</td>
<td>affirming President</td>
<td>georgesoros, do_your_job, trump_war_room, mitchellvii</td>
</tr>
<tr>
<td></td>
<td>admiring President</td>
<td>blessed, grateful, spirit, best_president_ever, patriot, wonderful, sacrifice</td>
</tr>
<tr>
<td>Criticize President</td>
<td>impeach President</td>
<td>resignnow, jilevin, impeachpence, impeachnow</td>
</tr>
<tr>
<td></td>
<td>racism</td>
<td>divisive, spewed, xenophobia, bigotry, demagogue, sexism, racial</td>
</tr>
<tr>
<td></td>
<td>satire</td>
<td>eating, cheese, hamberders, juice, drink, snort, cheeseburger, pizza</td>
</tr>
<tr>
<td></td>
<td>sexual assault</td>
<td>assaulted, teenager, sexually, pregnant, incest, infant, raping, lesbian, daughter</td>
</tr>
<tr>
<td></td>
<td>President investigations</td>
<td>cover-up, perjury, prosecution, McCabe, investigation, russiagate, robertmuller, indictment</td>
</tr>
<tr>
<td>Public policy</td>
<td>financial struggles</td>
<td>contractor, charity, loan, mortgage, stiffed, property, reimbursement, owes, salary,</td>
</tr>
<tr>
<td></td>
<td>illegal immigration</td>
<td>illegal_aliens, coyote, smuggling, sanctuary_cities, Yuma, Guatemala, Honduras</td>
</tr>
<tr>
<td></td>
<td>economy</td>
<td>deregulation, wage, industry, layoff, stagnant, growth, gdp, manufacturing, export, economy, inflation, income, taxation, recession, subsidy, revenue, farm</td>
</tr>
<tr>
<td>Next election</td>
<td>election 2020</td>
<td>2020election, beat, gerrymandering, challenger, candidate, senatorial, ballot, berniesanders, voter, ewarren, andrewgillum, incumbent</td>
</tr>
<tr>
<td>World news</td>
<td>foreign policy</td>
<td>Erdogan, Pakistan, Turkey, Maduro, jzarif, india, Syria, Kurd, Afghan, Cuba, Tehran, Mullah, Japenese</td>
</tr>
<tr>
<td>Arguments</td>
<td>arguments</td>
<td>although, confused, obviously, honestly, opposite, impressed, saying, neither, unfortunately, genuinely, complicated, admit, besides, weird</td>
</tr>
<tr>
<td>Legislation</td>
<td>Congress news</td>
<td>fake_emergency, veto, override, resolution, rand_paul, senator_collins, unanimously, senate_majldr, legislating, senate, government_shutdown</td>
</tr>
<tr>
<td></td>
<td>bill process</td>
<td>require, function, procedure, legislative, statute, limiting, institution, mandatory, enact, administer, obligation, governmental, banning</td>
</tr>
<tr>
<td>referrings media</td>
<td>Democrats’ media</td>
<td>huffpost, columnist, newsweek, editorial, newyorker, latimes, buzzfeednews</td>
</tr>
<tr>
<td></td>
<td>Republicans’ media</td>
<td>gov_mike_dewine, Shulkin, Toomey, elizabeth_pipko, james_aydelott</td>
</tr>
<tr>
<td></td>
<td>diplomatic visits</td>
<td>guest, hosted, palace, Buckingham, Orlando, dinner, ceremony, Manchester, visiting, scheduled, Miami, invitation, delegation, Fayetteville, trumpukvisit</td>
</tr>
<tr>
<td></td>
<td>talking media</td>
<td>shadow_banning, retweets, algorithm, censored, suspended, user, content, unfollowed, Parler, trollbots, satire, feedback, facebook, pinned, fakenews, Koppel</td>
</tr>
<tr>
<td></td>
<td>reporting events</td>
<td>December, 10th, October, alert, date, decodes, percent, estimated, margin, consecutive, neverforgotten, posted, apprehended</td>
</tr>
</tbody>
</table>
Share of topics over time
A Topic-aware Galaxy Example

President Trump, Tweets collected from Sep 9th to 11th in 2019
Stance-aware Galaxies
Case Study: Hydroxychloroquine

- First, collected all the tweets that contained “Hydroxychloroquine” in the first 5 months of 2020 as the core (~400k).
- Next, using some fields we extended our dataset to collect all the replies (5 millions)
- Our results show 29% support, 38% oppose, and 33% neutral messages
- Criticism increased over time
- Neutral language did not generate any cross-cutting discussion with significant majority of one side in the generated discussions
Tweets posted from March 20th to April 4th in 2020 containing ~47000 tweets.
Text Mining and NLP Applications for Decision-Making

Fernanda Capela
fcapela@stevens.edu

August 2020
Risk Panel System – Problem Scope

- Assess the risk of a company to **lose competitiveness** in its market segments, based on investments and acquisitions.

- **Decision support system** that recommends technological applications for companies in the security industry.

  - Who are my major competitors?
  - What is my position compared to my competition?
  - What technologies do they have that I don’t?
  - What new technologies from my market segment are being commercialized? Which ones are being developed?
  - Which technologies will give me the best competitive advantage return?
Who are my major competitors?

- Companies with similar technology portfolio that compete with me for market share.
- Technologies that can overcome mine – Technology Kill Chain

Which technologies will give me best competitive advantage return?

- Define competitive advantage.
- Create an index to represent the concept quantitatively - MOAT
Risk Panel System – Data Gathering

• What type of data do we need to answer each one of the problem’s questions?
• Consider time period! (Tech industry changes fast!)
• What is the data source? Is it enough? Do we need to combine different datasets? – Web scrapping, APIs, etc.

“DATA PIPELINE”

Variables: Companies, Technologies and Market Segments

Source: Newspapers, tech blogs, patents and academic papers

Analytics: Updated solutions and recommendations

Database
Risk Panel - Analysis and Recommendations

Meshing Capability and Threat-based Science & Technology Resource Allocation

Sector Assessment

Excluding a tech-driven sector can be similar to optimizing an investment portfolio as it is a decision and trade-off regarding expected return and risk.

![Graph showing technological return versus risk](image1)

Notes: Sector(s) in red are recommended as they offer the highest expected return for a defined level of risk, or the lowest risk for a given level of expected return. Sector(s) in grey are other sectors. You may adjust your preferences in the "Preferences Inputs" page.

Sector Recommendation

Based on the technological assessment scores on the left and your risk preferences in the second tab, we suggest general sector for the following sector:

![Graph showing sector recommendations](image2)

Tracking and monitoring of data trends, giving insights about the market.

![Graph showing data trend monitoring](image3)

Enabling Technology

Explore the major enabling technology maturity for the selected sector. Technology maturity score is currently defined as Room Theory based distance.

![Graph showing technology maturity score](image4)

Augmented Reality

![Graph showing technology maturity score across years](image5)

Market Environment

Gather competitors information. Colors represents companies, and the number currently is derived from revenue.

![Graph showing market environment](image6)
Risk Panel - Analysis and Recommendations

Players and Strategy
Make selections to interact with the system.

Select Your Company:
The system will make recommendations for the selected company.
- SAAB AB-B

Select Competitors:
Add or remove competitors.
- • AOT INC
- • ALLIEDGION PLC
- • ASSABLOY AB-B
- • BRINKS CO
- • DOFINABABA HOLDING
- • G4S PLC
- • GENTILDO
- • GEO GROUP INC/TH
- • LOOMIS AB-B
- • PALO ALTO NETWORK
- • 5-1 CORPORATION
- • SAAB AB-B
- • SECURITAS AB-B
- • STANLEY BLACK & DECKER
- • SYMANTEC CORP
- • UNITED TECH CORP

Select Strategy:
- Technology Portfolio evaluates technology ownership and recommends new technologies based on major competitors' technology portfolio.
- Competitive Advantage considers time of ownership as key pillar of competitive advantage and recommends technologies to optimize increase in competitive advantage.

Interactive: User can make selections and play with different scenarios.
Risk Panel - Analysis and Recommendations

Comparative Strengths and Weaknesses with Competition
Your strategic position compared to competitors. Your major competitors are highlighted.

Select a Technology to Divest:
Smart Safe Technology

Ranking Loss for Divesting in Technology
Change in strategic position after dropping investment on selected technology. Your major competitors are highlighted.

Loss in Competitive Advantage if Divest in Technology

Estimation of Impact of Investing and Divesting in Technologies.