

A pilot of a cognitive computing system to analyze immunization data

Marija Stanojevic

PhD Student

Zoran Obradovic

Mentor

Temple University

US-Serbia & West Balkan Data Science Workshop

August 26-28, 2018

Utilities of Cognitive Computing Systems

- Leverage automation for quantitative and qualitative data analysis
- Goal of immunization programs is to maintain or improve vaccination coverage to prevent diseases
- **Quantitative data**
 - Disease surveillance data
 - Vaccination coverage surveys
 - Policy surveillance
 - Annual reports
- **Qualitative data**
 - Funding applications
 - Site visit reports
 - Twitter
 - News reports

} Formal

} Informal

Data

➤ Formal language data

- Vaccines for Children (VFC) policy and procedures, immunization-related websites, journals, and legislation
- Scrapers and parsers used to automate data extraction
- Location for each VFC Awardee

➤ Informal language data

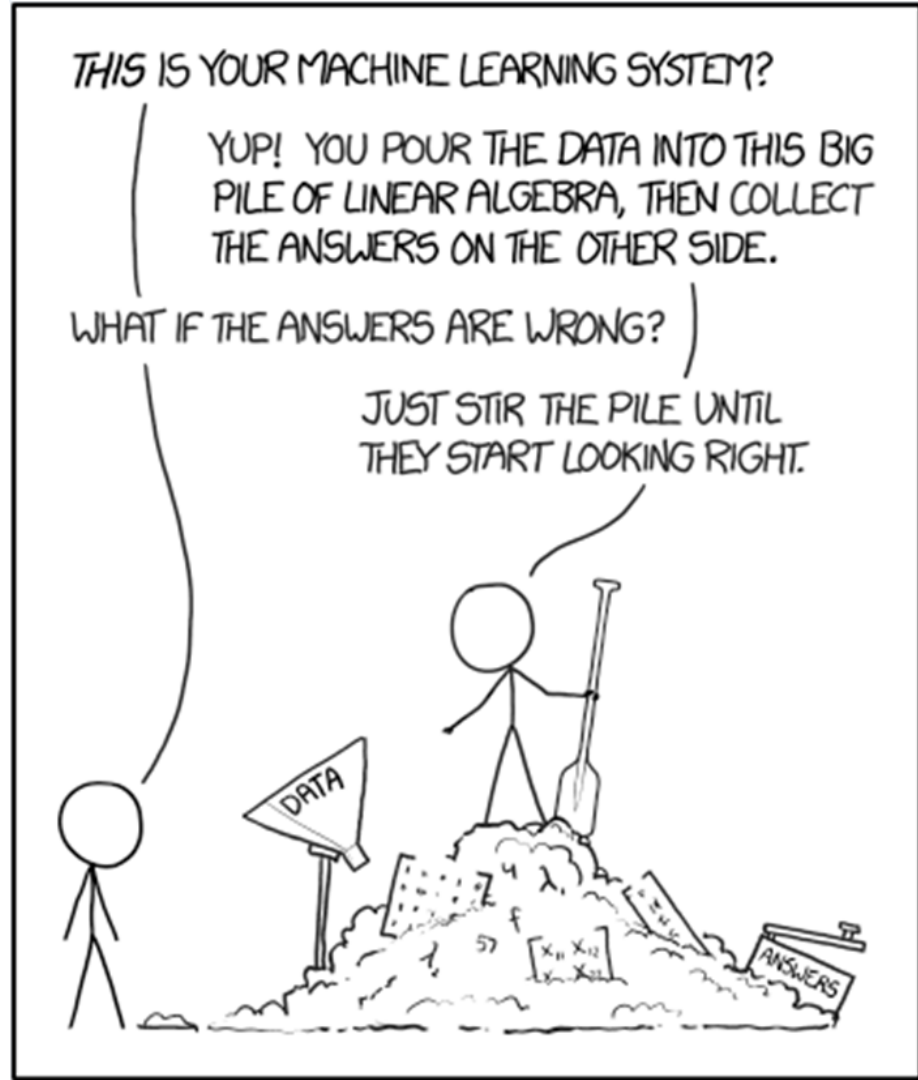
- Tweets, blogs, forums, news articles
- Collected via a social media search platform, Sysomos
- Input from approximately November 2016 to May 2018
- Location information available for at least 60% of informal data.
- Related to the user-profile, not the location of the Tweet

Data preprocessing

- Automated through use of scripts
 - Removed links and Twitter usernames
 - Converted text to lowercase
 - Recognize a given word regardless of letter case (uppercase or lowercase)
 - Tokenized
 - Split text into “tokens”, units for analysis
 - Removed stop words
 - Lemmatization
 - Remove inflectional endings of words to reduce redundancy in results.

Algorithms

- Infer meaning by assessing similarity
- IPCSS algorithms
 - Word2vec and doc2vec
 - Global vectors for word representation (GloVe)
 - Word Topic Mixture (WTM) Model
 - Paragraphs similarity
 - Paragraphs aggregation per state



IPCCS: Informal data paragraphs search

Informal lexicon - words

Informal lexicon - paragraphs

Formal lexicon - words

Formal lexicon - paragraphs

Logout

Informal lexicon - paragraphs

Query:

Results number:

Location:

Start date:

End date:

Word2vec

1. SOURCE New York Blood Center NYBC Urges the Public to Donate to Replenish the Community's Critically Low Blood Supply NEW YORK To donate blood or for information on how to organize a blood drive: Call Toll Free: 1-800-933-2566 Visit: www.nybloodcenter.org/blood (Please see attached list to find the nearest blood drive in your area) In order to maintain a safe blood supply, a seven-day inventory of all types must be continually replenished. Companies, organizations, and community groups are also encouraged to step up and host a blood drive in July or August to help rebuild the blood supply. Hosting a blood drive is easy, and NYBC staff will help you every step of the way. O negative blood donors are considered "universal," and their blood type is needed most readily in trauma situations and emergency departments across the country. Due to its high demand, O negative blood is in short supply, and NYBC encourages individuals with this blood type to consider stepping forward and donating today. Our local blood supply has reached a critically low level, with under a two-day supply of O negative, B negative, and A negative blood. As we head into the summer months, we are reminded of how essential it is that our communities maintain steady participation in blood donation. The best preparation for an unpredictable tragedy is having blood on hospital shelves in advance. This is key to potentially saving lives. "By spreading the word or even hosting your own blood drive, inviting friends, family, and community organizations, you may save lives in your community," said Andrea Cefarelli Historically, during the summer months, blood centers have had to focus on building up the community's blood supply, as it tends to diminish due to seasonal factors.

Glove

1. RT @GodlessApeMan: Dumbest? #YEC #MAGA #ResearchFlatEarth #AntiVax: 0.9679; [Link](#); **Time**: 2018-02-19 21:17:41; **Sentiment**: NEUTRAL; **Author**: J5_Project; **Reach**: 504
2. 01:02 PM CDT: 0.9693; [Link](#); **Time**: 2017-04-05 16:41:26; **Sentiment**: NONE; **Author**: <http://www.arkansasmatters.com>; **Reach**: 8640
3. RT @bengoldacre: Just to reiterate: Andrew Wakefield, struck-off fraudulent anti-vaccine godfather, is at Trump's inaugural ball...: 0.9741; [Link](#); **Time**: 2017-01-21 10:10:28; **Sentiment**: NEGATIVE; **Author**: LiuLab4Virology; **Reach**: 168
4. RT @ChelseaClinton: Ironic that George Washington is on an anti-science anti-vax banner - he had the Continental Army vaccinated against...: 0.9758; [Link](#); **Time**: 2017-03-31 21:16:10; **Sentiment**: NEGATIVE; **Author**: TinaMorphis; **Reach**: 8231
5. RT @NFIDvaccines: #FF During #HeartMonth #GetVaccinated to #FightFlu @Texas_Heart @H_eHA @everettclinic @DrBGellin @iCubed_URI...: 0.9858; [Link](#); **Time**: 2017-02-10 20:03:30; **Sentiment**: NEUTRAL; **Author**: ARAdultImmDoc; **Reach**: 1642
6. Spies, Scandals, and naughty liaisons:A RECKLESS REDEMPTION

WTM

1. The Flu Vaccine Is Working Better Than Expected, C.D.C. Finds <https://t.co/w0apyR12lv>: 1.3159; [Link](#); **Time**: 2018-02-16 16:09:22; **Sentiment**: POSITIVE; **Author**: buckyball360; **Reach**: 334
2. RT @D_J_Oz4: @CDCgov @OANN STOP telling ppl to get their flu shots! It's only going to #KILL people who HAVE the flu! They've ALREADY declared that they predicted the WRONG STRAIN of flu #vaccine What is the point? #RefugeeCrisis Side Effects: 1.3171; [Link](#); **Time**: 2018-02-18 04:31:55; **Sentiment**: NEGATIVE; **Author**: dwlove9; **Reach**: 2335
3. RT @doritim: #BeHPVfreeFL #FLImmysummit Dr. Michael Brown: we can't guarantee no flu if get #vaccines. Working on it - none yet. But better than none...: 1.3206; [Link](#); **Time**: 2017-02-03 21:15:37; **Sentiment**: NEUTRAL; **Author**: ARAdultImmDoc; **Reach**: 1642
4. I added a video to a @YouTube playlist <https://t.co/VWXRYuHfwl> The Undying Mythology of Tetanus - Dr Tim O'Shea: 1.3242; [Link](#); **Time**: 2018-01-03 23:01:15; **Sentiment**: NEUTRAL; **Author**: SteveCherry3000; **Reach**: 124
5. RT @MicrobiomDigest: The Flu Vaccine Is Working Better Than Expected, C.D.C. Finds <https://t.co/rFIXQ0Rxil>: 1.3259; [Link](#); **Time**: 2018-02-16 00:10:44; **Sentiment**: POSITIVE; **Author**:

IPCCS: Formal data words search

Social media lexicon - words

Social media lexicon - queries

Formal data lexicon - words

Formal data lexicon - queries

Logout

Formal data lexicon - words

Query:

Results number:

Location:

Word2vec

1. nonmedical: 0.83
2. philosophical: 0.8
3. religious: 0.8
4. waiver: 0.73
5. objection: 0.73
6. philosophic: 0.72
7. conscientious: 0.71
8. process: 0.7
9. safe: 0.69
10. pbe: 0.69

Glove

1. et: 3.48
2. al: 3.47
3. vaccination: 3.07
4. required: 3.05
5. religious: 2.62
6. regulation: 2.58
7. 1: 2.54
8. health: 2.46
9. 2: 2.44
10. 2009: 2.42

WTM

1. religious: 0.89
2. required: 0.86
3. philosophical: 0.85
4. nonmedical: 0.84
5. philosophic: 0.84
6. conscientious: 0.83
7. must: 0.83
8. homeschooling: 0.82
9. homeschool: 0.82
10. timetable: 0.82

Acknowledgements

➤ **CDC:**

- NCIRD/ISD - Stacie Greby, Yoonjae Kang, Cynthia Knighton
- NICIRD/HCSO - Allison Fisher
- NCIRD/OD - Pamela Srivastava
- CSTLTS (proposed)/PHLP - Alexandra Bhatti
- OCOO – Charles Martin

➤ **Abt Associates:**

- Sarah Ball, William Campbell, Alison Thaung, Ben Holland, Jason Brinkley, Sarah Hamad

➤ **Temple University:**

- Fang Zhou, Zoran Obradovic