

Data Mining Techniques in Online Health Communities

Cali Sweitzer & Cassandra Mikkelson

cali.sweitzer@students.Jefferson.edu

cassandra.mikkelson@students.Jefferson.edu



Meet the Presenter: Cali Sweitzer

- 1st Year PhD Student
- GC550: Foundations in Biomedical Sciences
- Vaccine Research at Boston Children's Hospital and Harvard Institutes of Medicine

Cell Biology &
Regenerative Medicine

Outline

- What is data mining?
- Data Mining Methods:
 - Machine Learning, Natural Language Processing, Sentiment Analysis
- Applications of data mining in healthcare
 - Online health communities
 - FDA databases
- Challenges and Proposed Augmentations



What is Data Mining?

- Data mining in the healthcare setting
- Patient Sentiment
 - emotions, opinions, and attitudes of patients



Prof. Les Sztandera, PhD, Thomas Jefferson University - lecture notes based on *Real-World Data Mining* by Dursun Delen, PhD (0133551075) Copyright © 2015 Pearson Education, Inc. All rights reserved

Online Health Communities

- Informal: Facebook, Reddit, Yelp
- Curated: Patientslikeme, Health Union, Healthboards, etc.
- Government: USA Food and Drug Administration (FDA)
 - FAERS & VAERS
- Output: an abundance of self-generated data on patient sentiment

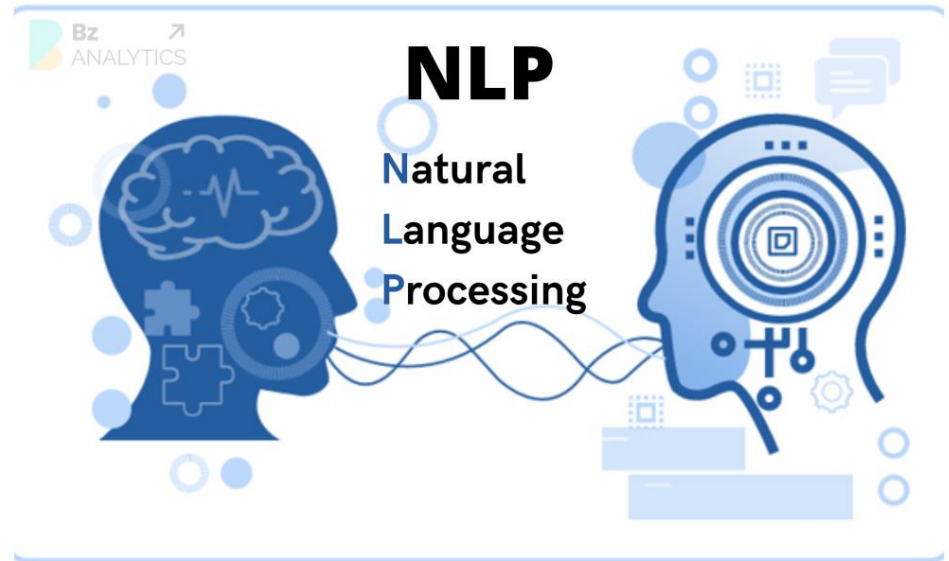


Data Mining Methods

- **Natural Language Processing (NLP):**
 - Machine learning approach
 - Artificial Intelligence (AI)



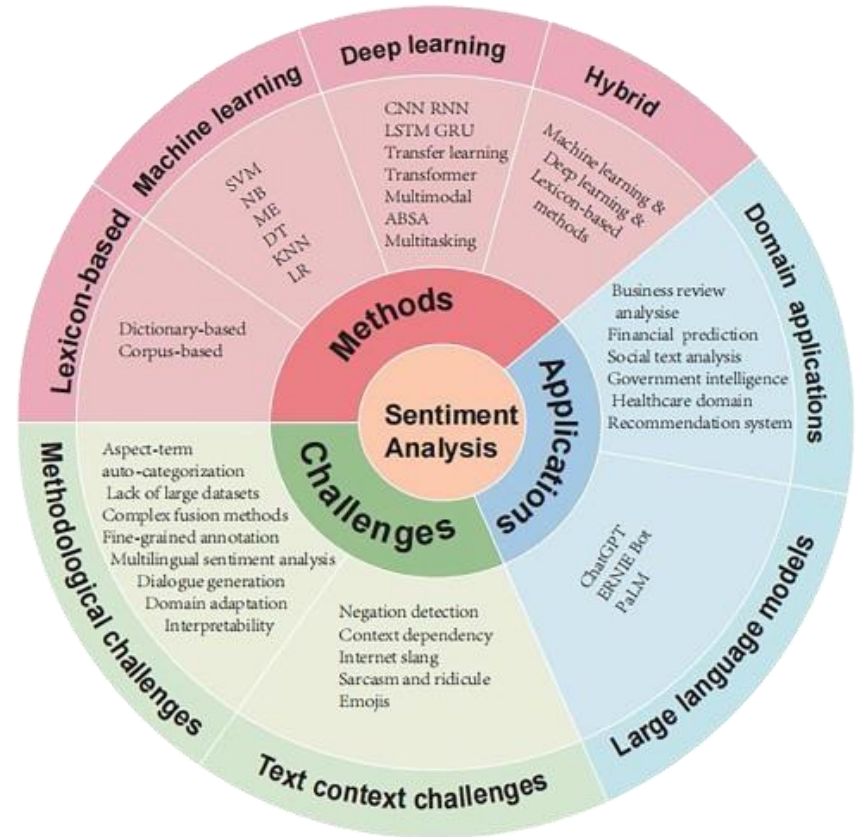
- **Sentiment Analysis**
- **Clustering Algorithms**



<https://endevsols.com>

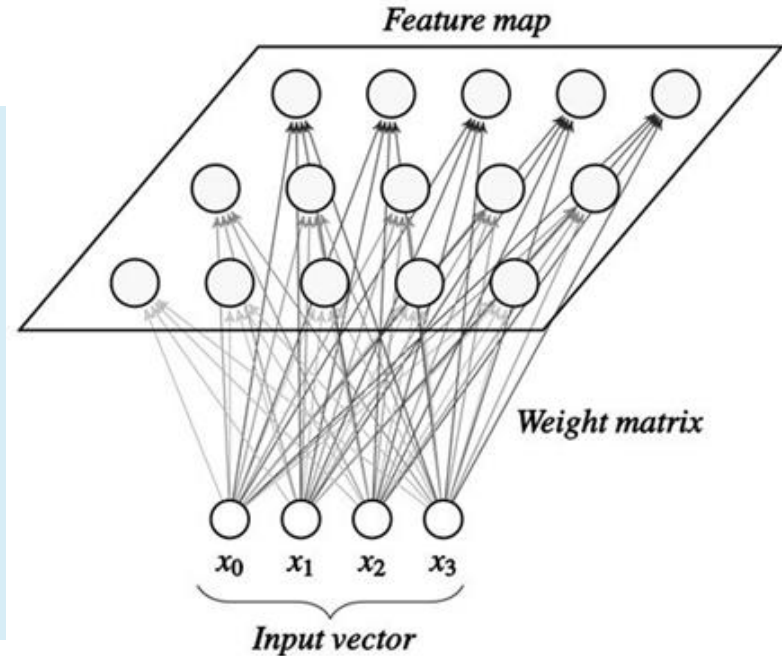
Data Mining Methods (cont.)

- **Sentiment Analysis:** natural language processing to automatically classify sentiment (opinions, emotions, attitudes) expressed in text
 - **Content Analysis:** assessing high frequency terms to deduce concepts from a text
 - **Topic Analysis:** identifying overarching topics within a text based on a probabilistic model



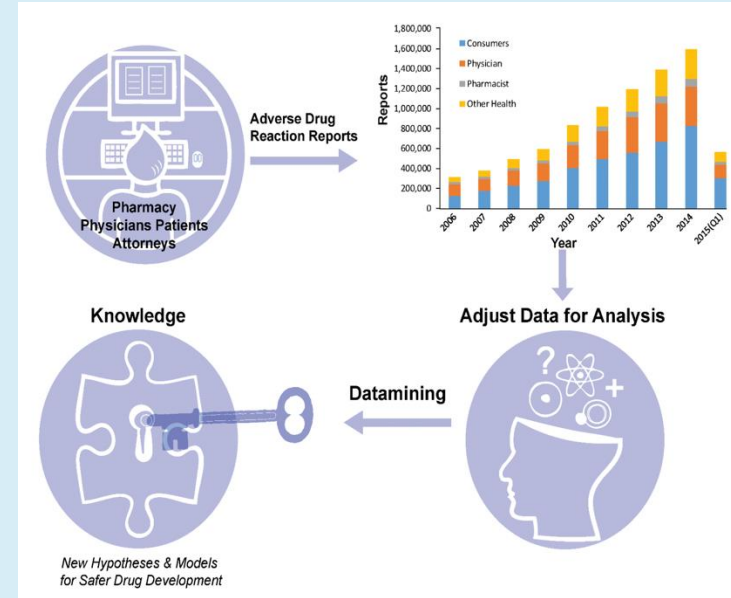
Data Mining Methods (cont.)

- **Self Organizing Map (SOM):** neural network technique that organizes and visualizes high-dimensional data



U.S. Food and Drug Administration (FDA)

- Post- Market Adverse Drug Reactions (ADRs)
- FDA Adverse Event Reporting System (FAERS)
- Vaccine Adverse Event Reporting System (VAERS)
- Proportional Reporting Ratio (PRR)



Bone and Houck. *eLife* 2017;6:e30280. DOI: <https://doi.org/10.7554/eLife.30280>

Challenges and Proposed Augmentations

- Robust datasets of unformatted, unedited language
 - Web scraping tools
- Data and Privacy Laws
 - Social Media vs Government

Conclusions

- **Data mining is a powerful tool** for gathering insights from online health communities
- **Patient-centered research** can benefit from these methods, influencing drug and device development
- **Existing databases like FAERS and VAERS** already use data mining for adverse event detection, highlighting its applicability
- **Proposed techniques are relevant but come with challenges**, requiring further refinement and augmentation
- **Ultimately, biomedical research will be more patient-driven**

References

- [1] M.-G. Fayn, V. des Garets, and A. Rivière, “Collective empowerment of an online patient community: Conceptualizing process dynamics using a multi-method qualitative approach,” *BMC Health Services Research*, vol. 21, pp. 1-19, 2021.
- [2] M. Fallah and S. R. N. Kalhori, “Systematic review of data mining applications in patient-centered mobile-based information systems,” *Healthcare informatics research*, vol. 23, no. 4, pp. 262-270, 2017.
- [3] R. Alnashwan, A. O’Riordan, and H. Sorensen, “Multipleperspective data-driven analysis of online health communities,” in *Healthcare*, MDPI, vol. 11, 2023, p. 2723.
- [4] Y. Mao, Q. Liu, and Y. Zhang, “Sentiment analysis methods, applications, and challenges: A systematic literature review,” *Journal of King Saud University-Computer and Information Sciences*, p. 102 048, 2024.
- [5] M. S. Jawad, W. Adi, A. Salem, and M. Doiher, “Implementation of data mining from social media for improved public health care,” *Future Technologies Conference*, pp. 234-240, 2017.
- [6] D. Isa, V. Kallimani, and L. H. Lee, “Using the self organizing map for clustering of text documents,” *Expert Systems with Applications*, vol. 36, no. 5, pp. 9584-9591, 2009.
- [7] A. Bone and K. Houck, “The benefits of data mining,” *Elife*, vol. 6, e30280, 2017.
- [8] U. Food and D. Administration, *Data mining at FDA*, U.S. Food and Drug Administration website, 2018.
- [9] A. A. Dror et al., “Vaccine hesitancy: The next challenge in the fight against covid-19,” *European journal of epidemiology*, vol. 35, no. 8, pp. 775-779, 2020.
- [10] M. D. Rousculp et al., “Burden and impact of reactogenicity among adults receiving covid-19 vaccines in the United states and Canada: Results from a prospective observational study,” *Vaccines*, vol. 12, no. 1, p. 83, 2024.
- [11] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application.,” *International Journal of Advances in Soft Computing & Its Applications*, vol. 13, no. 3, pp. 144-168, 2020.
- [12] M. Marmot and R. Wilkinson, *Social determinants of health*. Oxford Academic, 2005.

Thank you!



TM

Jefferson

Thomas Jefferson University

HOME OF SIDNEY KIMMEL MEDICAL COLLEGE