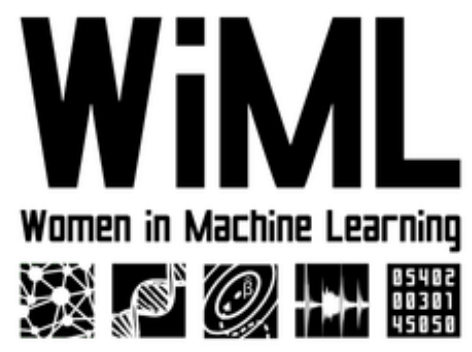


What Can 1.5 Million Tweets Reveal About Breast Cancer Patient Concerns?



Swati Rajwal¹, Avinash Kumar Pandey², Zhishuo Han², Abeed Sarker³

¹Department of Computer Science, Emory University

²Goizueta Business School, Emory University

³Department of Biomedical Informatics, Emory University

INTRODUCTION



EHRs capture health info, not daily habits/interests



Patient-generated meta data can be captured in social media



Chatter available on social media platforms like Twitter (X)



Social media-based NLP improve patient-centered outcomes



Keywords: "cancer", "breastcancer", "tamoxifen", "survivor", & their hashtag equivalents



Patterns were developed by manual analysis of Tweets



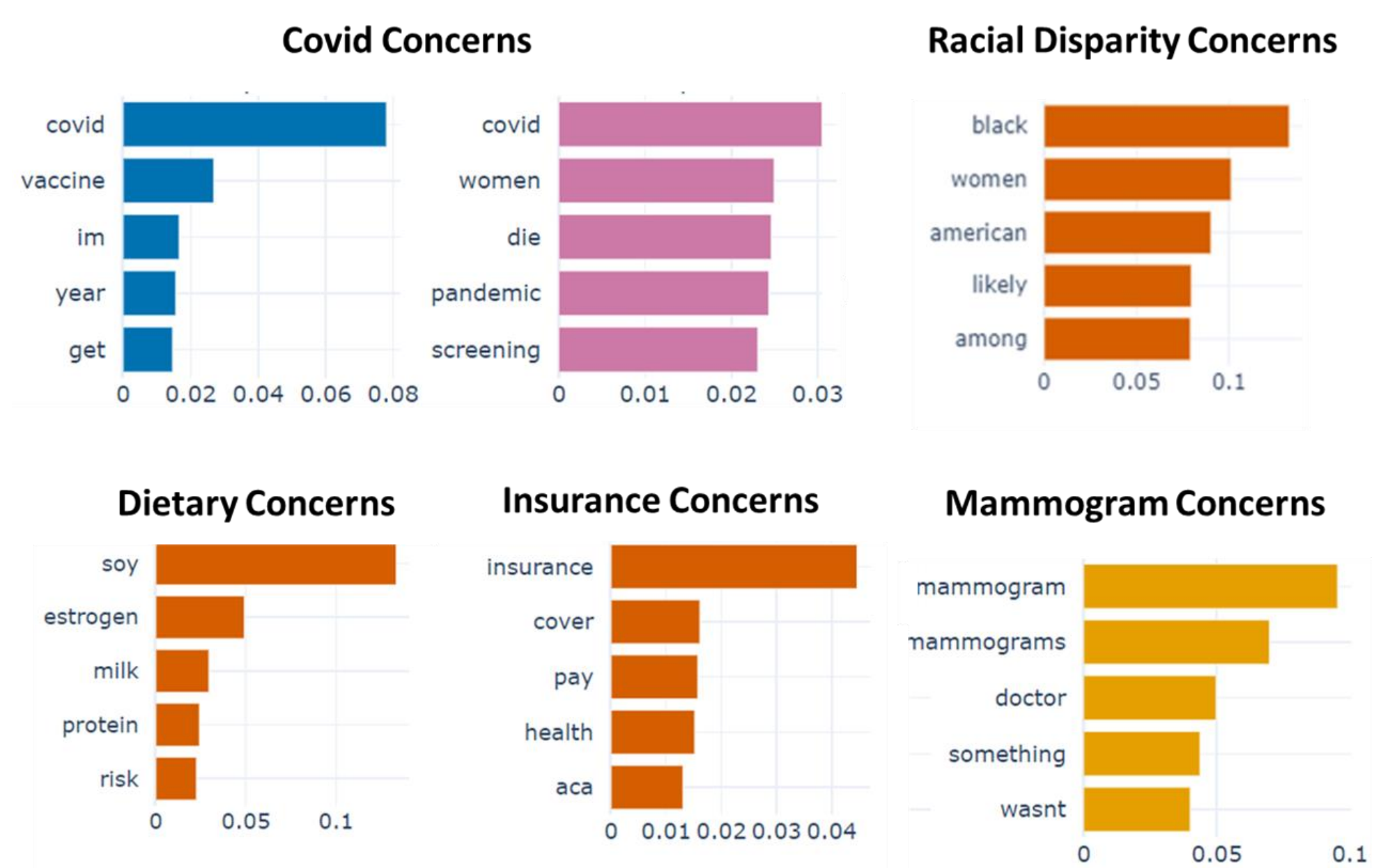
1,454,638 Breast Cancer (BC) Tweets



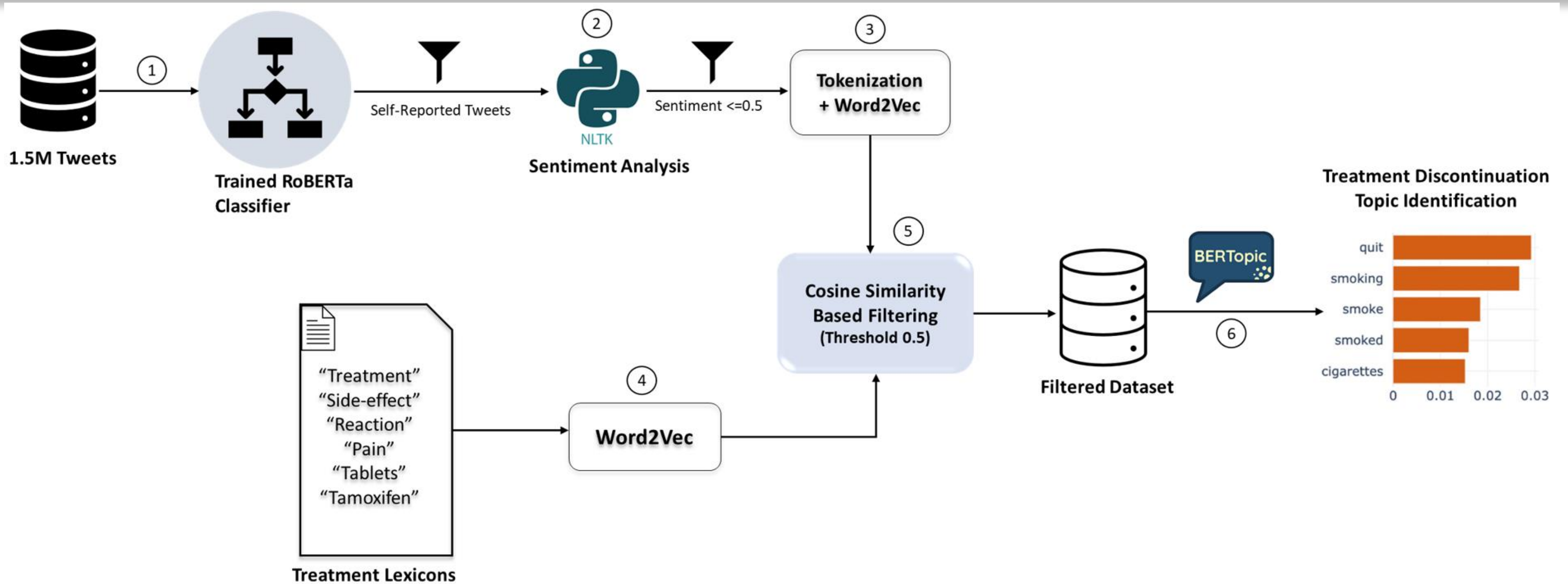
manually-annotated **3513** Tweets (self-reported BC or not)

RESULTS

Model name	f1_micro	f1_macro	f2_micro	f2_macro	logloss
Decision Tree	0.778	0.608	0.778	0.596	0.734
Log. Regression	0.772	0.576	0.772	0.570	0.464
Naive bayes	0.745	0.427	0.745	0.468	0.568
Random Forest	0.752	0.476	0.752	0.498	0.652
RoBERTa	0.894	0.853	0.894	0.841	0.332



METHODOLOGY



INLINE MEDIA

The New York Times: "Some Older Women Need Extra Breast Scans. Why Won't Medicare Pay?"

CBS News: "Breast cancer diagnosis for younger women leading to big medical debt"

CONCLUSION

- Transformer Based **RoBERTa** performed well for self-reported BC tweets identification
- Study covers broader **Societal, Psychological, and Medical** Concerns
- Insights useful for: Physicians, Policymakers, Breast Cancer Awareness Programs