

# Unveiling Voices: Identification of Concerns in a Social Media Breast Cancer Cohort via NLP

---

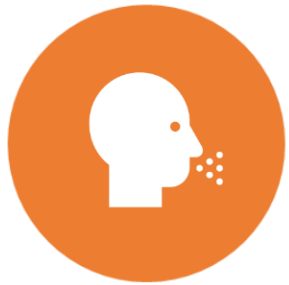
**Swati Rajwal\*** (CS PhD, Emory University)  
Avinash Kumar Pandey  
Zhishuo Han  
Abeed Sarker

# Learning Objectives



- NLP-framework to Identify Self-Reported Breast Cancer on Twitter
- Study Patient Reported Concerns related to Breast Cancer

# Social Media: A Health Data Resource!



EHRs capture health info,  
not daily habits/interests



Such “meta-data” are  
captured in patient-  
generated social media  
data<sup>1</sup>



There is chatter available on  
social media platforms like  
Twitter



Social media based NLP  
improve patient-centered  
outcomes in cohort studies

# Breast Cancer Cohort (1,454,638 Tweets)



## Keywords for Tweet Collection<sup>2</sup>

"cancer"  
"breastcancer" (one word)  
"tamoxifen"  
"survivor"  
and their hashtag equivalents



## Patterns for Tweet Filtering:

Tweet contains [#]breast & [#]cancer & [#]survivor; OR  
Tweet contains [#]breastcancer & #survivor; OR  
Tweet contains [#]tamoxifen AND ([#]cancer OR [#]survivor)  
Tweet contains a personal pronoun (e.g., 'my', 'I', 'me', 'us') AND  
[#]breast & [#]cancer

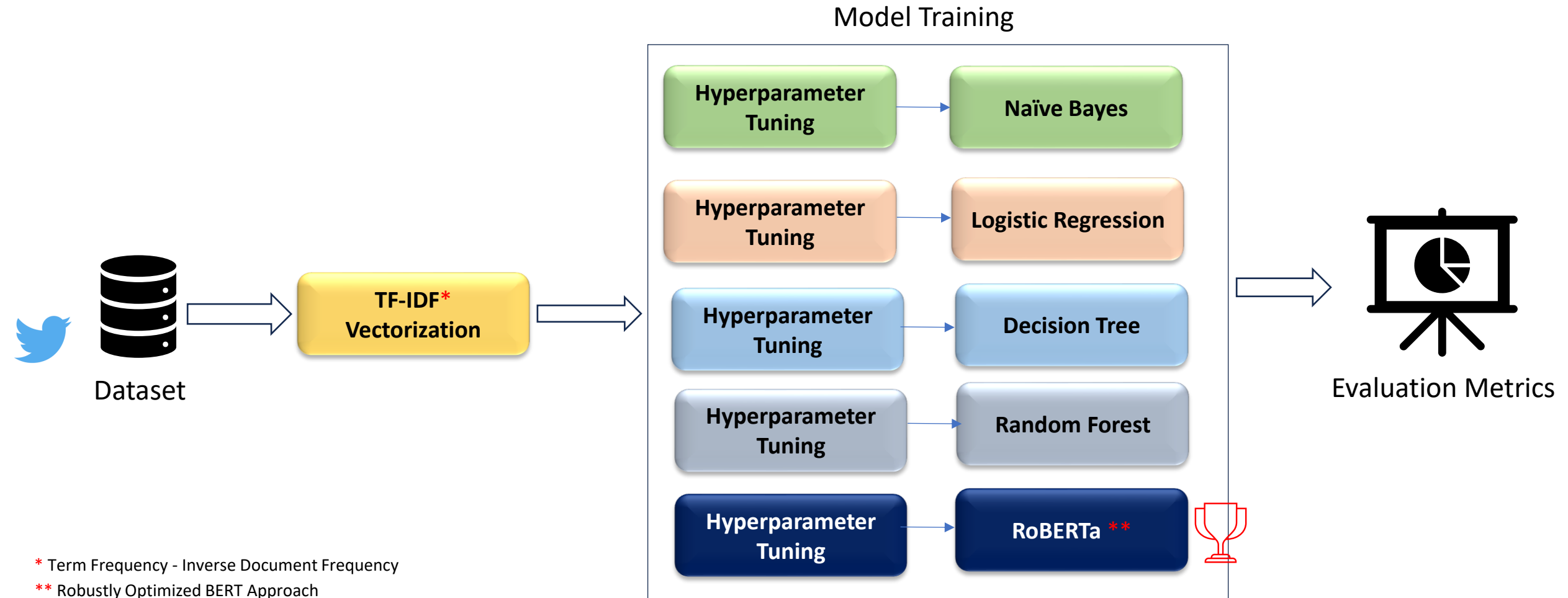


Patterns were developed by manual analysis of Twitter chatter

- N = 1,454,638 Breast Cancer (BC) Tweets
- manually-annotated data (if a Tweet is self-reported BC or not)
  - Train=3513,
  - Dev=302,
  - Test=1204

# Phase 1: Self Reported Breast Cancer Tweet Classifier

# Self Reported Breast Cancer Tweet Classifier



\* Term Frequency - Inverse Document Frequency

\*\* Robustly Optimized BERT Approach

# Evaluation Metrics Results on Test Set

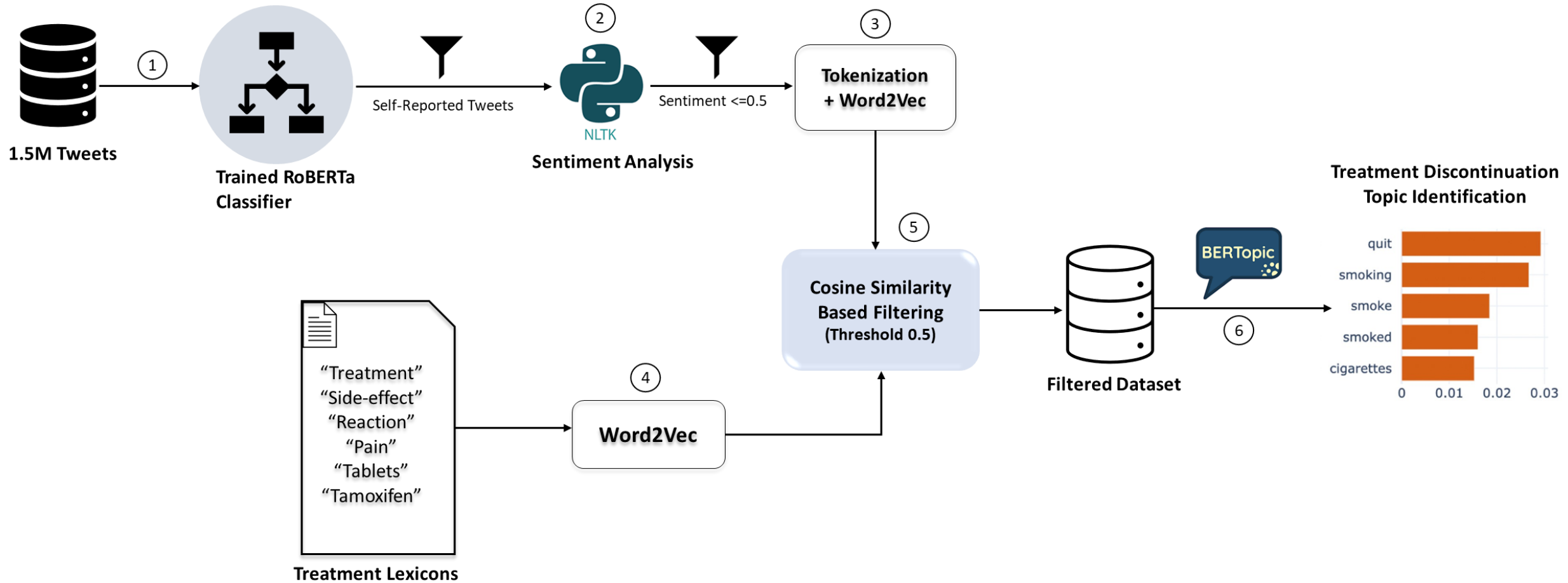
| Model name                 | Hyperparameters                    | f1_micro     | f1_macro     | f2_micro     | f2_macro     | logloss      |
|----------------------------|------------------------------------|--------------|--------------|--------------|--------------|--------------|
| <b>Decision Tree</b>       | criterion='gini',<br>max_depth=10  | 0.778        | 0.608        | 0.778        | 0.596        | 0.734        |
| <b>Logistic Regression</b> | C=10, penalty='l2'                 | 0.772        | 0.576        | 0.772        | 0.570        | 0.464        |
| <b>Naive bayes</b>         | alpha=0.1                          | 0.745        | 0.427        | 0.745        | 0.468        | 0.568        |
| <b>Random Forest</b>       | max_depth=None,<br>n_estimators=50 | 0.752        | 0.476        | 0.752        | 0.498        | 0.652        |
| <b>RoBERTa</b>             | epochs=7,<br>batch_size=16         | <b>0.894</b> | <b>0.853</b> | <b>0.894</b> | <b>0.841</b> | <b>0.332</b> |



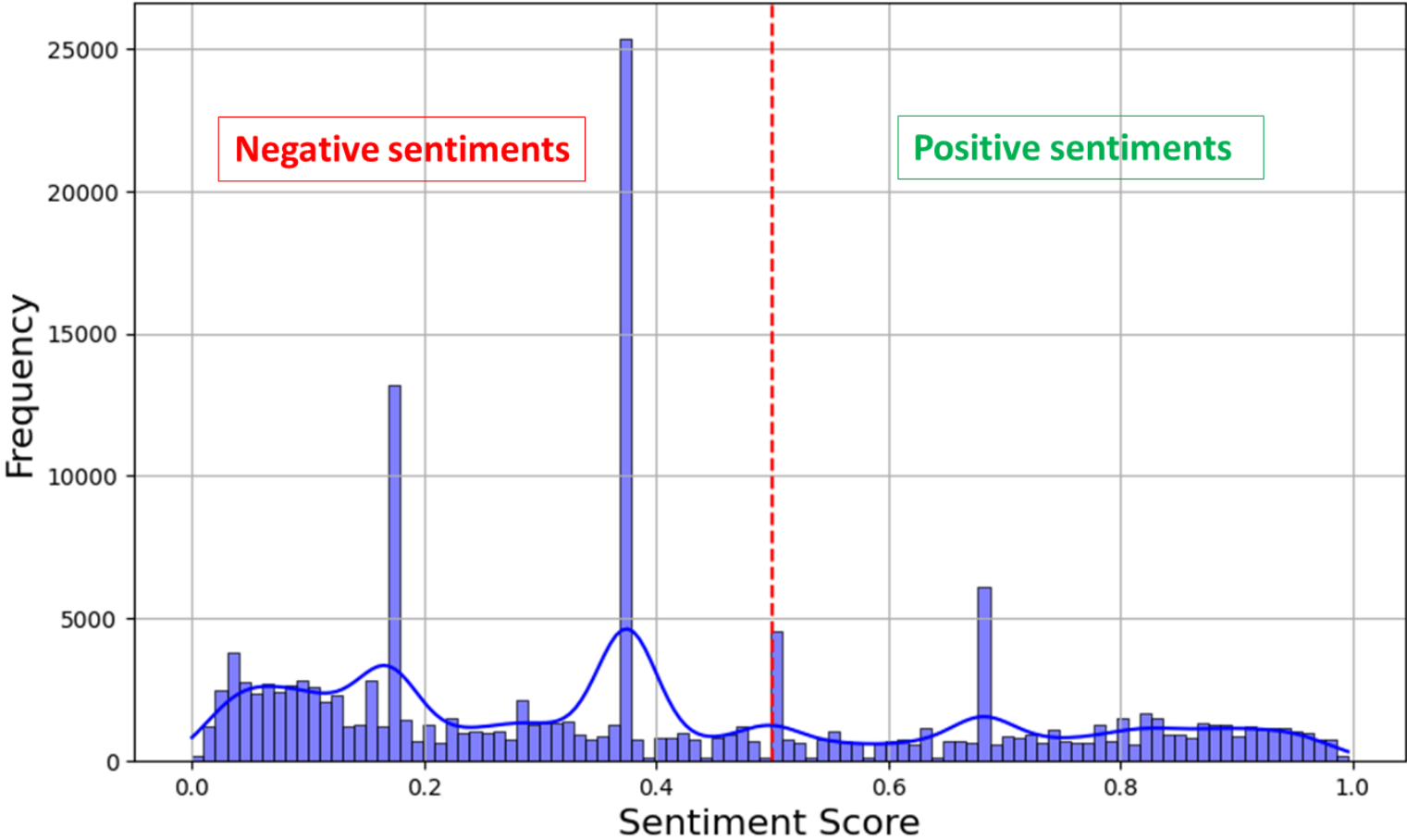


# Phase 2: Breast Cancer Cohort Concern Identification

# Methodology

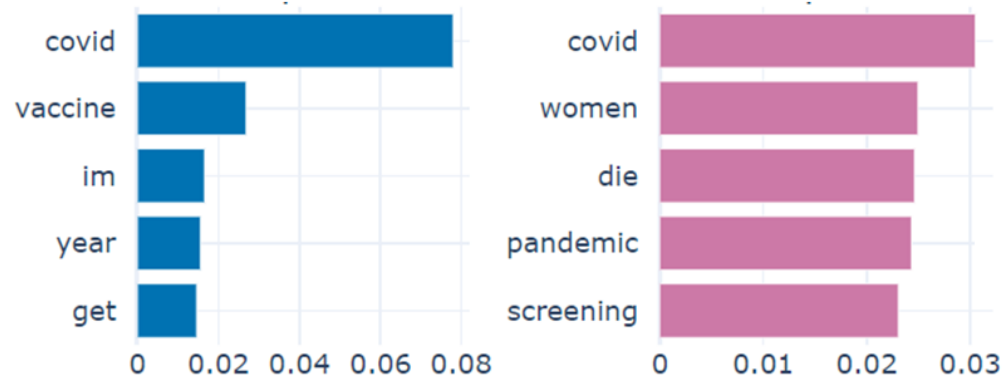


# Sentiment Score Distribution

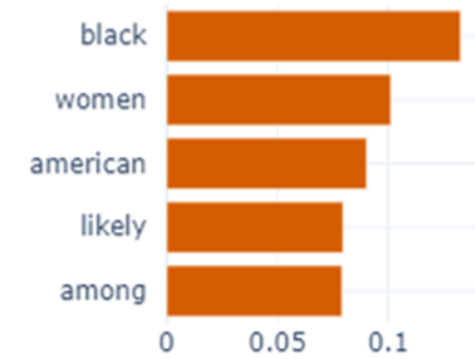


# Breast Cancer Concerns Topics

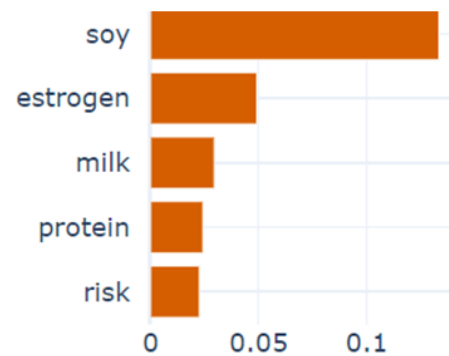
### Covid Concerns



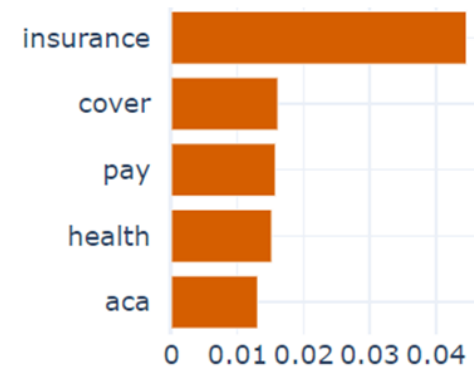
### Racial Disparity Concerns



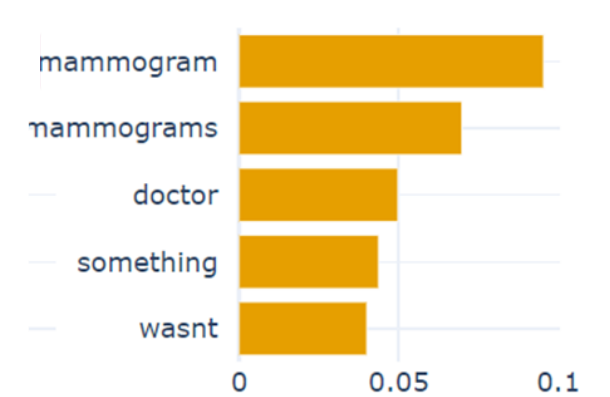
### Dietary Concerns



### Insurance Concerns



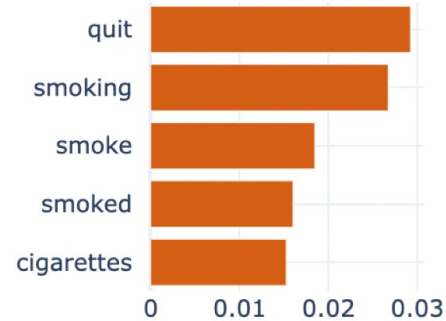
### Mammogram Concerns



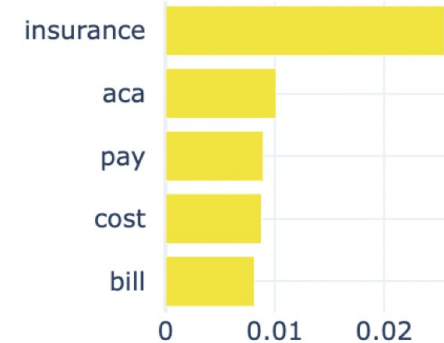
# Treatment Discontinuation Topics



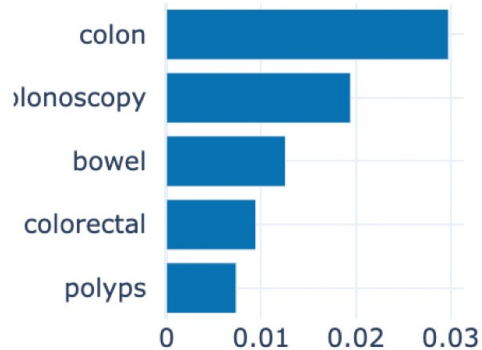
Hair/Baldness



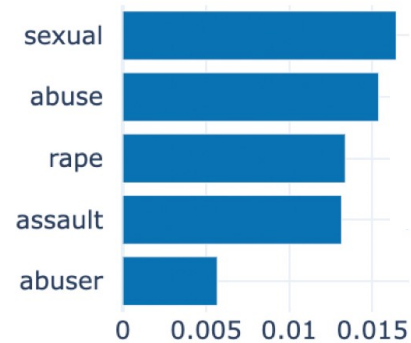
Smoking Restrictions



Insurance Issues



Bowel/Colon Issues



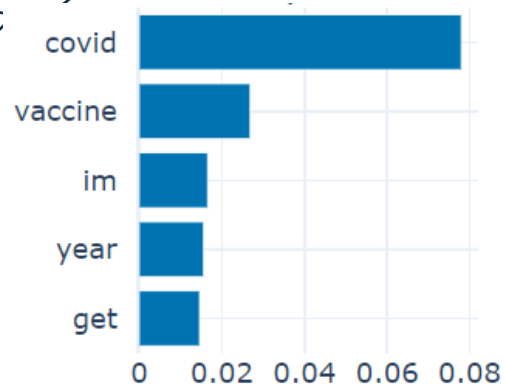
Sexual Abuse Concerns



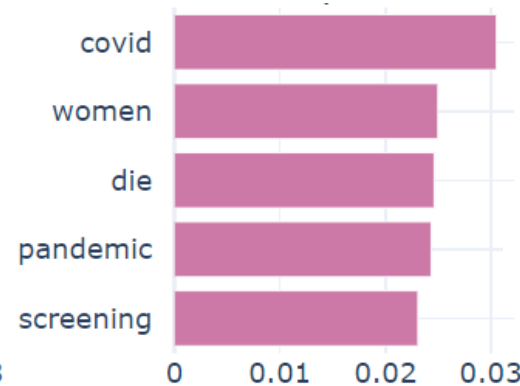
Mental/Emotional Health

# Results: Covid-19 affected Breast Cancer Patients

my **fear of being locked up again** is one of two contributors to my **refusal to seek** medical help for possible **breast cancer**.



**pandemic** leads women to **delay mammograms**; experts fear future rise in breast cancer deaths <url> via <user>



<user> always had a normal mammogram. this year she was scared to go in due to **covid**. she pushed it back a few months. when doctors found stage <number> **breast cancer**.

CDC [3]

## Studies Show Effects on Cancer Screening

Many preventive services, including cancer screenings, were put on hold because of the pandemic. Scientists in CDC's Division of Cancer Prevention and Control have done research to find out how the COVID-19 pandemic has affected cancer screening.

One [study](#) looked at whether the COVID-19 test positivity was linked to changes in cancer screening among women in CDC's National Breast and Cervical Cancer Early Detection Program. The study found that—

- During April 2020, breast and cervical cancer screening declined by 87% and 84%, respectively, compared with the previous 5-year average for that month.
- Some states were able to screen as many people during the pandemic as they did before the pandemic. Further studies will be done to find out how.

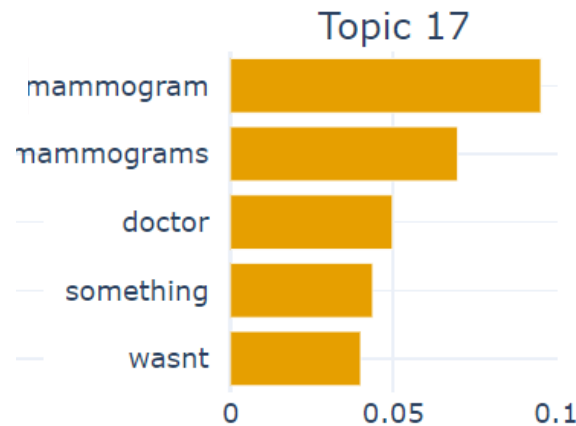
# Results: Anxiety associated with Mammogram Screening

thousands of women under <age> with **breast cancer** showing up on their **mammograms**. i think another huge issue is lumps being **fobbed off as hormonal changes** - natasha was only referred because she put her foot down



i had this **conversation with my mom** in the summer when she was **scared to get her mammogram**. she went and she did have breast cancer that fortunately **was caught early**.

same with **mammograms**. I am so **scared** to get them and i have <allcaps> to because my **family** has a **history of breast cancer**



also, this isn't doctors, it's the **insurance**, but i'm pissed that i can't get a **mammogram** when my **granny literally has breast cancer**..... they'll do an ultrasound but not a mammogram. the f\*\*k is that?

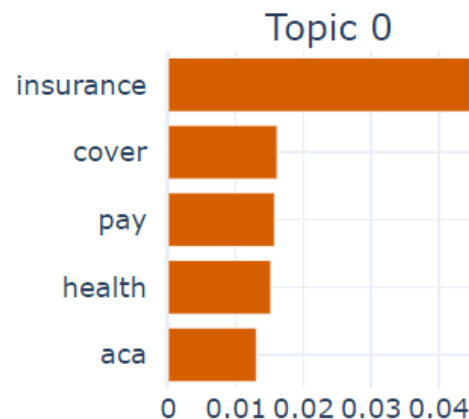
# Results: Concerns regarding Insurance Policies

the last thing a **breastcancer patient** needs to **worry about is if their insurance will cover treatment.** unfortunately, this is the reality for most facing this disease. help us eliminate barriers to care



I am <number> just **found a lump in my breast**, got a referral for a **mammogram** from my pcp, got a letter from **insurance** promptly <allcaps> telling me **,not all of this will be covered by us just fyi**, now i have to worry...& maybe have to pay out of pocket to find out

so i,am mad because **my insurance doesn't cover a well women, No exam nor a mammogram till i,am <number>** i mean **breast cancer** can affect women of all ages you don't think late teens deserve to check on their health.



im supposed to have **mammograms** on the reg bc of family history but **insurance barely or just doesn't pay for them.** im lucky to have insurance at all but I m terrified my broke ass is gonna get breast cancer; not be able to do shit about it



# Inline Media Discussions

The New York Times

## Some Older Women Need Extra Breast Scans. Why Won't Medicare Pay?

Mammography can miss tumors in women with dense breasts, so their doctors often include ultrasound or M.R.I. scans. Patients often wind up paying the bill.

Share full article



A mammography machine at a mobile cancer screening clinic in Lancaster, Calif. Sarah Reingewirtz/Media Services Group for The New York Times

April 2024, New York Times [6]

abc 10 NEWS WFTS TAMPA BAY

Watch Now

## Breast cancer diagnosis for younger women leading to big medical debt

Pinellas woman's mammogram not covered by insurance because of her age



Prev

Next

On Tuesday, Brittany McArthur snapped a picture in a hospital gown, still smiling despite the fact she was about to undergo a double mastectomy. In January, at just 30 years old, she was diagnosed with breast cancer. "It's not something I ever imagined to be facing this early in life," McArthur said, turning 31 in March.

April 2024, ABC [7]

A screenshot of the CBS News website. The top navigation bar includes "CBS NEWS" and "NEW YORK". Below the navigation bar, there is a "LOCAL NEWS" section with the headline "Some patients with Medicare are now being charged for breast ultrasounds, physicians say". The byline reads "By Zinnia Maldonado February 24, 2024 / 11:30 PM EST / CBS New York". Below the text is a video player showing a woman in a white lab coat looking at medical images on a computer monitor. The CBS News logo and "02 CBS NEWS NEW YORK" are visible in the bottom right corner of the video player.

NEW YORK -- Local physicians say patients are now noticing new charges following annual breast cancer screenings.

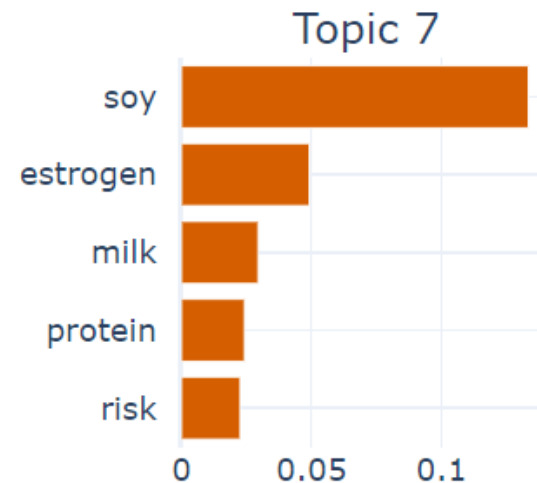
Feb 2024, CBS [8]

# Results: Confusions Regarding Dietary Habits

<user> <user> i'd like to switch to **soy milk**. i'm **worried** that it's been linked to **breast cancer** due to the presence of **estrogen**



proponents say that **soy** can help prevent heart disease, **breast cancer**, and more. detractors worry that soy might interfere with thyroid function and block nutrient absorption. **who's right?**



## Will eating soy increase my risk of breast cancer?

Answer From [Katherine Zeratsky, R.D., L.D.](#)

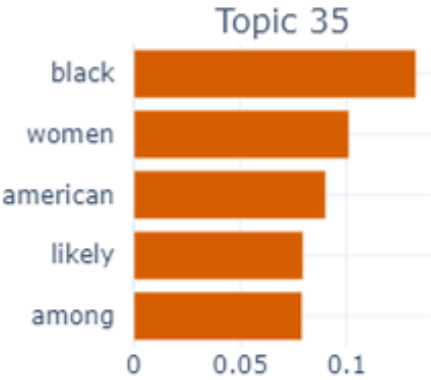
Studies show that a lifelong diet rich in soy foods reduces the risk of breast cancer in women. This protective effect is less dramatic for women who eat less soy or who start eating soy later in life. Soy contains protein, isoflavones and fiber, all of which provide health benefits.

Report by Mayo Clinic [5]

It was once thought that soy foods increase the risk of breast cancer. **However, eating a moderate amount of soy foods does not increase risk of breast cancer — or other types of cancer. A moderate amount is one to two servings a day of whole-soy foods, such as tofu, soy milk and edamame.**

# Results: Racial Disparity Concerns

*i visited the cdc website to learn about cancer disparities between **black and white american women**, and i was extremely shocked to learn about the development of **breast cancer** so present in young black women when it compares to young white women.*



***black women** have a <number>% higher death rate from breast cancer. we shouldn't be surprised that **communities with unequal health outcomes** have unequal trust in vaccines.*

***black women**, specifically young black women, are more susceptible to breast cancer. Don't be afraid to self-exam often!*

### Disparities in Breast Cancer Deaths

[Español \(Spanish\)](#) [Print](#)

Breast cancer deaths are decreasing in the United States overall, but disparities still exist.

**Breast cancer** is a leading cause of cancer death among American women. A [study](#) found that breast cancer deaths among women in the United States dropped 1% per year from 2014 to 2018. Another [study](#) found that non-Hispanic **Black women had a 41% higher death rate from breast cancer than non-Hispanic White women from 2011 to 2015.** CDC scientists wanted to know if these trends had changed.



You can help lower your breast cancer risk by taking care of your

CDC [4]



# Key Takeaways

- Transformer Based **RoBERTa** performed well for self-reported BC tweets identification
- This study captures broader
  - **Societal** Concerns (Racial, Sexual),
  - **Psychological** Concerns (Screening Anxiety, Pandemic, Complex Insurance Policies), and
  - **Medical** Concerns (Diet, Hair-loss, and other)
- Such Insights can be useful for:
  - Physicians
  - Policymakers
  - Breast Cancer Awareness Programs

# References



1. <https://pubmed.ncbi.nlm.nih.gov/31613225/>
2. [https://link.springer.com/chapter/10.1007/978-3-030-59137-3\\_10](https://link.springer.com/chapter/10.1007/978-3-030-59137-3_10)
3. <https://www.cdc.gov/cancer/dcpc/about/covid-19.htm>
4. <https://www.cdc.gov/cancer/dcpc/research/articles/disparities-breast-cancer-deaths.htm>
5. <https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/expert-answers/soy-breast-cancer-risk/faq-20120377>
6. <https://www.nytimes.com/2024/04/19/health/dense-breast-mammograms-medicare.html>
7. <https://www.abcactionnews.com/news/region-pinellas/breast-cancer-diagnosis-for-younger-women-leading-to-big-medical-debt>
8. <https://www.cbsnews.com/newyork/news/breast-ultrasound-cancer-screening-charge-medicare/>

# Thank You

<https://swati-rajwal.github.io/>



**Read More**

## Team Members



**Avinash Pandey**  
PhD Student



**Zhishuo Han**  
PhD Student



**Abeed Sarker**  
Associate Professor