Massachusetts Health Quality Partners

Together for good measure

Using NLP Software to Classify CG-CAHPS Patient Comments in Massachusetts

October 7, 2021

Disclosures

We have no conflicts of interest, financial interest, or sponsorships relevant to this activity to disclose.

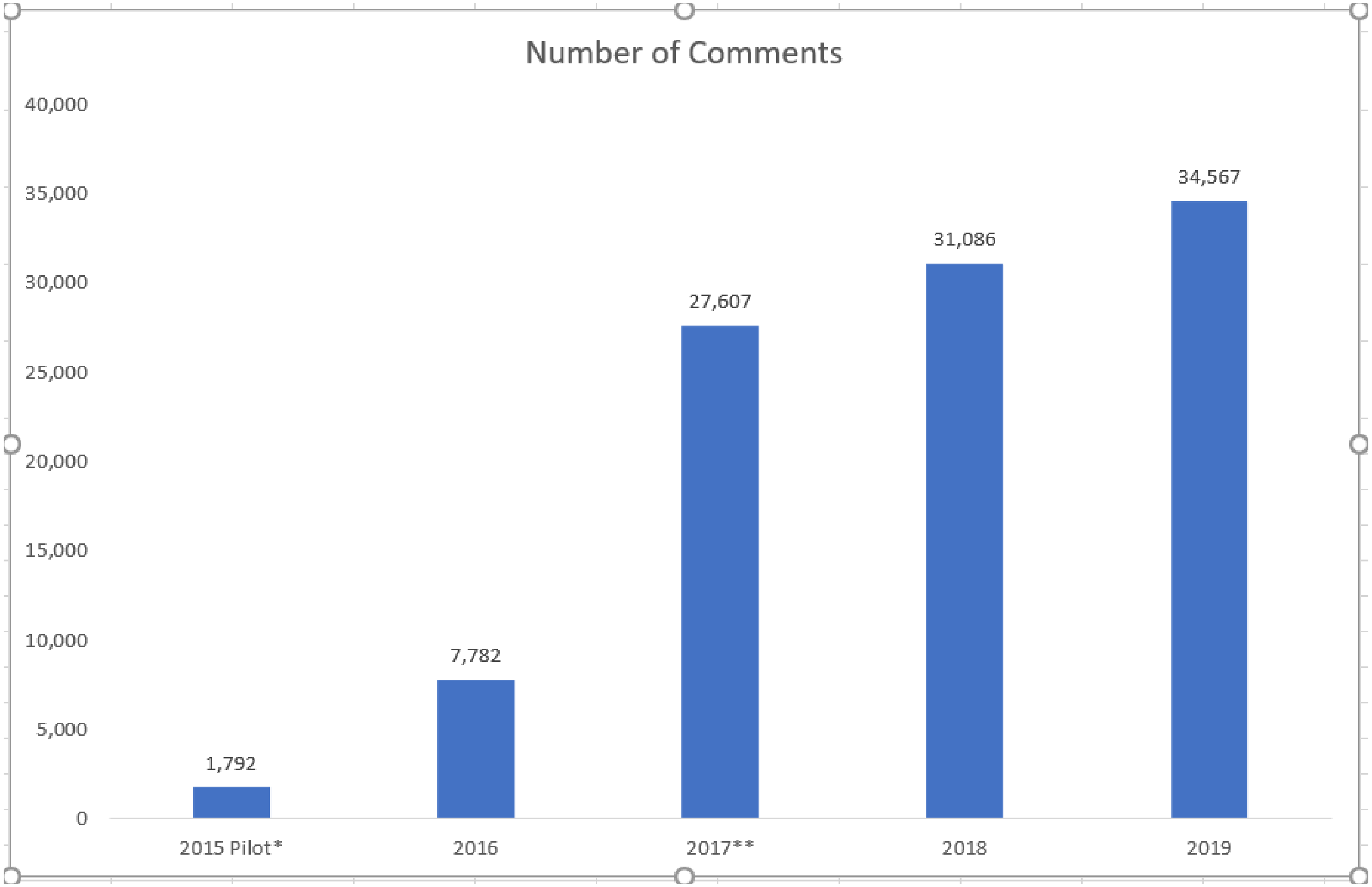
Background/Impetus for Collecting Patient Comments

- In 2015, MHQP held a workgroup series with its stakeholders to develop recommendations for modernizing our commercial Patient Experience Survey (PES) program
- One main area of focus was to enhance the richness and value of survey data by adding patient narratives to privately reported survey results
- MHQP began collecting patient comments in a large-scale pilot project conducted with the California Healthcare Performance Information System (CHPI), funded by the Center for Healthcare Transparency. We field tested two different sets of open-ended questions, a 5-item and 3-item protocol respectively
- MHQP decided to incorporate the 5-item protocol which became known as the CAHPS Patient Narrative Item Set into its statewide patient experience survey, taken online or through email invitation

Project Overview and Goals

- From 2016 to 2019 MHQP shared comments with participating provider organizations and reported them in various ways
- Over time, MHQP recognized that the collection of large numbers of narratives requires some type of processing to efficiently extract the meaning of the information emphasized in the comments
- We understood the need to report patient narrative information back to provider organizations in a way that was easy, effective and actionable for improvement purposes
- In 2020, MHQP stakeholders requested that the team work on categorizing the comments by composite, which would allow provider organizations to compare patient comment feedback with the composite quantitative scores they received for that given survey year
- We needed to find a Natural Language Processing Tool to classify comments

Total Number of Unique/Individual Comments by Year (Adult & Pediatric)



*Collection of Adult Survey Only
**Revised introduction to open-ended questions

Sample Comments

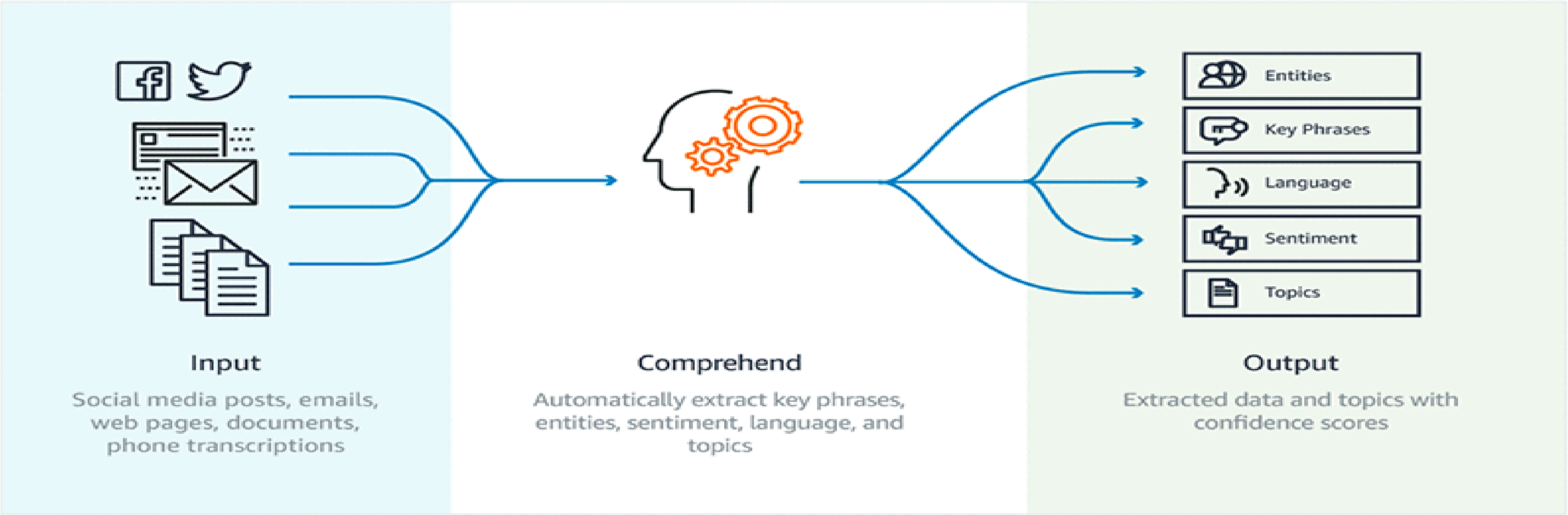
- *“Dr. T has been my primary physician for a **very long time**. She has always been there for me when I need medical attention. She also asks about and cares about my **stress level and mental health**. She is a wonderful doctor because she works with the **whole patient mind body and soul**. She is truly an outstanding doctor, and all of her patients are very fortunate to have her in their corner. She is a true gem in her profession.”*

- *“I called my provider to ask for a specific blood test. Staff was **respectful listened and coordinated** an appointment as soon as possible. **Provider listened** to why I asked for the appointment and blood test and was also **respectful** and **ordered the test immediately**.”*

Amazon Comprehend Tool

- Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents.
- Amazon Comprehend processes any text file for machine training – Requires two columns. One is “Classification” and Two is “Actual Comment”
- It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in a document.

Source: https://console.aws.amazon.com/comprehend/v2/home?region=us-east-1#welcome



Train Machine | Predict Patient Narratives | Distribute the Feedback

Process of Coding, Classification and Machine Learning

Adult Composites	Child Composites
Communication	Communication
Integration of Care	Integration of Care
Knowledge of Patient	Knowledge of Patient
Self-Management	Self-Management
Adult Behavioral Health	Pediatric Preventive Care
Organizational Access	Organizational Access
Office Staff	Office Staff
	Pediatric Development

The machine learning tool uses the trained file to predict the classification for a given narrative.

Inclusion Criteria: In the initial iterations of training the machine, the following inclusion criteria were implemented:

- Comments with 150 or more characters per narrative
- Limited to CAHPS Narrative Item Set Q2-Q6. We excluded Q1 comments as these narratives reflect expected "preferences" rather than actual "experiences"

Terminology used:

- When referring to the manual process of coding the comments, we use the word "code"; when referring to the process of Amazon Comprehend classifying comments, we use the word "classify" (i.e., code=manual; classify=machine)
- Predicted file = machine output, Training file = manually coded

Process of Coding

Coding and Classification Steps:

Team manually coded 300 adult comments and 300 child comments. Coded comments were run through NLP for machine learning.

- The team coded more comments and adapted various comments to meet the 50-comment threshold per composite classification. Once the minimum was met, the comments were run through NLP for machine learning. The team reviewed the output.

▪ Repeated the iterations of training the machine multiple times to improve the machine learning and patient narratives classification process for both Adult/Pediatric comments.

Confusion Matrix

- When a custom classifier model is **trained**, Amazon Comprehend creates a **confusion matrix output file** that provides metrics on how well the model performed in training.
- This enables you to assess how well the classifier will perform when run.
- This matrix shows a matrix of labels as predicted by the model compared to actual labels and is created using **10 to 20 percent** of the documents submitted to test the trained model.

Confusion Matrix

Adult Confusion Matrix - Iteration 4		
[3, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0]	Adult Behavioral Health	50.00%
[2, 30, 2, 4, 6, 3, 2, 0, 0, 0, 0]	Communication	61.22%
[0, 3, 8, 0, 0, 2, 1, 0, 0, 0, 0]	Integration of Care	57.14%
[0, 2, 2, 18, 1, 1, 1, 0, 0, 0, 1]	Knowledge of Patient	69.23%
[0, 2, 0, 2, 42, 5, 1, 0, 0, 0, 0]	Not a Composite	80.77%
[0, 4, 1, 0, 2, 20, 3, 0, 0, 0, 0]	Office Staff	66.67%
[0, 1, 3, 0, 2, 1, 20, 0, 1, 0, 0]	Organizational Access	71.43%
[1, 0, 0, 0, 1, 0, 1, 3, 0, 0, 0]	Referral	50.00%
[0, 2, 0, 2, 0, 0, 0, 0, 8, 0, 0]	Self-Management Support	66.67%
[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	Technology	0.00%
[0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0]	Trust	0.00%

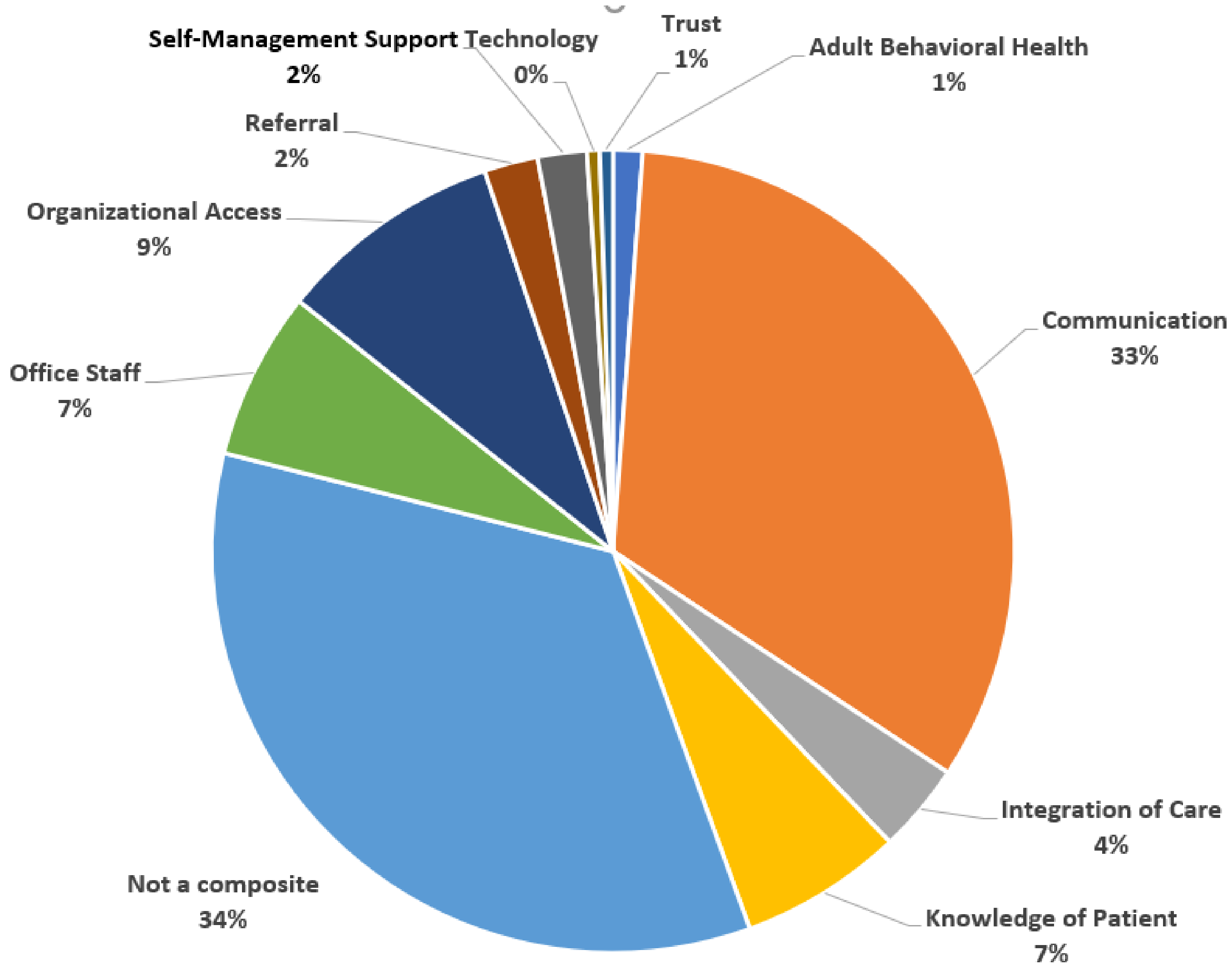
- After several iterations, we have improved our current training model accuracy.
- We had an internal target of 50% or higher.

NLP Prediction Output

- Amazon Comprehend **assigns** a weight to predicted classifications for every comment.
  - The weights represent a probability distribution over the words in each classification. Since we have more than 10 classifications, they may not sum to 1.
- In many cases, comments are assigned multiple classifications. For example, a comment may mention communication with the provider, issues about office staff, and access challenges. For each comment, the composite classification with the highest weight assigned is selected.

Comment	Predicted Classifications
Comment 1	Communication/0.9854; Integration of Care/0.0042; Office Staff/0.0024
Comment 2	Organizational Access/0.5562; Communication/0.3917; Referral/0.0181

PES 2019 Adult Comments Predicted Classification



Current Work and Next Steps

- Restructuring patient comment reports
  - Comments classified into MHQP’s traditional survey composites

- Potential inclusion of new domains found in patient narratives
  - Trust
  - Referrals
  - Technology
  - New telehealth domain to align with 2021 PES instrument (based on CAHPS Clinician & Group Visit Survey)

- Continuous training to improve accuracy and prediction

Questions