

Collecting COVID-19 Mask Mandate Information Using Artificial Intelligence to Contextualize 2022 NAEP Results

2023 AERA Annual Meeting

April 15, 2023 | Chicago, Illinois

Sadaf Asrar

Director - Data Science

 **OPTIMAL**
SOLUTIONS GROUP, LLC

Optimal Solutions Group, LLC: Who We Are



Innovative, Non-Partisan,
Rigorous Public Policy
Research Firm
UMD Research Park



International Experience in
40+ countries



Real-Time Framework for
Public Policy Decision
Making

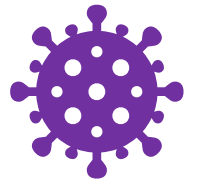
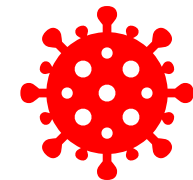
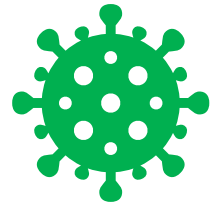
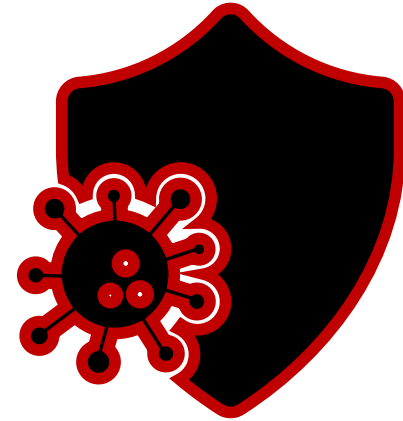
Working with the National Center for Education Statistics

- Since the early 2000s, have worked with the National Center for Education Statistics (NCES) to support statistical analysis, reporting, and dissemination activities for the National Assessment of Educational Progress (NAEP).
- Administered at the national, state, and district levels, NAEP measures what students know and can do in the United States (U.S.).



COVID-19 Pandemic and K-12 Education

- The COVID- 19 pandemic has disrupted all aspects of life.
- Particularly affected are children with regards to access to and quality of education.
- Disruptions have taken place in the form of school closures and the transition to remote or hybrid learning.
- One way schools in the U.S. have decided to manage COVID-19 outbreaks is through imposing mask mandates.



School District Level Mask Mandates

- Many states that did not have a universal mask mandate at the state level allowed districts to adopt their own masking policies.
- Similarly, many school districts in states with mask mandate bans adopted mask mandates.

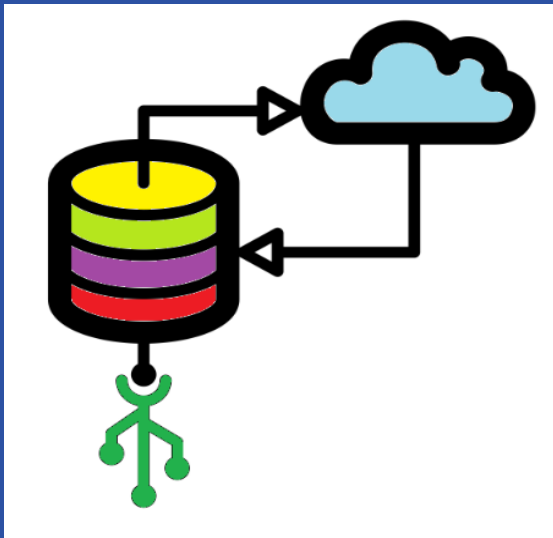


School District Level Mask Mandates

- To better understand the prevalence of masking in schools, Optimal:
 - Developed a robotic process automation tool to query and collect information on mask mandates at the school district level across the U.S. using the Google Search Engine;
 - Developed a validation code to filter out irrelevant search results; and
 - Trained supervised and unsupervised Natural language Processing (NLP) machine learning algorithm to classify and cluster the search results respectively.
- The data will be used to contextualize and help understand the performance results from the 2022 NAEP administration.



Data Collection



Google

🔍 Bexley City School District Mask Mandate Ohio



Google Search

I'm Feeling Lucky

BeautifulSoup



bexley city school district ohio mask mandate



[All](#) [News](#) [Images](#) [Shopping](#) [Maps](#) [More](#) [Tools](#)

About 497,000 results (0.60 seconds)

<https://www.bexleyschools.org> > protected > ArticleView

Our Masking Protocol for the 2021-22 School Year - Bexley ...

Additionally, the **City of Bexley** updated its guidance to recommend that all individuals wear **masks** when in indoor **public** spaces on July 30th. Ensuring students ...

<https://www.bexleyschools.org> > protected > ArticleView

Bexley City School District Provides Update Regarding ...

To assist families, the **District** will purchase two **Bexley City Schools**-branded 3-ply **masks** made from non-woven, high-efficiency filter material. These **masks** ...

<https://www.nbc4i.com> > news > local-news > bexley-ci...

Bexley City Schools drops mask mandate | NBC4 WCMH-TV

Mar 1, 2022 — In a unanimous vote during a **school board** meeting Tuesday, **masks** will now be optional starting Friday.

School District Level Mask Mandates



bexley city school district ohio mask mandate



All News Images Shopping Maps More Tools

About 497,000 results (0.60 seconds)

<https://www.bexleyschools.org> > protected > ArticleView

[Our Masking Protocol for the 2021-22 School Year - Bexley ...](#)

Additionally, the [City of Bexley](#) updated its guidance to recommend that all individuals wear **masks** when in indoor **public** spaces on July 30th. Ensuring students ...

<https://www.bexleyschools.org> > protected > ArticleView

[Bexley City School District Provides Update Regarding ...](#)

To assist families, the **District** will purchase two **Bexley City Schools**-branded 3-ply **masks** made from non-woven, high-efficiency filter material. These **masks** ...

<https://www.nbc4i.com> > news > local-news > bexley-ci...

[Bexley City Schools drops mask mandate | NBC4 WCMH-TV](#)

Mar 1, 2022 — In a unanimous vote during a **school board** meeting Tuesday, **masks** will now be optional starting Friday.

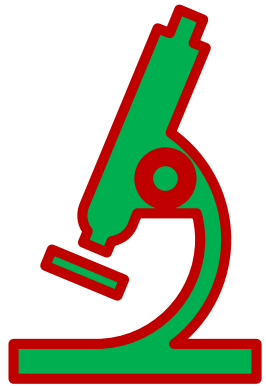
School District Level Mask Mandates

School District Level Mask Mandates

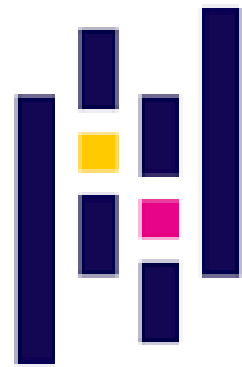
	A	B	C	D	E
1	# District	Title	Description	URL	
2	1	Bexley City	Our Masking Protocol for the 2021-22 School Year	Additionally, the City of <i>Bexley</i> updated its guidance to recommend that all individuals wear <i>masks</i> when in indoor <i>public</i> spaces on July 30th. Ensuring students ...	https://www.bexleyschools.org/protected/ArticleView.aspx?iid=52BYG2&dasi=3AP2
3					
4					

- Optimal created a Python program to scrape this information using a list of school districts in the U.S.
- The Python program used the BeautifulSoup package to extract this information by specifying the Cascading Style Sheets (CSS) selectors of the Google Search Engine.

Data Validation



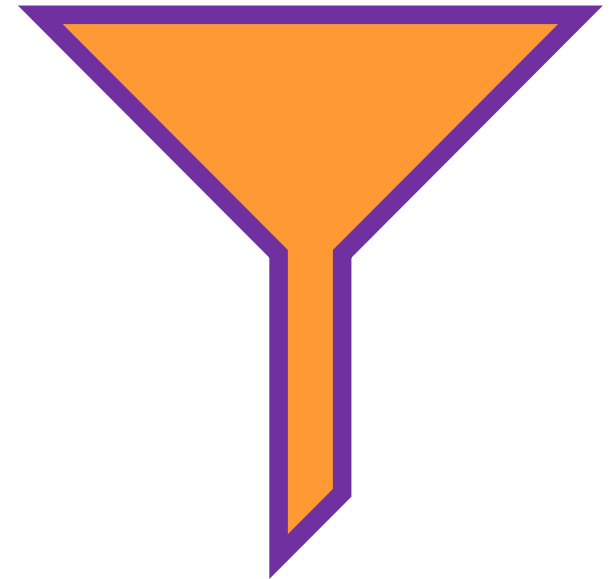
ANACONDA®



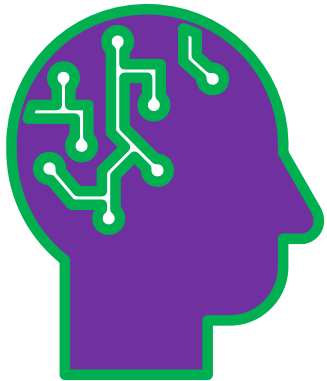
pandas

Filtering Out Irrelevant Results

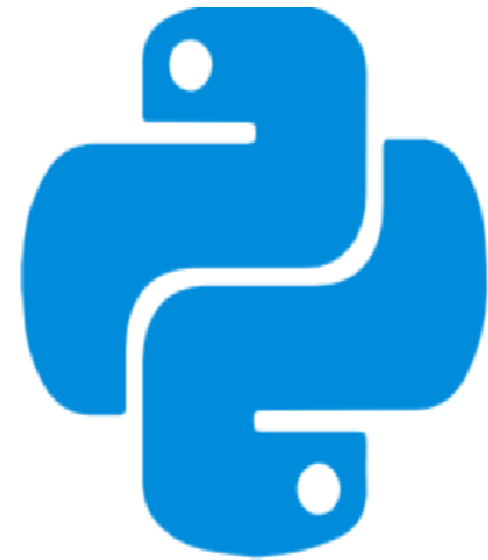
- Optimal performed validation checks following categories:
 - Result is for the correct school district;
 - Result is for the correct state;
 - Result has information on mask mandates;
 - Result is from a valid source; and
 - Result has a valid date.
- Optimal manually reviewed and labeled the validated Google Search results data from Michigan, Vermont, and Alaska into three categories:
 - No Mandate | Mask Mandate | Recently Repealed Mask Mandate.
- The total number of records was 834.



Text Classification

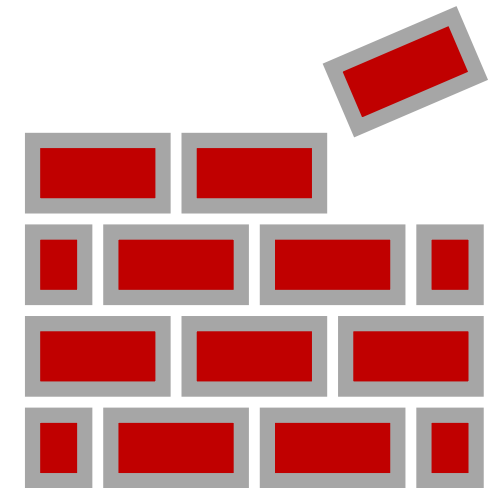


ANACONDA®



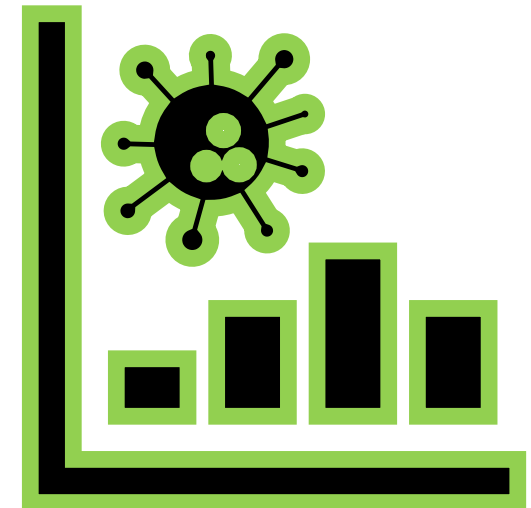
Feature Engineering

- Optimal implemented the following feature engineering techniques to convert the free text Google Search results data into vectors for training the algorithm:
 - Count vectors;
 - TF-IDF at the word, N-Gram, and character levels; and
 - Text / NLP based features.
- The raw text was also normalized and stop words were removed for data processing.

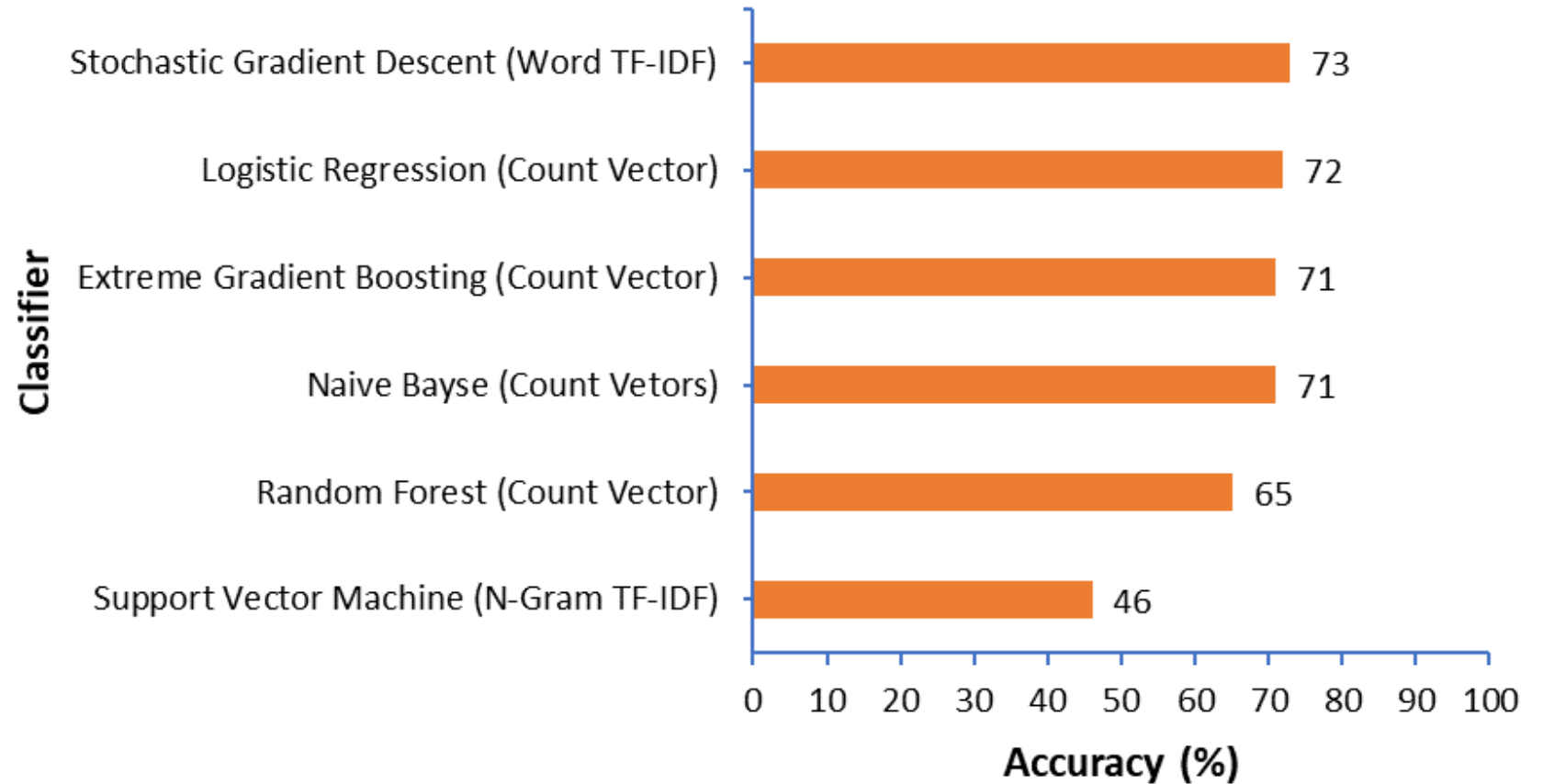


Natural Language Processing Model

- Optimal developed the following text classification frameworks to train the NLP model:
 - Support Vector Machine;
 - Naive Bayes;
 - Logistic Regression;
 - Random Forest;
 - Stochastic Gradient Decent; and
 - Extreme Gradient Boosting.

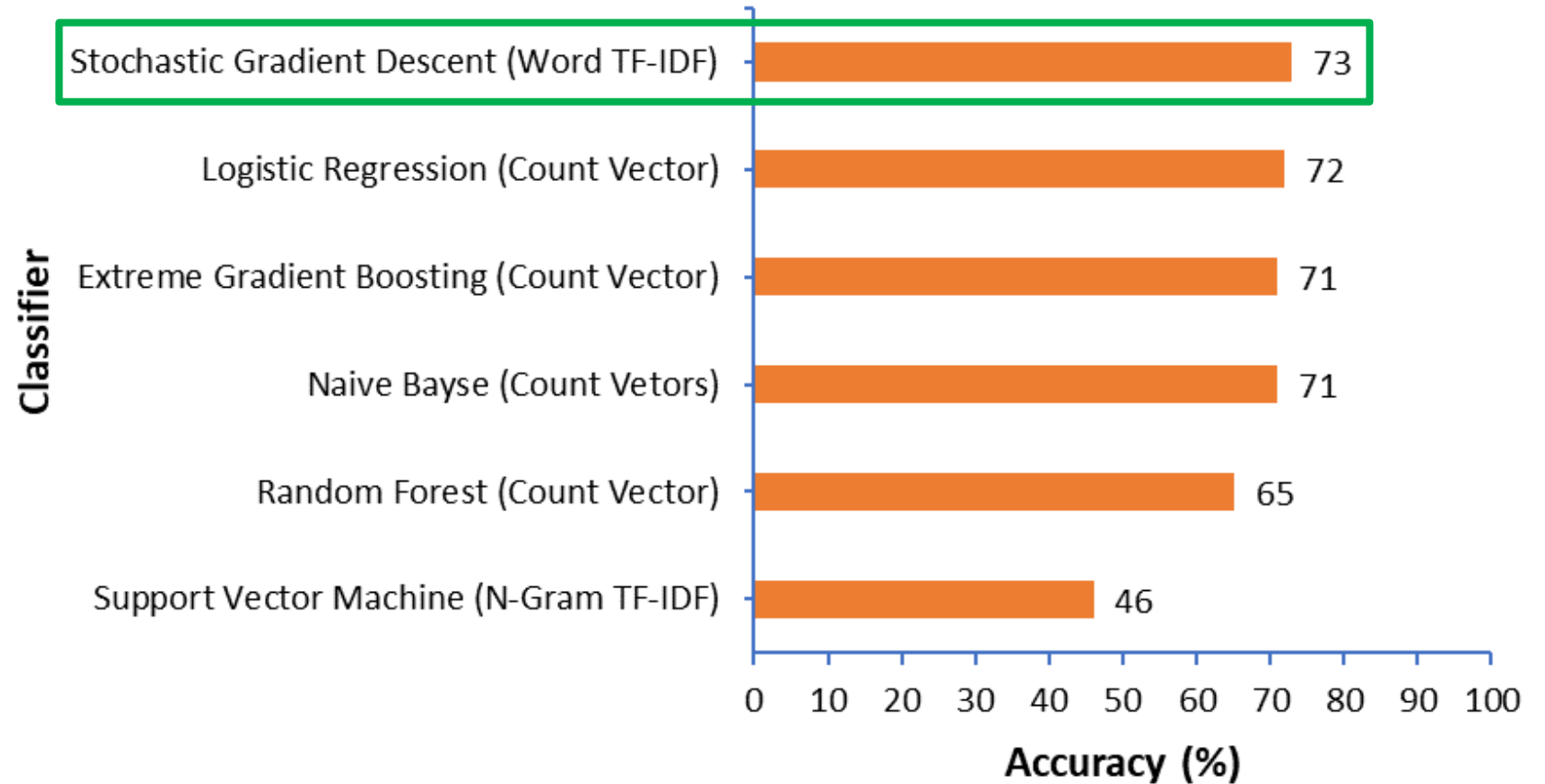


Model Performance



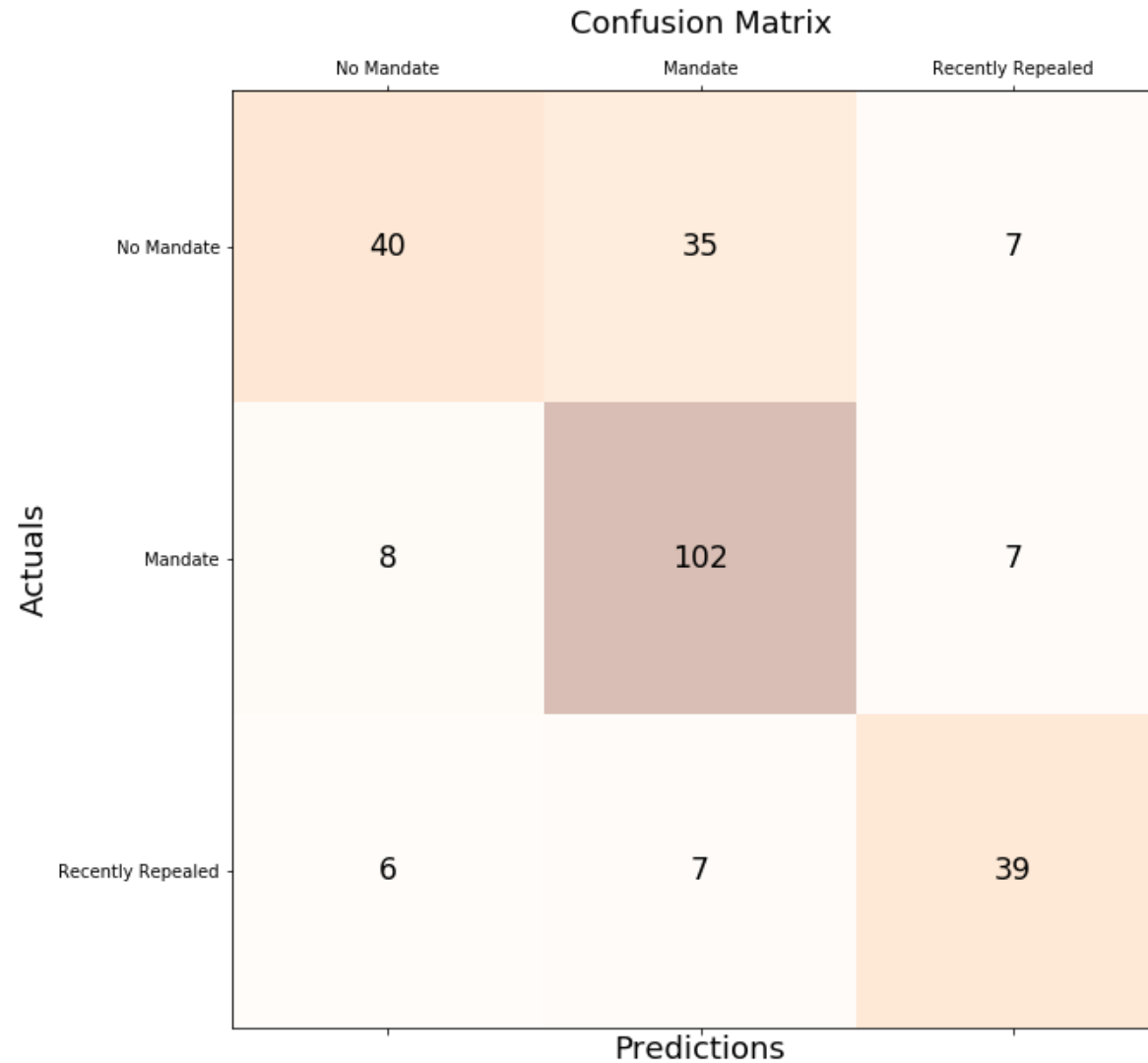
NOTE: The feature engineering technique used to process the data for each classifier is listed in parenthesis.

Model Performance



NOTE: The feature engineering technique used to process the data for each classifier is listed in parenthesis.

Final Model Results



- Overall accuracy of the model on the test data was 72%.
- Prediction accuracy for:
 - Mask Mandate - 87%;
 - Recently Repealed -75%; and
 - No Mandate - 49%.

Final Model Results

Confusion Matrix

		Predictions		
		No Mandate	Mandate	Recently Repealed
Actuals	No Mandate	40	35	7
	Mandate	8	102	7
	Recently Repealed	6	7	39

- Overall accuracy of the model on the test data was 72%.
- Prediction accuracy for:
 - Mask Mandate - 87%;
 - Recently Repealed -75%; and
 - No Mandate - 49%.

Final Model Results

Confusion Matrix

		Predictions		
		No Mandate	Mandate	Recently Repealed
Actuals	No Mandate	40	35	7
	Mandate	8	102	7
	Recently Repealed	6	7	39

- Overall accuracy of the model on the test data was 72%.
- Prediction accuracy for:
 - Mask Mandate - 87%;
 - Recently Repealed -75%; and
 - No Mandate - 49%.

Final Model Results

Confusion Matrix

		Predictions		
		No Mandate	Mandate	Recently Repealed
Actuals	No Mandate	40	35	7
	Mandate	8	102	7
	Recently Repealed	6	7	39

- Overall accuracy of the model on the test data was 72%.
- Prediction accuracy for:
 - Mask Mandate - 87%;
 - Recently Repealed -75%; and
 - No Mandate - 49%.

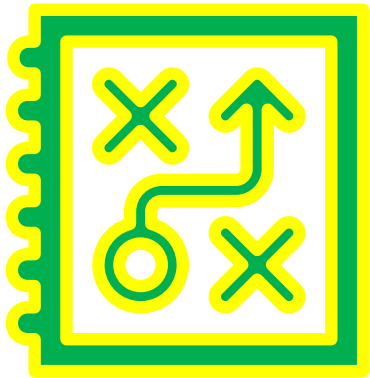
Challenges for Classification



Title: *Flint school district requires masks for fall classes, then ...*

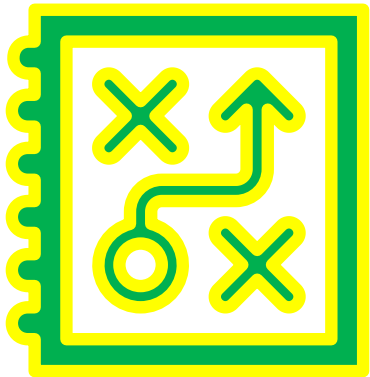
Description: *7 days ago, masks are now merely recommended, not required, in the district that gave every employee a \$22k covid bonus.*

Next Steps



- Increase the size of the training data by:
 - Manually reviewing and labelling the search results for a few additional states.
- Train additional machine learning and deep learning models.
- Explore unsupervised or self-supervised approaches such as:
 - Latent Dirichlet Allocation (LDA);
 - Word embeddings using word2vec; and
 - Bidirectional Encode Representations from Transformers (BERT).

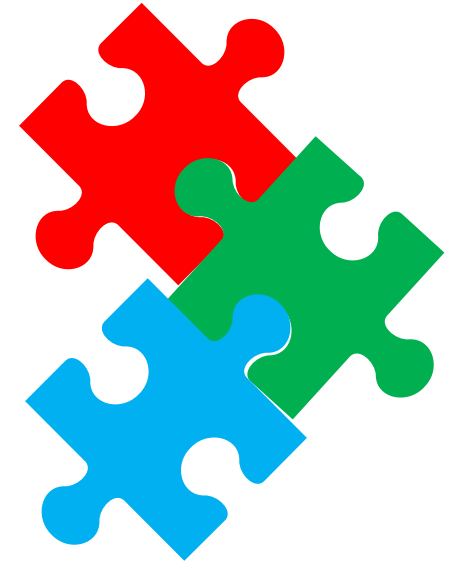
Next Steps



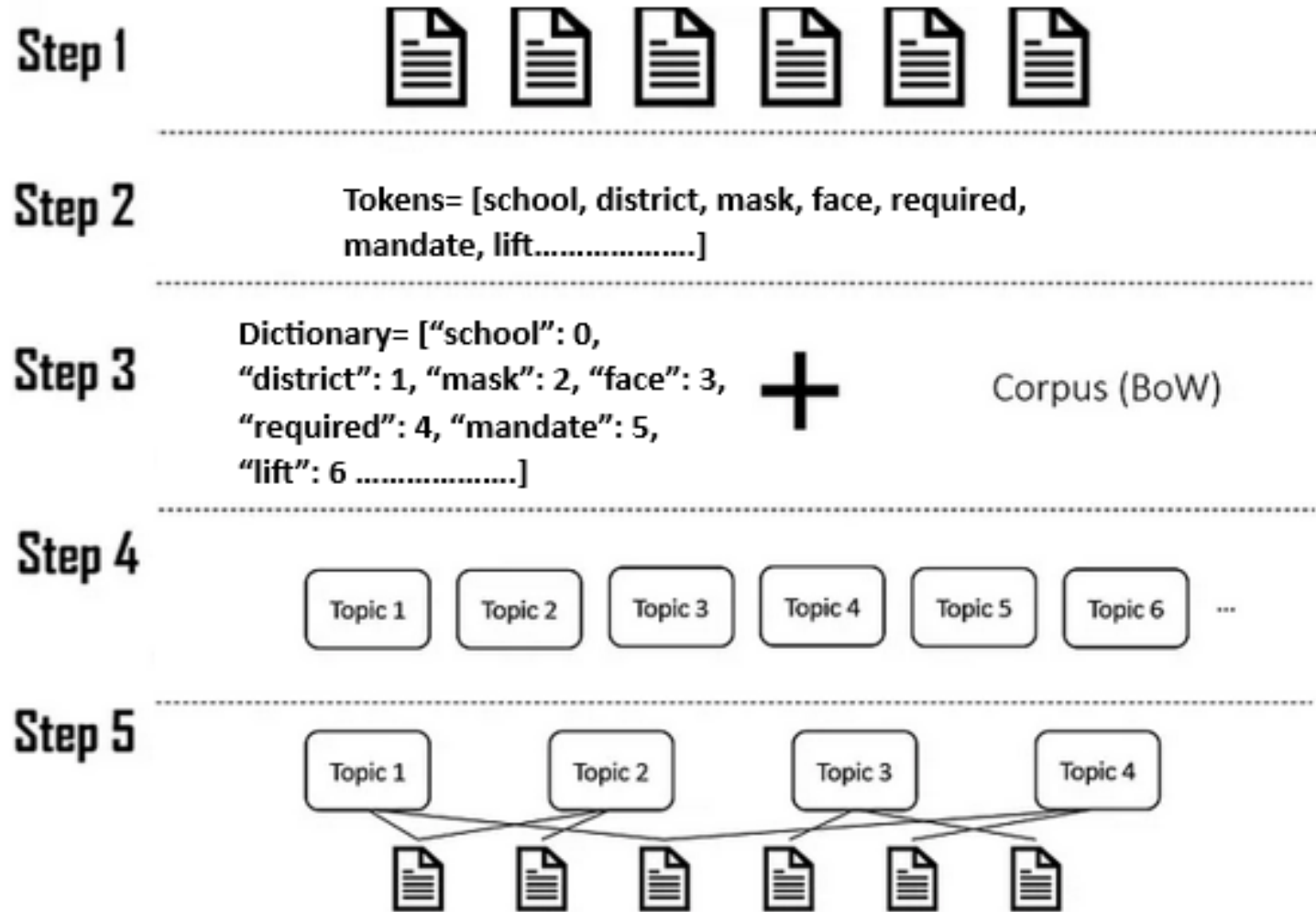
- Increase the size of the training data by:
 - Manually reviewing and labelling the search results for a few additional states.
- Train additional machine learning and deep learning models.
- Explore unsupervised or self-supervised approaches such as:
 - **Latent Dirichlet Allocation (LDA);**
 - **Word embeddings using word2vec;**
and
 - Bidirectional Encode Representations from Transformers (BERT).

Latent Dirichlet Allocation (LDA)

- LDA is a generative probabilistic model that shares similarities with Naive Bayes.
- LDA represents topics as word probabilities.
- It allows for uncovering hidden or latent topics when it clusters words based on their co-occurrence in a given document.

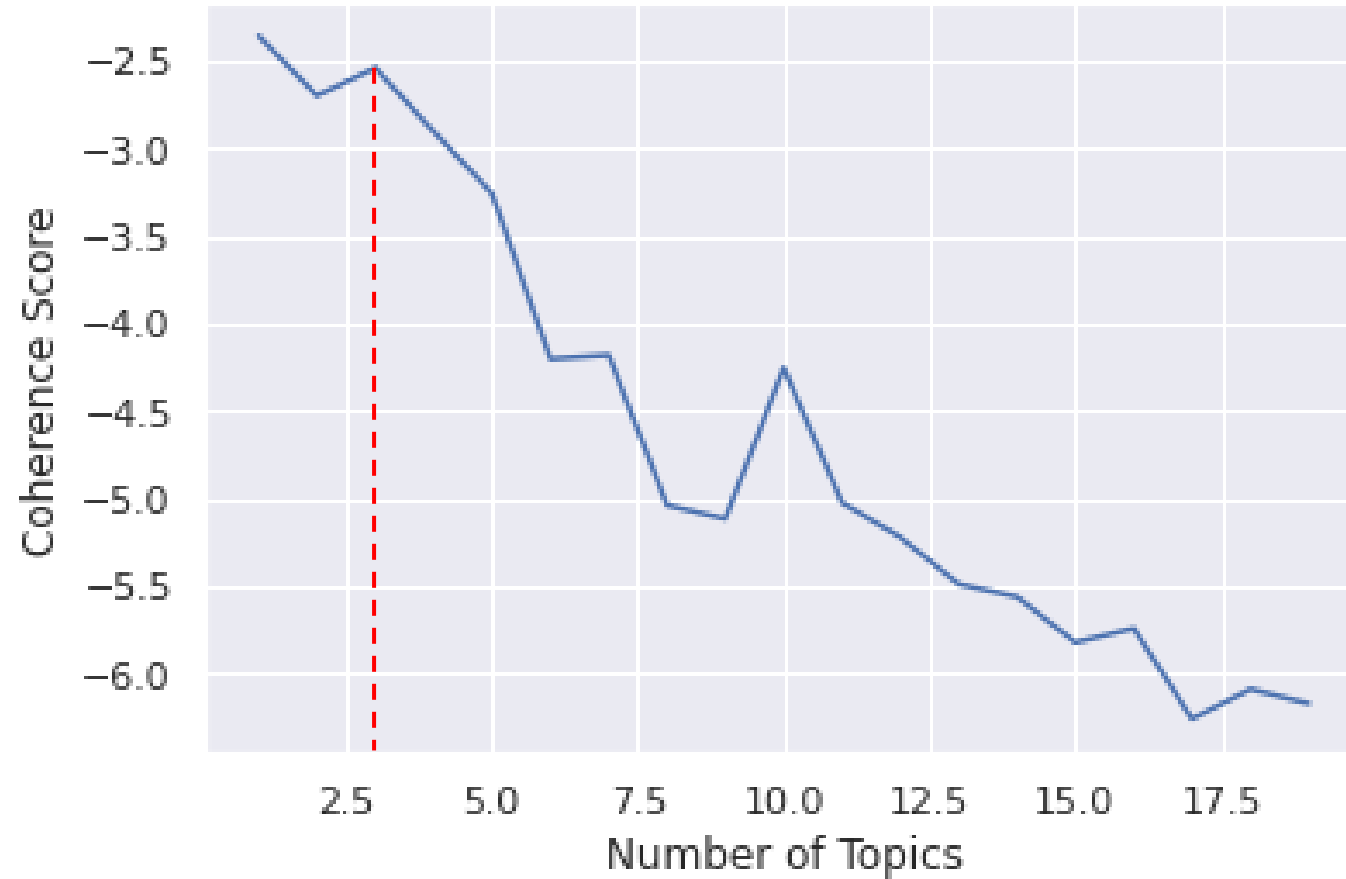


LDA Approach

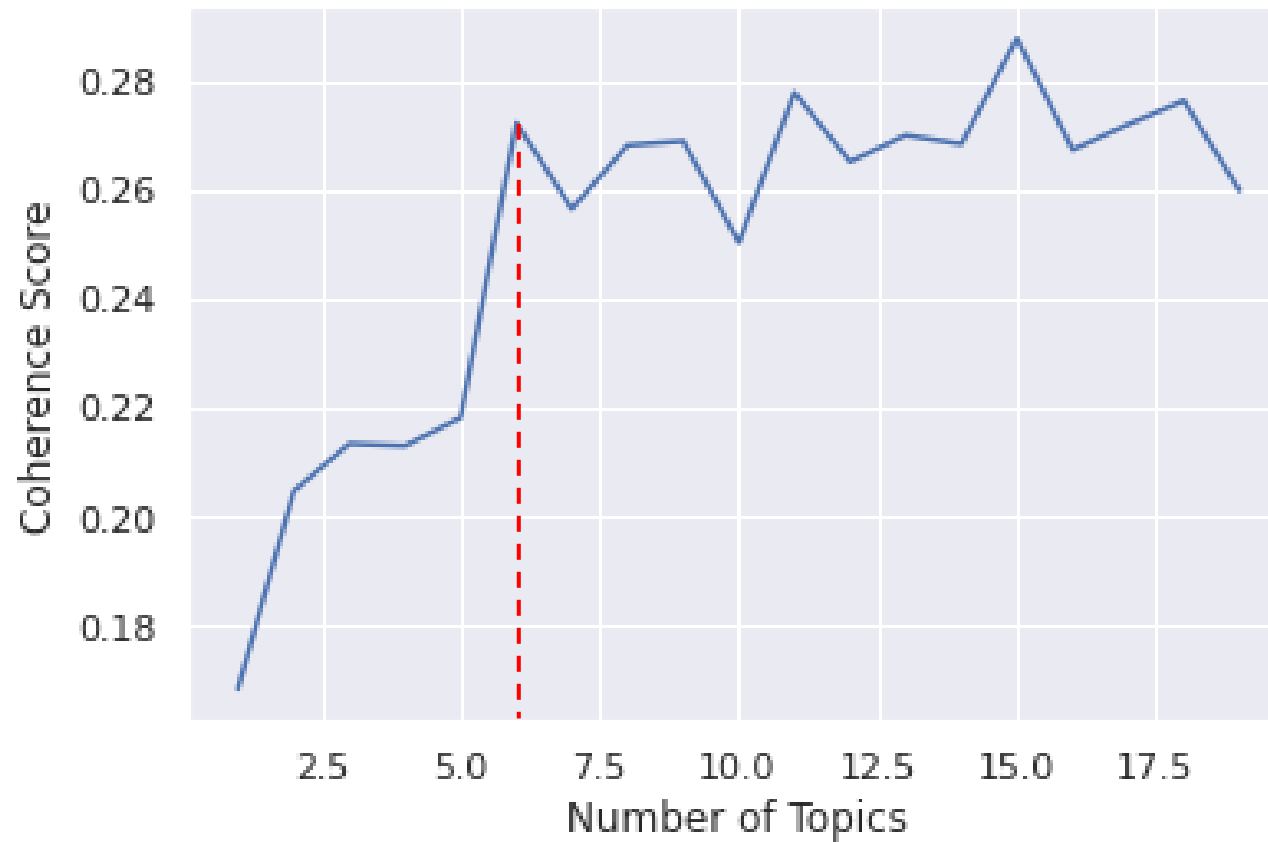


NOTE: The figure was developed based on a diagram published in the article titled, "Topic Modelling in Python with spaCy and Gensim", in The Towards Data Science website. URL: <https://towardsdatascience.com/topic-modelling-in-python-with-spacy-and-gensim-dc8f7748bdbf>

LDA Results: Coherence Score (C_umass)

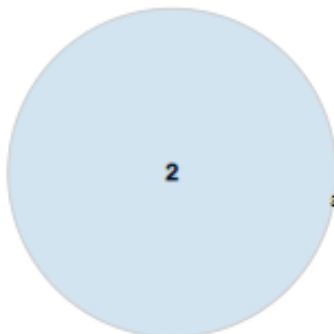
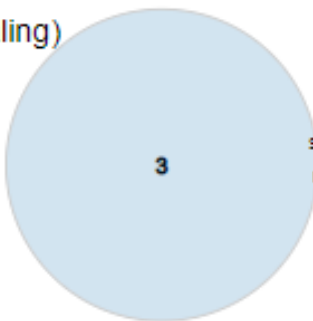
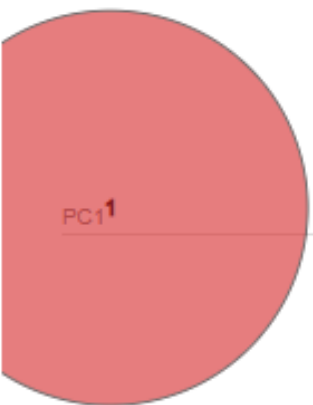


LDA Results: Coherence Score (C_v)



Setting number of topics = 3

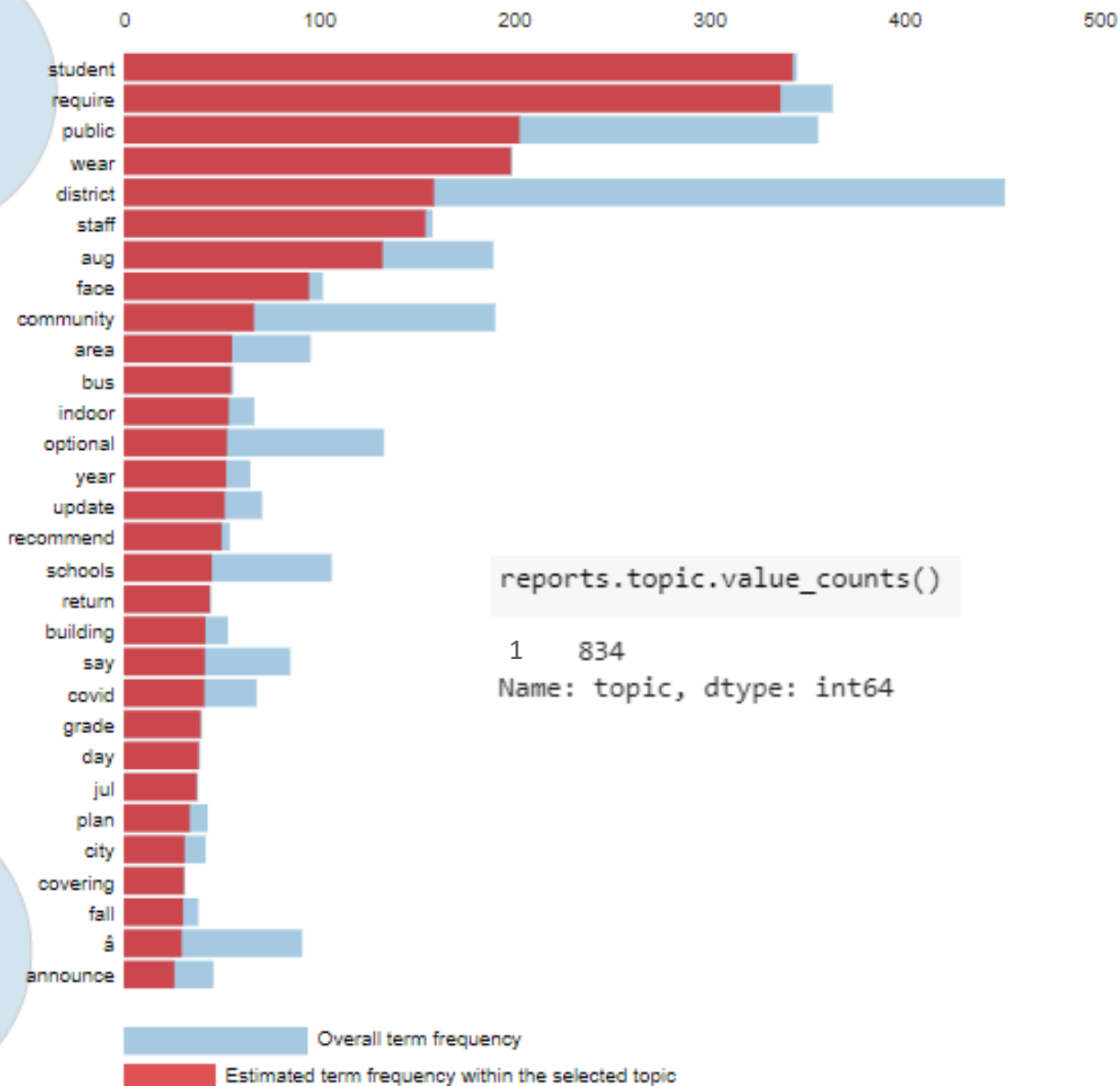
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (43.2% of tokens)



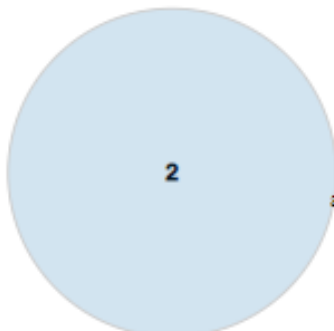
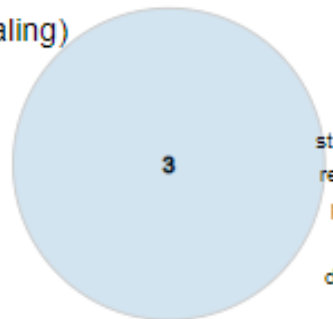
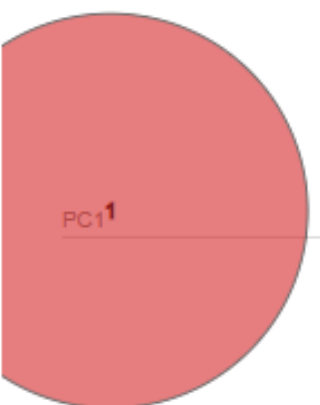
```
reports.topic.value_counts()
```

```
1    834
Name: topic, dtype: int64
```

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

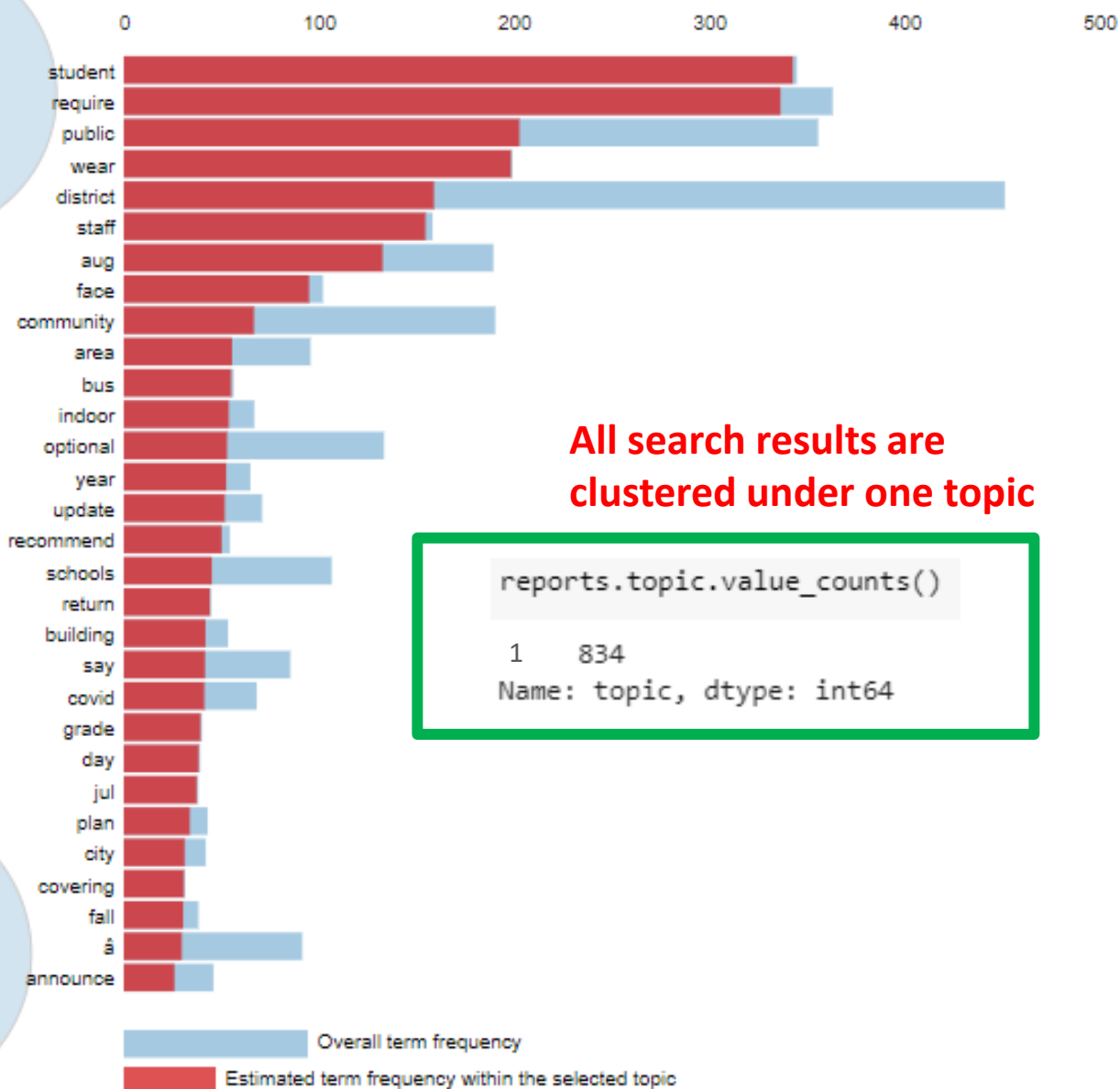
Setting number of topics = 3



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (43.2% of tokens)



All search results are clustered under one topic

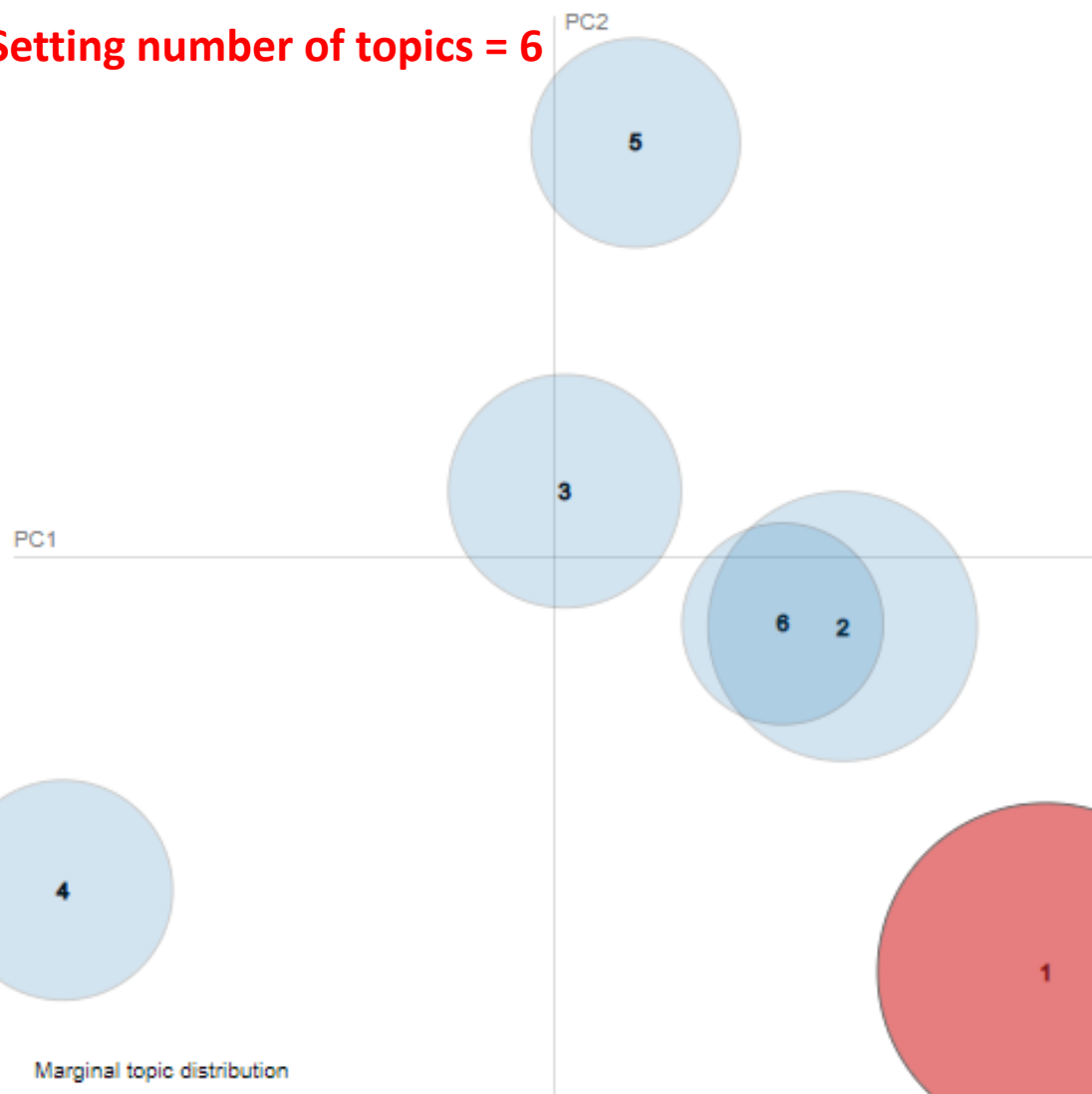
```
reports.topic.value_counts()

1    834
Name: topic, dtype: int64
```

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

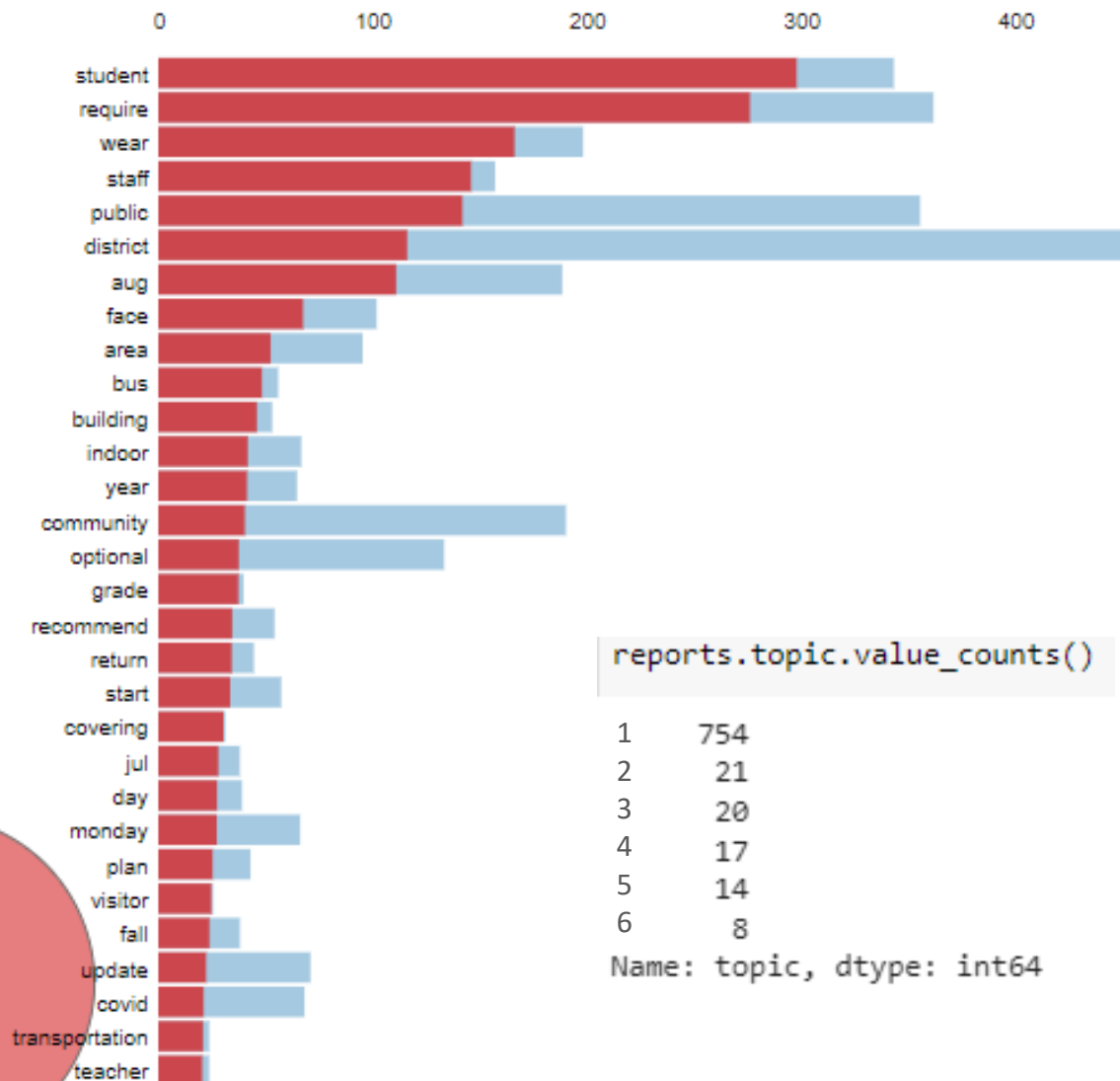
Setting number of topics = 6



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (30.5% of tokens)



```
reports.topic.value_counts()
```

1	754
2	21
3	20
4	17
5	14
6	8

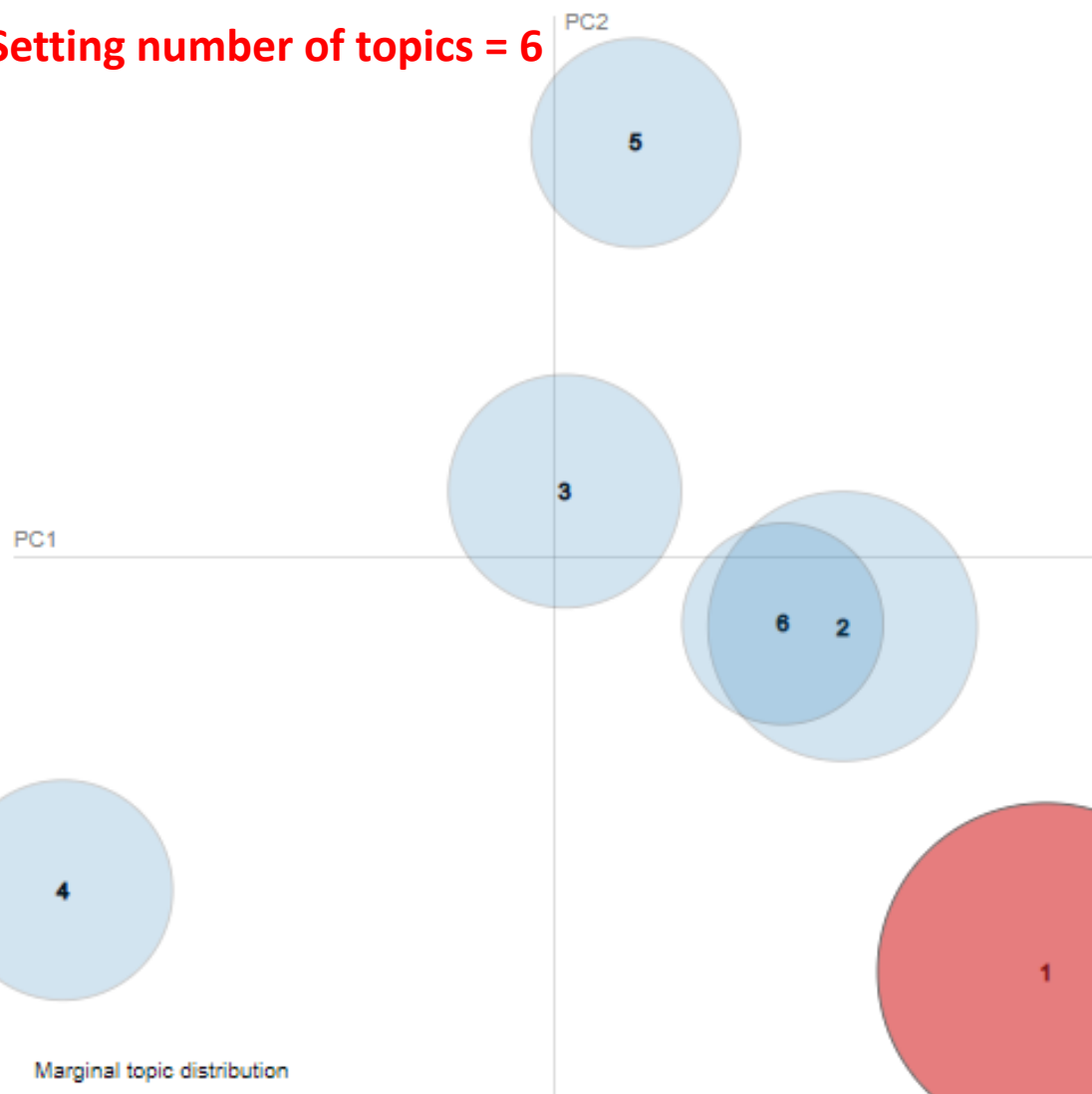
Name: topic, dtype: int64

Overall term frequency (blue bar)
 Estimated term frequency within the selected topic (red bar)

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

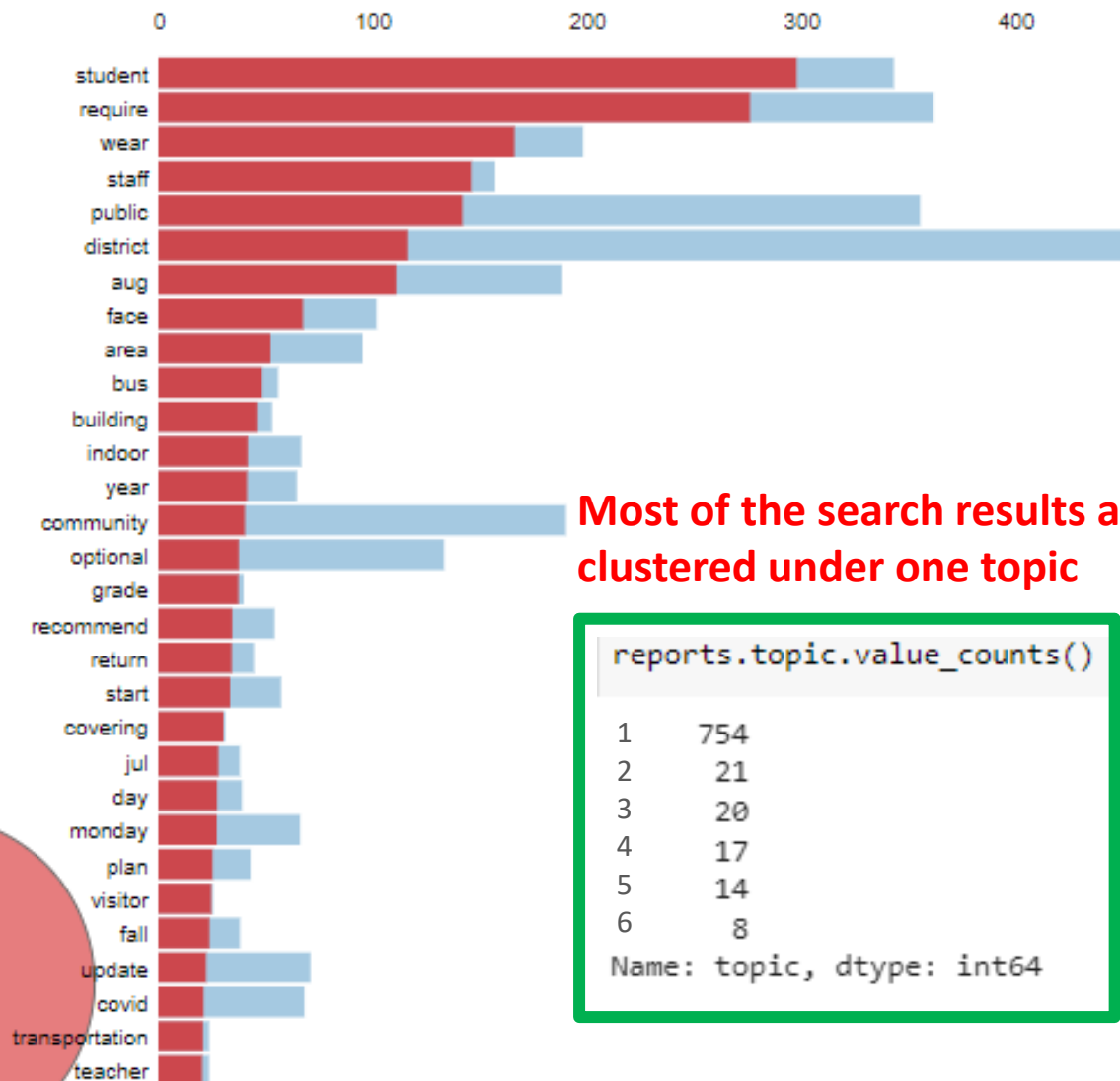
Setting number of topics = 6



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 1 (30.5% of tokens)



Most of the search results are clustered under one topic

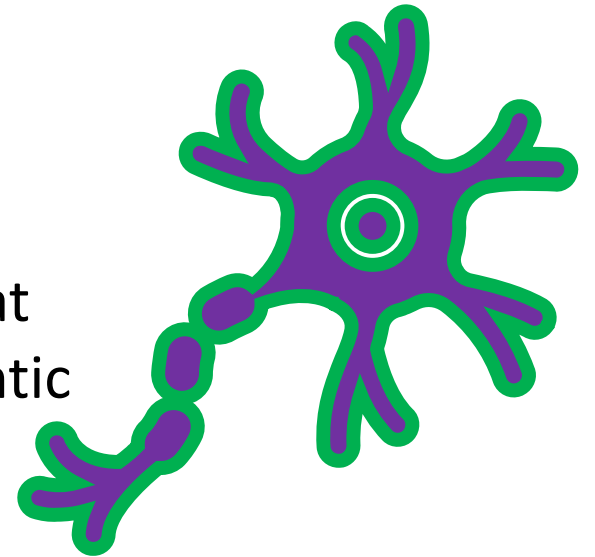
```
reports.topic.value_counts()
1    754
2     21
3     20
4     17
5     14
6      8
Name: topic, dtype: int64
```

Overall term frequency (blue bar)
Estimated term frequency within the selected topic (red bar)

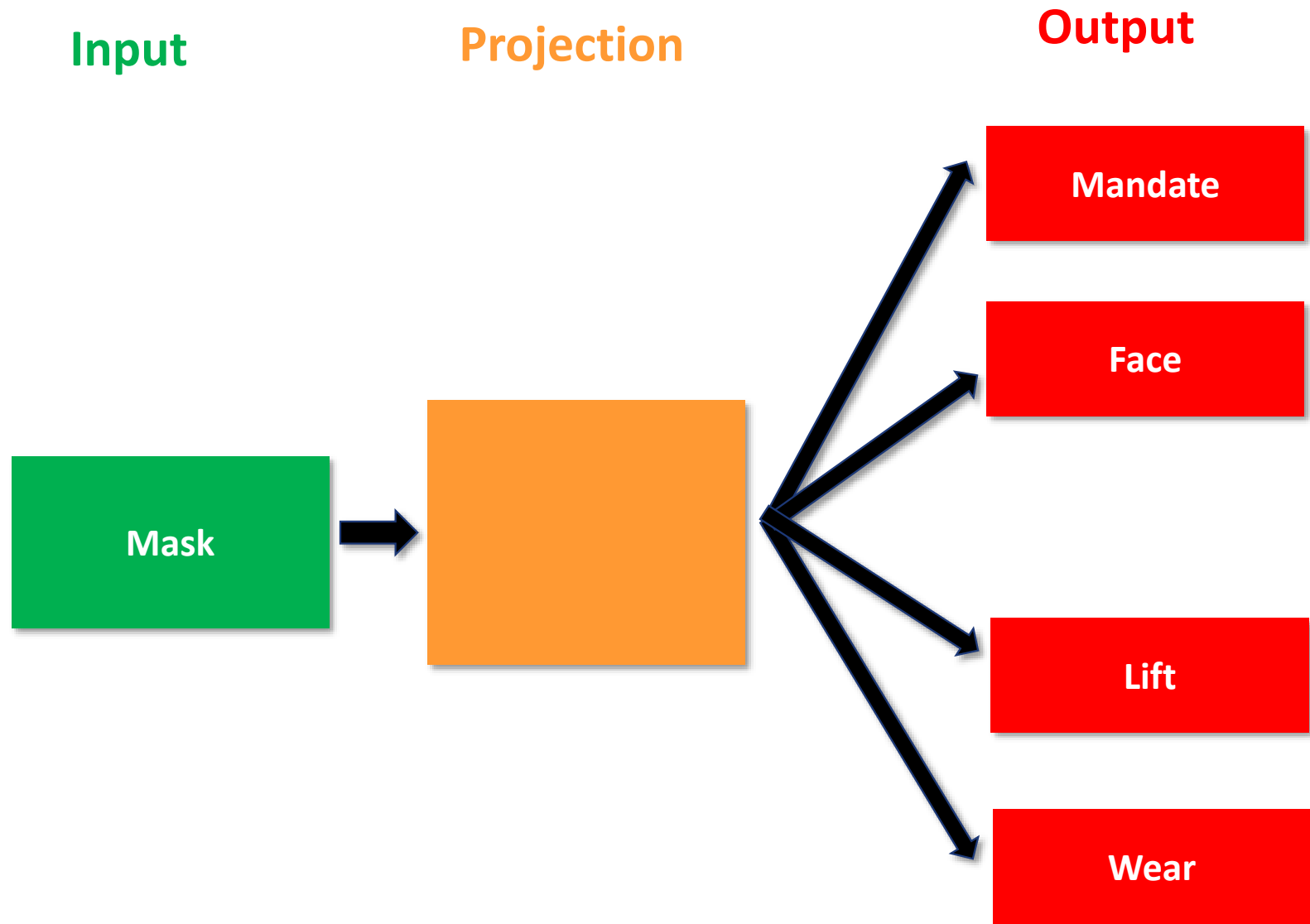
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Word2vec

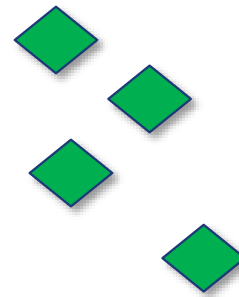
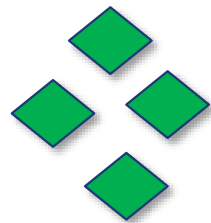
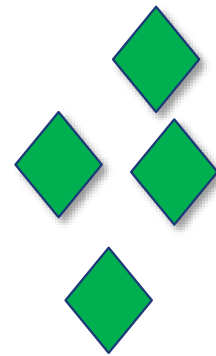
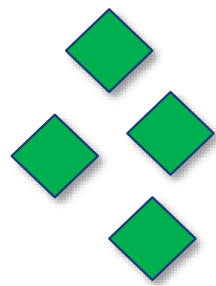
- The word2vec word embedding algorithm uses an artificial neural network to learn word associations from a given corpus. The model can detect similar words or suggest additional words in a phrase.
- Word2vec characterizes each unique word with a vectors of real numbers, where the vectors are selected so that they capture the syntactic and semantic qualities of words.
- Semantic similarity between words represented by those vectors can be detected using mathematical functions such as cosine similarities.



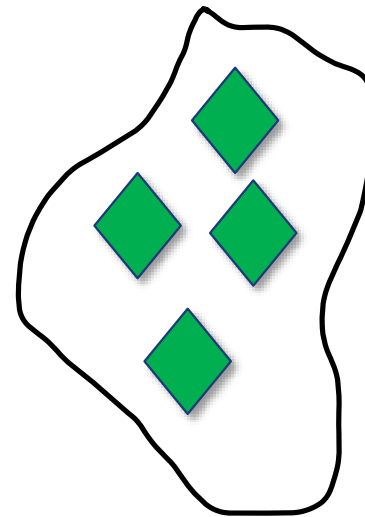
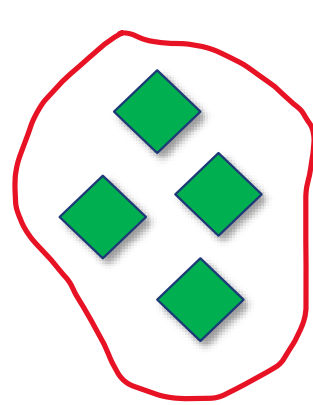
Word2vec Approach



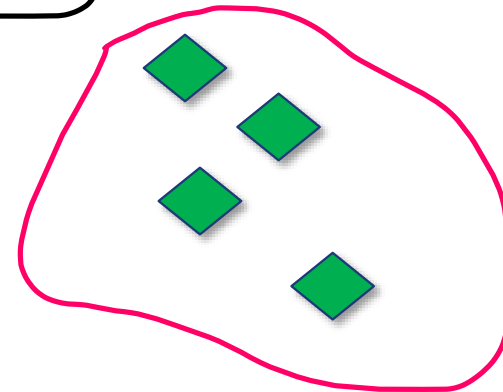
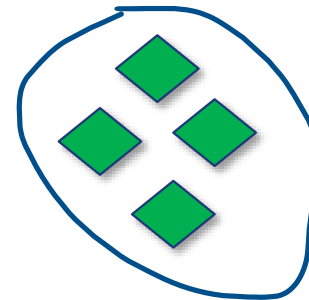
Word2vec + Clustering (K- means)



Word2vec + Clustering (K- means)



$$\underline{\underline{K = 4}}$$



Word2vec + Clustering Results: 4 Clusters

```
For n_clusters = 4  
Silhouette coefficient: 0.52  
Inertia:6.648079171763502  
Silhouette values:  
Cluster 2: Size:261 | Avg:0.55 | Min:0.05 | Max: 0.76  
Cluster 0: Size:109 | Avg:0.53 | Min:0.06 | Max: 0.68  
Cluster 3: Size:314 | Avg:0.53 | Min:-0.00 | Max: 0.74  
Cluster 1: Size:148 | Avg:0.47 | Min:0.01 | Max: 0.65
```

Word2vec + Clustering Results: 4 Clusters

```
For n_clusters = 4  
Silhouette coefficient: 0.52  
Inertia:6.648079171763502  
Silhouette values:
```

Cluster 2: Size:261		Avg:0.55		Min:0.05		Max: 0.76
Cluster 0: Size:109		Avg:0.53		Min:0.06		Max: 0.68
Cluster 3: Size:314		Avg:0.53		Min:-0.00		Max: 0.74
Cluster 1: Size:148		Avg:0.47		Min:0.01		Max: 0.65

**Search results are clustered
across the four topics**

Word2vec + Clustering: Most Representative Results

covid-19 dashboard -
kalkaska public schools
sep 14, 2021 " 315 s.
coral street, kalkaska, mi
49646 ... wear a mask
(covering mouth and
nose) when around
people with whom you
do not live.

masking update - mesick
consolidated schools the cdc
will be updating its guidance
for schools in the coming
weeks. this updated guidance
will inform schools for the
planning of the 2021-2022
school year. fully vaccinated
people will no longer need to
wear masks indoors or
outdoors

Cluster 0

danville implements mask
mandate, business owners
oppose dec 3, 2021 " "
danville is the latest
vermont town to put a
mask mandate in place. the
rule went effect last night
after a unanimous vote by
the select " ...

Word2vec + Clustering: Most Representative Results

covid-19 dashboard -
kalkaska public schools
sep 14, 2021 " 315 s.
coral street, kalkaska, mi
49646 ... **wear a mask
(covering mouth and
nose) when around
people** with whom you
do not live.

masking update - mesick
consolidated schools the cdc
will be updating its guidance
for schools in the coming
weeks. this updated guidance
will inform schools for the
planning of the 2021-2022
school year. **fully vaccinated
people will no longer need to
wear masks indoors or
outdoors**

Cluster 0: **Mask Wearing or Mask Mandate**

**danville implements mask
mandate**, business owners
oppose dec 3, 2021 " "
**danville is the latest
vermont town to put a
mask mandate** in place. the
rule went effect last night
after a unanimous vote by
the select " ...

Word2vec + Clustering: Most Representative Results

masks will be optional in
brandon valley schools after
parents ... aug 16, 2021 "â€"
25, after the school board
voted unanimously to keep
masks ... or opposition to, a
mask mandate for the school
district east of sioux fallsÂ ...

masks optional at newaygo
public schools this year jul
28, 2021 "â€" masks will be
optional for both staff and
students, the school's
superintendent jeff wright
said in a letter sent to
newaygo public schoolsÂ ...

Cluster 1

romulus community schools -
schoolandcollegelistings mar
14, 2022 "â€" updated mask
mandate dear romulus
community school district
students, families, students
and staff: as you may be
aware, the wayne county has
lifeted its mask mandate

Word2vec + Clustering: Most Representative Results

**masks will be optional in
brandon valley schools** after
parents ... aug 16, 2021 â€”
25, after the school board
voted unanimously to keep
masks ... or opposition to, a
mask mandate for the school
district east of sioux fallsÂ ...

**masks optional at newaygo
public schools** this year jul
28, 2021 â€” masks will be
optional for both staff and
students, the school's
superintendent jeff wright
said in a letter sent to
newaygo public schoolsÂ ...

Cluster 1: **Masks Optional or Mask Mandate Lifted**

romulus community schools -
schoolandcollegelistings mar
14, 2022 â€” updated mask
mandate dear romulus
community school district
students, families, students
and staff: as you may be
aware, the **wayne county has
lifted its mask mandate**

Word2vec + Clustering: Most Representative Results

all but battle creek public schools will make masks optional this fall aug 11, 2021 " but battle creek public schools (bcps) monday night released its preparedness and response plan that requires masking for all individuals indoors at all schools in the district starting when doors open Aug. 25

state, feds suggest masks in classrooms. michigan schools ... ann arbor public schools hasn't made a call on masks, according to a district spokesperson. kalamazoo public schools is planning to require all students and ...

Cluster 2

covid guidelines, fall 2021 - spring lake public schools aug 13, 2021 " michigan department of health and human services (mdhhs), and the michigan high school ... mask guidelines effective august 17, 2021: Masking is required in school buildings, including classrooms, hallways, offices, and common areas with the exception of time when students are eating lunch

Word2vec + Clustering: Most Representative Results

all but **battle creek public schools** will make masks optional this fall aug 11, 2021 " but battle creek public schools (bcps) monday night released its preparedness and response plan that **requires masking for all individuals indoors at all schools** in the district starting when doors open Aug. 25

state, feds suggest **masks in classrooms**. michigan schools ... ann arbor public schools hasn't made a call on masks, according to a district spokesperson. **kalamazoo public schools is planning to require all students and** ...

Cluster 2: **Masks Required**

covid guidelines, fall 2021 - **spring lake public schools** aug 13, 2021 " michigan department of health and human services (mdhhs), and the michigan high school ... mask guidelines effective august 17, 2021: **Masking is required in school buildings, including classrooms**, hallways, offices, and common areas with the exception of time when students are eating lunch

Word2vec + Clustering: Most Representative Results

are masks required at my kid's school? here's what districts in ... aug 18, 2021 " masks are recommended, but not required for students above sixth grade at kenowa hills public schools. the district says their current plan may

schools in midland county see uptick in cases among staff ... sep 28, 2021 " meridian public schools does not have a mask mandate, since the start of the school year the district has had 65 cases.

Cluster 3

livonia public schools fall 2021 mask & mitigation ... aug 25, 2021 " yes. per federal mandate, masks are required by any student, adult or volunteer while riding on public transportation, including school buses....

Word2vec + Clustering: Most Representative Results

are masks required at my kid's school? here's what districts in ... aug 18, 2021 " **masks are recommended, but not required** for students above sixth grade at kenowa hills public schools. the district says their current plan may

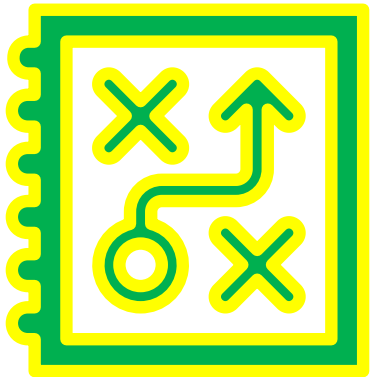
schools in midland county see uptick in cases among staff ... sep 28, 2021 " **meridian public schools does not have a mask mandate**, since the start of the school year the district has had 65 cases.

Cluster 3: **Masks Recommended or Not Required**

livonia public schools fall 2021 mask & mitigation ... aug 25, 2021 " yes. per federal mandate, **masks are required by any student, adult or volunteer while riding on public transportation, including school buses...**

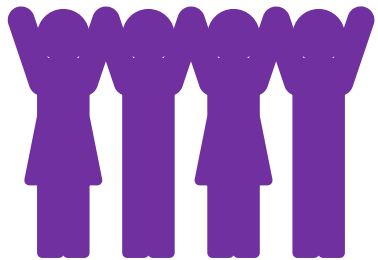
Overall, word embedding using word2vec and clustering using k-means provided very robust results

Further Applications



- This methodological approach could be used to collect, validate, and classify or cluster data on a wide range of topics related to education policy.
- Some topics Optimal is currently working on include:
 - Incidence of school shooting hoxes; and
 - Availability of mental health services in public schools.
- Overall, this approach can provide researchers a pathway to convert “non-data” data into rich tabular data that can be analyzed to better inform education policy decisions.

Thank you!



- **Contact us**

- Email:
essin@optimalsolutionsgroup.com
- Telephone: +1.301.306.1170



Visit our website!



