

# Web Scraping to Collect U.S. School District Level COVID-19 Mask Mandate Information for Contextualizing Results from the National Assessment of Educational Progress

2022 FCSM Research & Policy  
Conference

October 27, 2022 | Washington, DC

**Brian Cramer, NCES**

**Sadaf Asrar, *Optimal  
Solutions Group, LLC***



# National Center for Education Statistics

- The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education in the United States (U.S.) and other nations.
- NCES is located within the Department of Education and the Institute of Education Sciences.
- NCES fulfills a Congressional mandate to collect, analyze, and report complete statistics on the condition of American education, conduct and publish reports, and review and report on education activities internationally.



# Optimal Solutions Group, LLC



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



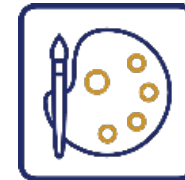
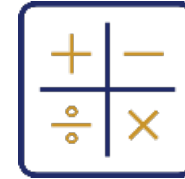
International Experience in  
40+ countries



Real-Time Framework for  
Public Policy Decision  
Making

# National Assessment of Educational Progress (NAEP)

- Assessment administered by the National Center for Education Statistics.
- Common measure of student achievement across the country in a variety of subject areas since 1969.
- Gold standard of large-scale assessments.
- Assesses students in a variety of subjects.
- Results are reported for major reporting groups and at the nation, state, and district (27) level.



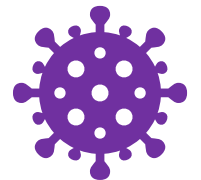
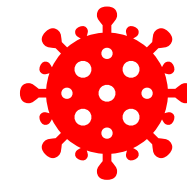
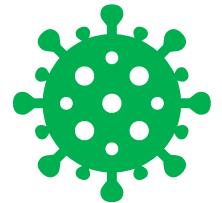
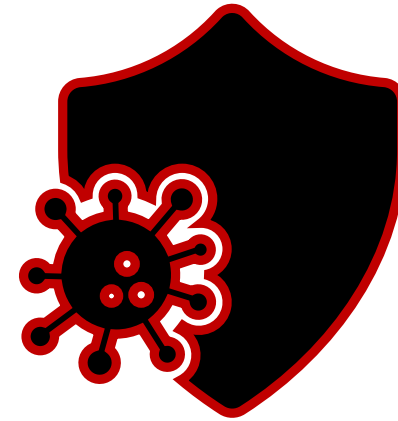
# Contextual Data for Understanding NAEP Student Achievement

- Contextual data utilized comes primarily from NAEP questionnaires (student, teacher, and school).
- Increasingly using NCES data sets, data sets from other federal agencies, and data from state education agencies and other sources.
- To collect data from states and other sources, NCES has been employing new methods.
- Two recent NASEM panels recommended that NCES utilize new data sources and data collection methods.



# 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.



- **During the 2021-2022 academic year:**
  - Several states imposed universal mask mandates.
  - Others imposed mask mandate bans.
  - Many left it up to the school districts to decide.



7

## 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.



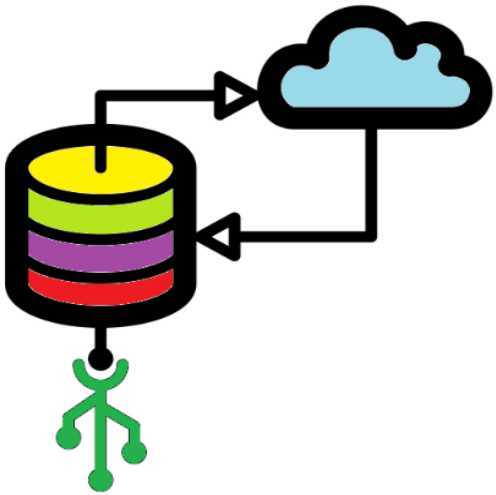


## School District Level Mask Mandates

- 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 a supervised Natural language Processing (NLP) machine learning algorithm to classify the search results.
- 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

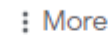
I'm Feeling Lucky

BeautifulSoup

# School District Level Mask Mandates



bexley city school district ohio mask mandate



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...

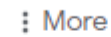
## Bexley City Schools drops mask mandate | NBC4 WCMH-TV

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



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](#)

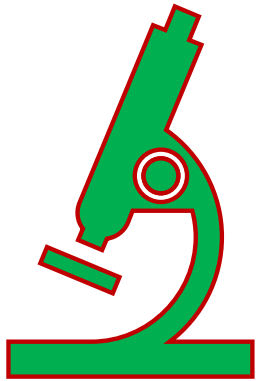
Mar 1, 2022 — In a unanimous vote during a **school board** meeting Tuesday, **masks** will now be optional starting Friday.

# School District Level Mask Mandates

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

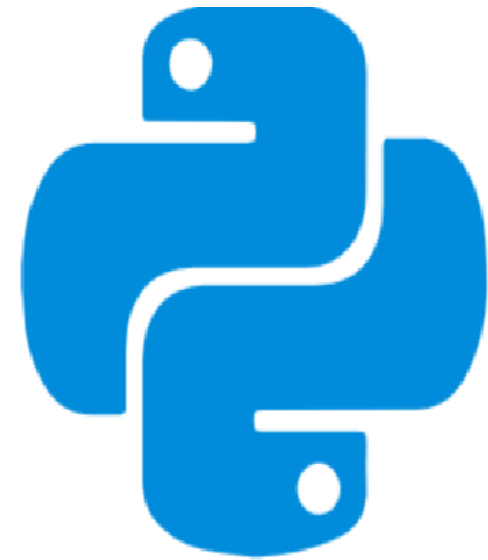
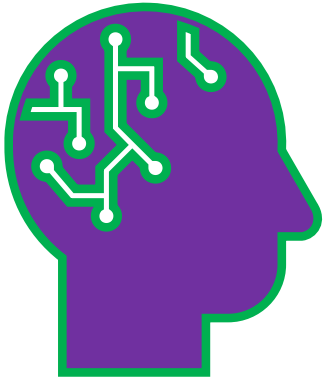


## Filtering Out Irrelevant Results

- 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.



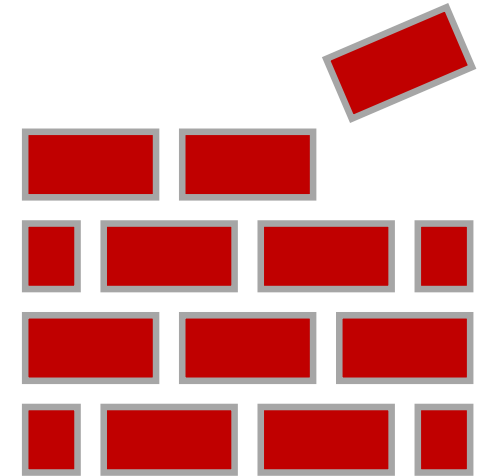
# Text Classification





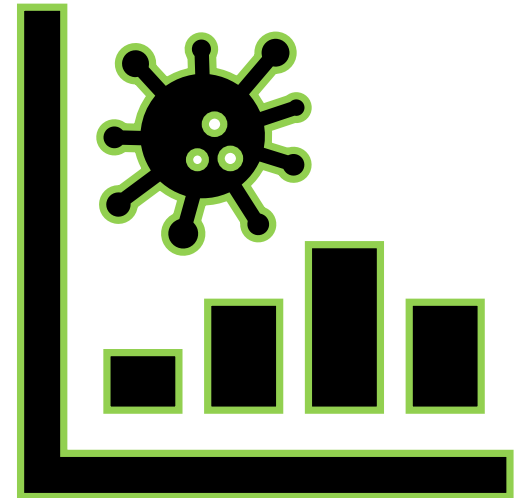
# Feature Engineering

- 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.

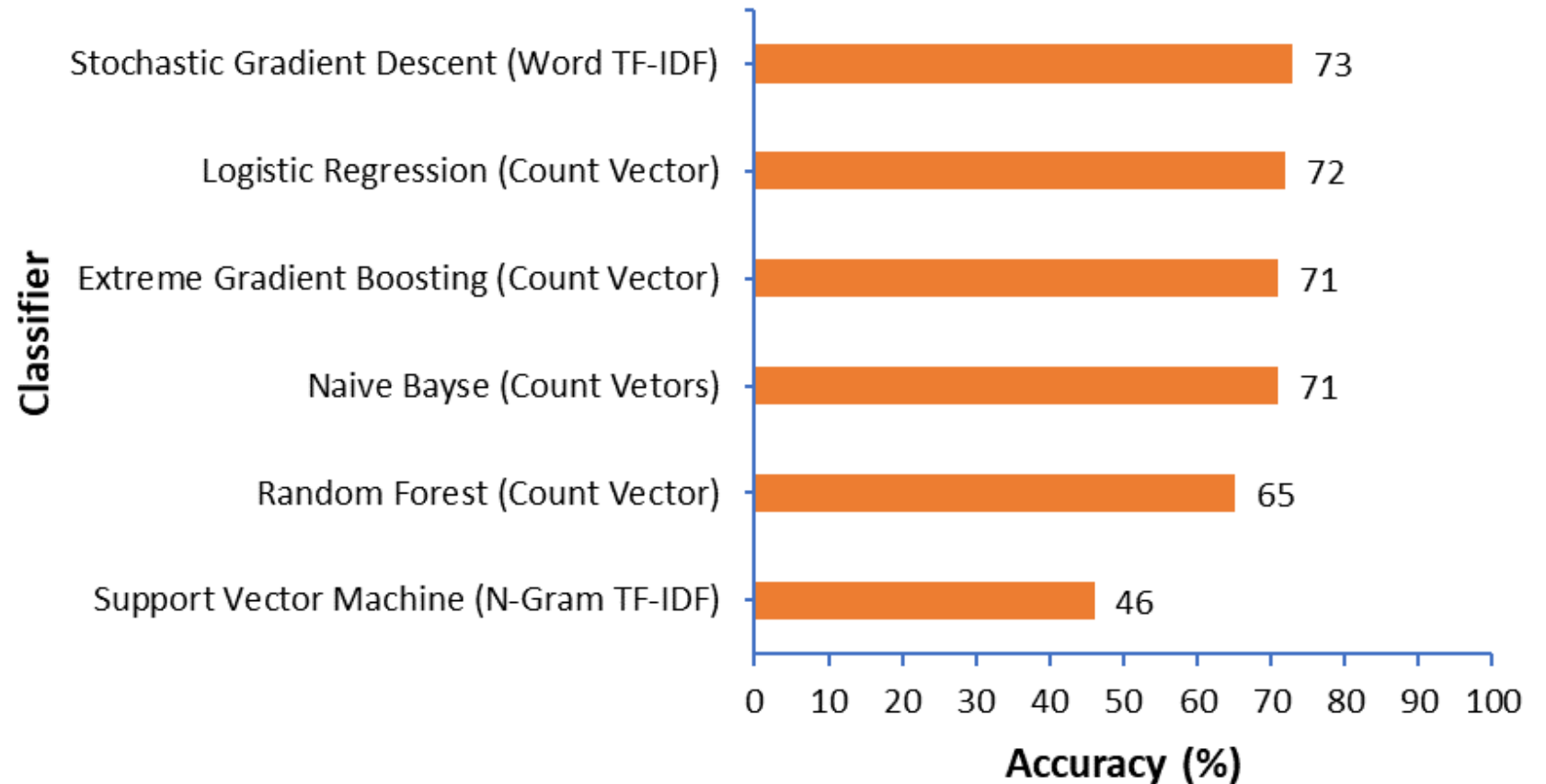


# Natural Language Processing Model

- 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.

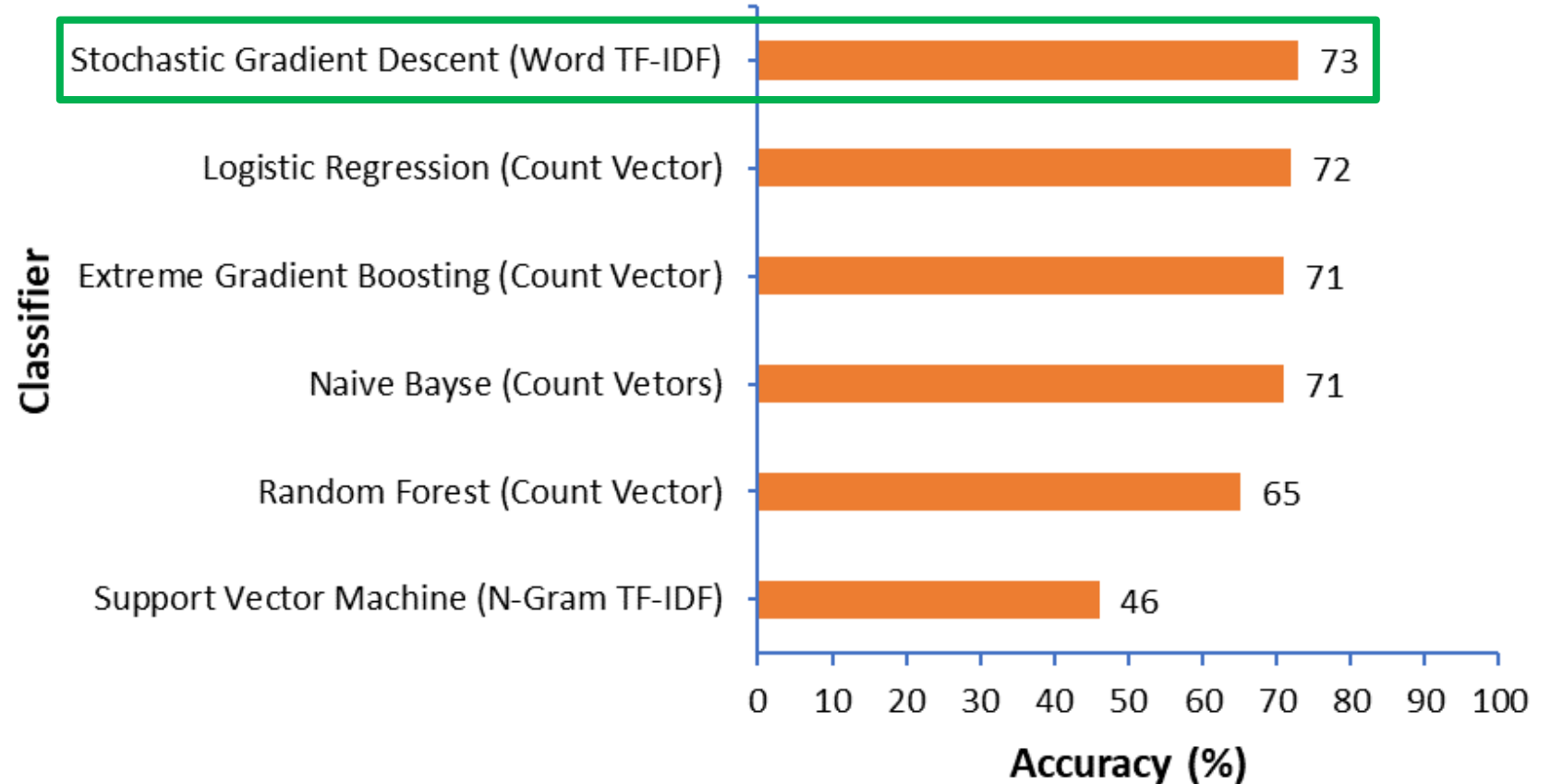


# Exploratory Model Performance



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

# Exploratory Model Performance



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

# Preliminary Results of Full Model

**Confusion Matrix**

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

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

# Preliminary Results of Full Model

**Confusion Matrix**

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

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

# Preliminary Results of Full Model

Confusion Matrix

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

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

# Preliminary Results of Full Model

Confusion Matrix

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

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



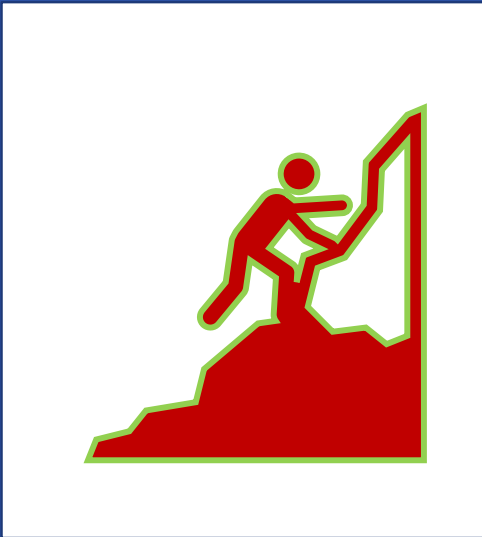
## Challenges for Classification



**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.*

## Steps to Improve the Approach



- Increase the size of the training data by:
  - Manually reviewing and labelling the search results for a few additional states.
- Balance the training dataset.
- Train additional machine learning and deep learning models.
- Explore unsupervised or self-supervised approaches such as:
  - Word embeddings;
  - Latent Dirichlet Allocation (LDA); and
  - Bidirectional Encode Representations from Transformers (BERT).
- Add human reviewers to bridge the accuracy gap for analysis on specific states and time periods.

# Thank you!



- **Contact us**

- Email:  
[essin@optimalsolutionsgroup.com](mailto:essin@optimalsolutionsgroup.com)
- Telephone: +1.301.306.1170

Visit our website!



