

# Using CATA and Machine Learning to Operationalize Old Constructs in New Ways: An Illustration Using U.S. Governors' COVID-19 Press Briefings

Organizational Research Methods  
1–28

© The Author(s) 2022

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/10944281221098607

[journals.sagepub.com/home/orm](https://journals.sagepub.com/home/orm)



Jason D. Marshall<sup>1</sup> , Francis J. Yammarino<sup>2</sup>,  
Srikanth Parameswaran<sup>2</sup>, and Minyoung Cheong<sup>3</sup>

## Abstract

Increased computing power and greater access to online data have led to rapid growth in the use of computer-aided text analysis (CATA) and machine learning methods. Using “big data”, researchers have not only advanced new streams of research, but also new research methodologies. Noting this trend and simultaneously recognizing the value of traditional research methods, we lay out a methodology that bridges the gap between old and new approaches to operationalize old constructs in new ways. With a combination of web scraping, CATA, and supervised machine learning, using labeled ground truth data (i.e., data with known inputs and outputs), we train a model to predict CIP (Charismatic-Ideological-Pragmatic) leadership styles from running text. To illustrate this method, we apply the model to classify U.S. state governors' COVID-19 press briefings according to their CIP leadership style. In addition, we demonstrate content and convergent validity of the method.

## Keywords

text analysis, CATA, web scraping, machine learning, leadership, COVID-19

---

<sup>1</sup>Heider College of Business, Creighton University, USA

<sup>2</sup>School of Management, Binghamton University, State University of New York, USA

<sup>3</sup>School of Graduate Professional Studies at Great Valley, Pennsylvania State University, USA

## Corresponding Author:

Jason D. Marshall, Heider College of Business, Creighton University, 602 N 20th Street, Omaha, NE 68102, USA.

Email: [jmarshall6112@gmail.com](mailto:jmarshall6112@gmail.com)

As the internet is crawling with data (pun intended), researchers have begun to leverage this “big data” source (Braun, Kuljanin, & DeShon, 2018) to advance theory in a variety of research contexts. Advanced computing capabilities and modern analytic techniques have made this possible. For example, with web scraping techniques, researchers can access data directly from a web page to be analyzed on their personal computer. In addition to numerical data, there is a plethora of textual data available on the internet. Accessing, extracting, and storing this data is as easy as executing a few lines of code in your favorite programming language (e.g., *Python* or *R*). This ease of access has corresponded with an increase in text analysis studies in recent years (Banks, Woznyj, Wesslen, & Ross, 2018). Specifically, computer-aided text analysis (CATA; McKenny, Aguinis, Short, & Anglin, 2018; McKenny, Short, & Payne, 2012; Short, Broberg, Coglisier, & Brigham, 2010) has grown more popular.

In brief, CATA provides rich insights into individual cognitions, values, and identities in ways that cannot be replicated via traditional research methods, such as self-report surveys (Pollach, 2012). CATA research typically uses deductive dictionary-based approaches to text analysis, which differs from open-language methods such as text mining (Kobayashi, Mol, Berkers, Kismihók, & Den Hartog, 2018), natural language processing (NLP; Pandey & Pandey, 2019), and neural network models (DeTienne, DeTienne, & Joshi, 2003; Minbashian, Bright, & Bird, 2010) that are also gaining traction. Whereas these open-language approaches to text analysis leverage machine learning algorithms and co-occurrence matrices to establish relationships among words in a bottom-up fashion, CATA predominantly takes a top-down approach, calculating word frequency within established categories or dictionaries. CATA dictionaries are comprised of words that are linked together in meaning, often theoretically developed, and psychometrically validated by subject matter experts (Short et al., 2010). Hence, with CATA’s psychometrically validated dictionaries, there is an untapped opportunity to utilize machine learning techniques, in conjunction with CATA, to gain some methodological advantages over open-language techniques.

There are two basic forms of machine learning: supervised and unsupervised. In unsupervised machine learning, the machine is not aware of any preexisting categories, dictionaries, or classifications of words, but merely uses a form of cluster analysis, based on co-occurrence matrices, to establish which groups of words cluster together in a corpus (i.e., a body of text). Text mining is an example of unsupervised machine learning applied to text analysis. Supervised machine learning, on the other hand, builds predictive models by learning from training examples where feature data is labeled according to its ground truth output (Zhou, 2018). Ground truth, in this case, refers to data that were human-coded or labeled in a prior research study, such as the CIP (Charismatic – Ideological – Pragmatic) labels assigned to U.S. Presidents in Yammarino and colleagues’ (2013) historiometric study. In other words, the researcher provides both input (independent) and output (dependent) variables, which allows the machine to learn how to best fit a model to match the two (Janasik, Honkela, & Bruun, 2009). We assert that supervised machine learning provides an avenue to leverage and extend existing ground truth output data to operationalize constructs in new ways.

With all the new advanced data analysis tools and techniques, and the relatively easy access to big data, one might be compelled to forget all about traditional research methods and move on to the shiny new toys. However, we offer an alternative perspective. Rather than take an “out with the old and in with the new” approach, we believe a better method is to leverage existing ground truth datasets (i.e., the old) and apply web scraping, CATA, and supervised machine learning techniques (i.e., the new) to advance new streams of research. This combination of techniques is an important methodological contribution to both micro- and macro-oriented research because it leverages and extends existing methodologies to allow for construct operationalization via text analysis. For example, the application of these techniques, in concert, would be particularly useful in upper echelons research (Hambrick & Mason, 1984) where traditional research methodologies (e.g., self-

report surveys, interviews) are more difficult to employ (Hambrick, 2007). As such, in this context, our operationalization could serve as a useful tool to test and/or develop theory in upper echelons research.

Of course, not all public-facing individuals occupy upper echelon positions in organizations. Fortunately, this approach could just as easily be applied at the individual level of analysis for movie stars, professional athletes, and social media influencers. Additionally, as we highlight throughout the manuscript, publicly available data and web scraping are not prerequisites for the execution of the CATA and machine learning portions of the method. As such, the method could be easily applied in conjunction with more traditional forms of research at the individual or group level. In other words, our recommended methodology could be applied at multiple levels of analysis and in a variety of contexts.

Another strength of this method, beyond convenience, is that it allows for a more nuanced examination of psychometric data over time. Aguinis and Bakker (2021) defined four aspects of time (*duration, frequency, timing, and sequence*) that can and should be specified when discussing time as an aspect of a construct. By operationalizing constructs through running text (e.g., speeches, interviews, tweets), the method we demonstrate in this manuscript allows researchers to capture each of these aspects of time. In other words, as running text often spans across time and contexts, this methodological combination makes it possible for researchers to examine underlying theoretical assumptions (e.g., stability, causal sequence) that were previously untestable or impractical to test. To demonstrate the utility of our recommended method, the current study illustrates how to use a combination of web scraping, CATA, and supervised machine learning techniques to classify categorical outcomes in an upper echelons context.

## **CIP Style in U.S. Governors' COVID-19 Press Briefings**

### *Current Study*

The particular categorical outcome we have chosen to model is the CIP (Charismatic-Ideological-Pragmatic) outstanding leadership style approaches (Lovelace, Neely, Allen, & Hunter, 2019; Mumford, 2006) of U.S. state governors. With the current COVID-19 pandemic, the effects of leadership (or lack thereof) have been magnified. As Bligh and colleagues noted, "People become increasingly susceptible to the leader and his or her vision in the wake of a crisis" (Bligh, Kohles, & Meindl, 2004: 215). The pandemic is certainly a crisis and leaders (e.g., U.S. state governors) have been frequently communicating their visions, beliefs, and solutions via press briefings. Accordingly, in this study, we seek to categorize the U.S. state governors' press briefings according to their CIP leadership style in hopes of not only advancing a new methodology, but also providing data analysis that may be of interest/value to the general public. The latter aspiration is a direct response to a call from Aguinis and colleagues (Aguinis, Suarez-Gonzalez, Lannelongue, & Joo, 2021) to reach an audience outside of academia.

### *CIP Model*

Mumford (2006) developed the CIP model of outstanding leadership based on early work from Max Weber (1924). According to the model, there are three different leadership styles that are all equally likely to produce outstanding leadership, and one key distinction among the styles is the time-orientation of the leader (Mumford, 2006): Charismatic leaders (i.e., the *C* in CIP) are future-focused, ideological leaders (i.e., the *I* in CIP) are past-focused, and pragmatic leaders (i.e., the *P* in CIP) are present-focused. Not only does time-orientation play a key role in differentiating among CIP leadership styles, but for charisma specifically, it also distinguishes Mumford's (2006) conception of the

construct from the more mainstream definitions of charisma (e.g., Antonakis, Bastardo, Jacquart, & Shamir, 2016; House & Howell, 1992; Shamir, Arthur, & House, 1994). In addition to time orientation, several other factors are used to differentiate among CIP styles. For example, charismatic leaders use positive emotional imagery and are concerned with motivating followers. Regarding problem solving, charismatics are best during the middle problem-solving phase when their ability to generate possibilities is most valuable (Lovelace et al., 2019). Ideological leaders are values-driven and use negative emotion in their communication (Hunter, Cushenbery, Thoroughgood, Johnson, & Ligon, 2011). During times of crisis, ideologues perceive situations as the causal mechanism, focus on changing the system, and may rally their base constituency (i.e., like-minded individuals), rather than attempt to appeal to the masses (Griffith et al., 2018). With regard to their problem-solving value, ideologues tend to be best during the late problem-solving phase when their ability to implement solutions is most needed (Lovelace et al., 2019). Finally, pragmatic leaders communicate using rational appeals and are best during the early phase of problem-solving when data are being gathered and evaluated. During times of crisis, pragmatic leaders adopt an interactionist approach, focusing on how the situation affects people and their behavior (Griffith et al., 2018; Lovelace et al., 2019).

### *Limits of Historiometric Method and Opportunity for New Methods*

Prior CIP studies have primarily used historiometric methods (Ligon, Harris, & Hunter, 2012) to assess the CIP style in outstanding historical leaders such as Presidents (Yammarino, Mumford, Serban, & Shirreffs, 2013), world leaders (Serban et al., 2018), and college/NFL football coaches (Hunter et al., 2011). In other words, these researchers have used historical documents, such as biographies, to code whether an individual is C, I, or P. Recently, Lovelace and colleagues (2019) stated that research on the CIP model has been over-reliant on the historiometric approach and claimed this reliance may be limiting the potential applications of the model. Although historiometric methods have a few limitations for the study of CIP in organizations, the biggest limitation is that historiometric methods are labor-intensive to employ (Lovelace et al., 2019). Researchers spend dozens of hours learning how to code the documents. Then, they spend countless hours reading over and manually coding documents. Then, the documents are checked for interrater reliability. All of this takes a tremendous amount of time and resources.

Thus, to address this limitation of the historiometric approach, upon which the CIP model has heavily relied, we advance a more granular, flexible, and timely operationalization of the CIP model that combines web scraping, computer-aided text analysis (CATA), and supervised machine learning techniques. Web scraping opens the door to numerous textual datasets, systematically displayed on publicly available websites. CATA not only allows researchers to code large volumes of text in a fraction of the time, but it also allows us to study running text (e.g., speeches, interviews, meeting transcripts, tweets) in current leaders, without waiting for biographies to be written after the fact. Finally, supervised machine learning, when used in conjunction with a ground truth dataset, can be used to train a model to predict outcome variables (e.g., CIP style in our case). Collectively, these methodologies provide a new operationalization for CIP research.

### **New Methods in Combination**

To illustrate the operationalization of our method, we leverage web scraping, CATA, and supervised machine learning techniques to classify CIP style of U.S. state governors. To help guide the reader through the process and enable other researchers to execute this method for their research topics of interest, whether micro, meso, or macro in orientation, we provide the full step-by-step process, including recommended *R* and *Python* packages, in Table 1.

**Table 1.** Steps to Execute the Illustrated Method.

Step	Action	Recommended Package/Program
1	Identify ground truth labels of interest (e.g., CIP style of U.S. Presidents, Yammarino et al., 2013)	
2	Identify textual data source to generate features for individuals in the ground truth dataset (e.g., website)	
3	Scrape textual data to generate features for a training dataset	<i>BeautifulSoup (Python)</i> <i>rvest (R)</i>
4	Perform computer-aided text analysis (CATA) on the textual data to create numeric features	LIWC
5	Merge LIWC output with ground truth labels to build a dataset for supervised machine learning	<i>merge (R)</i>
6	Compare supervised machine learning algorithms <ul style="list-style-type: none"> <li>■ Certain algorithms are designed for categorical outcome variables and others for continuous outcome variables</li> <li>■ Simply change the “model” in the <i>caret</i> script to run various algorithms</li> <li>■ Compare model accuracy and kappa among algorithms</li> </ul>	<i>caret (R)</i>
7	Model feature selection <ul style="list-style-type: none"> <li>■ Let theory and statistics guide the feature selection</li> <li>■ Depending on the algorithm, there are different ways to identify the important features, within the <i>caret</i> package <ul style="list-style-type: none"> <li>• For MLR, we used the <i>varImp</i> function</li> <li>• For other algorithms (e.g., RandomForests), the <i>rfe</i> function in the <i>caret</i> package can be used</li> </ul> </li> </ul>	<i>caret::varImp (R)</i> <i>caret::rfe (R)</i>
8	Train and validate a supervised machine learning model <ul style="list-style-type: none"> <li>■ Various cross-validation methods are available in the <i>caret</i> package (e.g., <i>k-folds</i>, repeated <i>k-folds</i>, bootstrap, bootstrap .632, leave-one-out)</li> </ul>	<i>caret::trainControl (R)</i>
9	Scrape textual data for model application (e.g., CIP style in U.S. state governors)	<i>BeautifulSoup (Python)</i> <i>rvest (R)</i>
10	Perform CATA on the textual data you wish to classify	LIWC
11	Apply the model to predict outcomes in new data	<i>predict (R)</i>
12	Model validation (if possible) via SMEs	

Note. LIWC stands for Linguistic Inquiry and Word Count (Pennebaker et al., 2015).

### Sample and Selection

As we use multiple datasets from multiple sources, in Figure 1 we provide sample data configurations to help distinguish among the datasets and provide a visual overview of the supervised machine learning process.

**Ground Truth Data.** To build a model using supervised machine learning, we needed a dataset with labels assigned to the outcome variables. In supervised machine learning, this labeled data is often referred to as *ground truth data*. For the present study, our ground truth labels were derived from Yammarino and colleagues’ (2013) historiometric study of United States Presidents, in which each President was classified as either C, I, or P. We used Yammarino et al.’s (2013) classifications to build our CIP ground truth dataset for all U.S. Presidents prior to Donald Trump.

**Training and Validation Data.** Through web scraping techniques, we collected Presidential address transcripts from “The American Presidency Project”, hosted by the University of California (UC)

<b>Ground Truth Labels</b> Yammarino et al. (2013) Washington - Obama	<b>President</b>	<b>CIP Style</b>		
	Abraham Lincoln	C		
	Ronald Reagan	I		
	Barack Obama	P		
<b>Training - Validation</b> <i>N</i> = 844 Addresses Scraped from The American Presidency Project	<b>President</b>	<b>Addresses</b>	<b>LIWC</b>	<b>CIP Style</b>
	Abraham Lincoln	1 <sup>st</sup> Inaugural Address	Numeric Data	C
	Abraham Lincoln	2 <sup>nd</sup> Inaugural Address	Numeric Data	C
	Ronald Reagan	1 <sup>st</sup> Inaugural Address	Numeric Data	I
	Ronald Reagan	State of the Union	Numeric Data	I
	Barack Obama	1 <sup>st</sup> Inaugural Address	Numeric Data	P
	Barack Obama	Correspondents' Dinner	Numeric Data	P
<b>Application Data</b> <i>N</i> = 835 COVID-19 Press Briefings Scraped from Rev.com	<b>U.S. Governors</b>	<b>COVID Press Briefings</b>	<b>LIWC</b>	<b>CIP Style</b>
	Andrew Cuomo	Press Briefings (160)	Numeric Data	?
	Ron DeSantis	Press Briefings (62)	Numeric Data	?
	Gretchen Whitmer	Press Briefings (47)	Numeric Data	?

**Figure 1.** Data sources and examples.

Note. CIP stands for *charismatic, ideological, and pragmatic*. LIWC stands for Linguistic Inquiry and Word Count (Pennebaker et al., 2015). Ground truth data were collected from Yammarino et al., 2013. Training-testing data were collected from “The American Presidency Project,” hosted by UC Santa Barbara (844 Presidential addresses total). Application data were collected from Rev.com (835 COVID-19 press briefings total).

Santa Barbara at <https://www.presidency.ucsb.edu/index.php>, to use as the raw data for model training and validation. The addresses included inaugural addresses, state of the union addresses, commencement addresses, remarks at the White House correspondents’ dinner, holiday addresses, remarks to the United Nations, remarks to U.S. Congress, remarks to foreign governments, and remarks on foreign affairs. The addresses spanned from George Washington, in 1789, through Barack Obama, in 2016. Data were also collected from Donald Trump; however, they were not used in this dataset because he was not included in the Yammarino et al. (2013) CIP classifications. As each President was classified by a single CIP style, in our ground truth data (Yammarino et al., 2013), we applied that same style to each Presidential address for a given President. For example, Barack Obama was classified as *P*, therefore, each of his Presidential addresses was labeled as *P*. In total, we collected 844 transcripts of U.S. Presidential addresses, which we analyzed using Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, Boyd, & Francis, 2015) and merged with the labeled ground truth CIP labels from Yammarino et al. (2013). After accounting for language change over time, which we systematically assessed (see Appendix A), our final training and validation sample consisted of 377 Presidential addresses that occurred after 1980.

**Application Data.** After a model has been trained and validated, it can be applied to predict outcomes in new data. Accordingly, our model was applied to a dataset containing 835 COVID-19 press briefing transcripts, representing 42 United States governors. Through web scraping techniques, we collected the press briefing transcripts from <https://www.rev.com/blog/transcript-tag/coronavirus-update-transcripts>. The press briefings spanned from February 27, 2020 to April 23, 2021. Regarding political affiliation, 549 of the briefings were from Democratic governors (65.74%) and

286 (34.25%) were from Republican governors. Females delivered 88 of the briefings (10.54%) and males delivered the other 747 (89.46%).

## Web Scraping

Web scraping refers to the automated process of obtaining data directly from a webpage. While some companies develop specific APIs (Application Programming Interfaces) that allow researchers to quickly and easily access data (Braun et al., 2018), many times APIs are not available. For websites without an API, Kobayashi and colleagues recommend using web scraping techniques to acquire the data (Kobayashi et al., 2018). Prior to beginning the web scraping process, it is important to consult the terms of use for the website from which you wish to acquire the data. For example, the terms of use for the Rev website, where we scraped the governors' press briefings, stated restrictions such that a user may not, "take any action that imposes an unreasonable or disproportionately heavy load on the Platform or its infrastructure or that negatively affects the ability of others to access or use the Platform" (Rev, 2020). This does not explicitly forbid web scraping per se, but rather necessitates the use of responsible web scraping practices, which we will describe in detail below. In addition to reviewing the terms of use, in this particular case, we also received written consent stating, "You're welcome to use the content as you please" (Lawson, 2020).

For this study, we followed Braun and colleagues' (2018) general web scraping process: a) identify the website address(es), b) identify the data on the website to be extracted, c) write a script to extract the data, and d) execute the script to download the data onto a computer.

*Identify the Website Address(es).* After you have identified the home page of the website, you must identify the specific pages that contain the data you wish to extract. For example, the website for the governors' COVID-19 press briefings is <https://www.rev.com/blog/transcript-tag/coronavirus-update-transcripts>. Many times, the data are not all available on this single page, but rather several pages. For example, at the time of this writing, there are 108 pages of COVID-19 press briefings on the Rev.com site. Fortunately, for web scraping purposes, most websites establish a stable pattern for labeling these pages (e.g., <https://www.rev.com/blog/transcript-tag/coronavirus-update-transcripts/page/2>; <https://www.rev.com/blog/transcript-tag/coronavirus-update-transcripts/page/3>; <https://www.rev.com/blog/transcript-tag/coronavirus-update-transcripts/page/4>). These patterns enable you to write a programming script, using your preferred programming language (e.g., *R* or *Python*), to access the web pages for scraping.

*Identify the Data on the Website to be Extracted.* Most web pages consist of some combination of text, images, tables, and/or hyperlinks to other websites or pages. On the front-end, the contents are often displayed in some aesthetically pleasing manner. On the backend, however, it is merely a combination of HTML (HyperText Markup Language) and CSS (Cascading Style Sheets). A full review of HTML and CSS is beyond the scope of this manuscript (for a detailed overview, see Mitchell, 2018), thus, we simply want to note that most websites use patterns, in the HTML and CSS, that enable researchers to write a programming script to extract the data from the web pages. There is an easy way to access the HTML and CSS to identify these patterns in the web pages: While viewing a web page on a web browser (e.g., Google Chrome), you can highlight a portion of the page you wish to extract, right click, and select "inspect". This will give you access to the HTML and CSS for that particular section of the web page. There you will find the HTML tags and classes needed to write the programming script. For a more detailed explanation, we provide a step-by-step tutorial of this process in Appendix B.

**Write a Script to Extract the Data.** After you have identified the web page(s) you want to visit and identified patterns in the HTML/CSS, you can write a script to extract the data. We used the *BeautifulSoup* package (Richardson, 2020) in *Python*; however, there are options in *R* (R Core Team, 2022) as well (for a summary of *R* packages for web scraping, see Braun et al., 2018). Put simply, *BeautifulSoup* reads and parses HTML so that data can be extracted from a website. For reference, two examples of our web scraping *Python* scripts are provided in Appendix B (governors' COVID-19 press briefings) and Appendix C (Presidents' state of the union addresses).

Writing the script follows the same general steps as the overall scraping process. After loading your packages (e.g., *BeautifulSoup*, *time*, *requests*, *pandas*) you create a list of the web pages you wish to visit. If the website has multiple pages, like our Rev.com example, you may have to create an empty list and use a *for loop* to recreate the pattern of the website's pagination to populate your website list. If the pages you visit merely hold preview content with links to the actual data (like our Rev.com example), you must create another empty list, write a *for loop* to visit each summary page, retrieve the hyperlinks to the pages with the actual data you wish to extract, and store the hyperlinks in the new list. Next, you write a *for loop* to visit each hyperlink that contains your desired data, retrieve the data from the specified HTML tags/classes, and store the data.

This is the phase of the process where responsible web scraping is critical. As noted in the Rev.com terms of service, websites do not want you to overload their servers with rapid requests. Doing so may not only violate the terms of use, but it may also result in your IP (internet protocol) address being blocked from accessing the website. This would hamper your data collection process. To avoid overloading the server, it is important to incorporate time delays in your web scraping script. The *time* package in *Python* allows you to put your script to "sleep" for a specified number of seconds before resuming the scraping process. This allows you to avoid overloading the server and, more pragmatically, avoid getting blocked from the website. Of course, this was a cursory overview of the script writing process. For a detailed overview of the entire process, using *BeautifulSoup*, see Mitchell's (2018) book.

**Execute the Script to Download the Data onto a Computer.** The final step in the web scraping process is executing the script and downloading the data onto your computer. This phase of the process will be largely dependent on the type of data you are scraping and the format in which the data is received from the website. Using a combination of lists to store the data and post processing techniques (e.g., strip text, split text, relabel variables), you can prepare your data in *Python* (or *R*) to be downloaded/exported in a common data file format such as CSV (Comma-Separated Values) or XLSX (i.e., Microsoft Excel spreadsheet). In *Python*, the *pandas* and *numpy* packages can be used in conjunction to create a data frame and export the data in your desired format. A similar process could be followed in *R*, storing the data in data frames and exporting to .csv or .xlsx files via the *openxlsx* package.

### Computer-Aided Text Analysis (CATA)

There are two broad categories of CATA: inductive and deductive. Inductive CATA, is often used to assist in qualitative research, such as grounded theory or thematic analysis (Short, McKenny, & Reid, 2018). Researchers may use computer-aided qualitative data analysis software (CAQDAS; O'Kane, Smith, & Lerman, 2021) specifically designed for inductive CATA, such as ATLAS.ti, NVivo, or QDA Miner, to visualize relationships between themes, capture text segments, or to simply assist in the organization of the coding process. As the name suggests, inductive CATA is an exploratory approach that is guided by human coding.

Our focus, however, is on deductive CATA, which takes a top-down approach and transforms textual data into quantitative data, often based on word frequency in specific dictionaries or dictionary categories (McKenny et al., 2018). The dictionaries are developed based on the premise that



specific words are related to certain higher-order concepts/constructs (Pang & Ring, 2020). These dictionaries are typically developed and validated using some variation of Short and colleagues' (2010) recommendations to 1) create a working definition of the construct of interest, using a *priori* theory, 2) assess construct dimensionality based on existing literature, 3) develop an exhaustive list of key words that capture the construct of interest, and 4) consult with content experts to validate the word lists and assess rater reliability. If the research question does not necessitate the development of a new dictionary, researchers may use pre-existing validated dictionaries that are often available through published research or software programs.

There are several deductive CATA software programs available, including CAT Scanner (McKenny, Short, & Newman, 2012), DICTION (Hart & Carroll, 2012), and Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015). Additionally, there are several CATA packages in *R*, such as *sentimentr*, *tm*, *stm*, *syuhzet*, and *quanteda*. For this study, we chose LIWC for three reasons: 1) there are specific validated dictionaries that map onto our CIP construct very well (e.g., *focuspresent*, *focusfuture*, *focuspast*), 2) LIWC's theoretically-derived dictionaries can be used to establish content validity of our operationalization and 3) LIWC has been used in high-quality journals to operationalize language in the political leadership context (e.g., Sergent & Stajkovic, 2020).

**LIWC.** We used LIWC2015 (Pennebaker et al., 2015) to analyze both the Presidential addresses (for model training and validation) and the governors' COVID-19 press briefings (for model application). LIWC2015 is capable of analyzing text in a variety of formats, including plain text, PDF, RTF, .csv, and .xlsx. The program processes words sequentially throughout the text, adding incremental value to the dictionary category (or categories) to which the word belongs. If a word is present in more than one dictionary category, it is counted in all applicable categories. There are approximately 90 variables calculated throughout this process, including 41 dictionary categories designed to tap psychological constructs, as well as grammar/punctuation variables and other descriptive variables (e.g., word count, words per sentence, percentage of words longer than six letters). For reference, in Table 2, we provide a list of the variables we used in this study, along with sample words and the total number of words in each dictionary category.

**Establishing validity.** While CATA offers many psychometric advantages over traditional measures, such as self-report surveys and archival data (McKenny et al., 2018), the validity of the measure must be established (Short et al., 2010). To establish content validity of our CATA-based measure, after analyzing the Presidential addresses via LIWC2015, we assessed the differences among CIP style, based on theoretically relevant dictionary categories. For example, we conducted a one-way ANOVA and subsequent post-hoc Tukey test (Tukey, 1977) to assess whether *C*, *I*, and *P* differed, as theory suggests, in time orientation (Mumford, 2006). As you can see in Figure 2, *C* is significantly higher than *I* and *P* in the *focusfuture* variable, *I* is significantly higher than *C* and *P* in the *focuspast* variable, and *P* is significantly higher than *I* and *C* in the *focuspresent* variable. As time orientation is one of the key differentiating aspects among CIP styles, the fact that these LIWC2015 variables also captured these differences suggests that the text-based measure is appropriately differentiating among the CIP styles, thus, providing supporting evidence for the content validity.

## Supervised Machine Learning

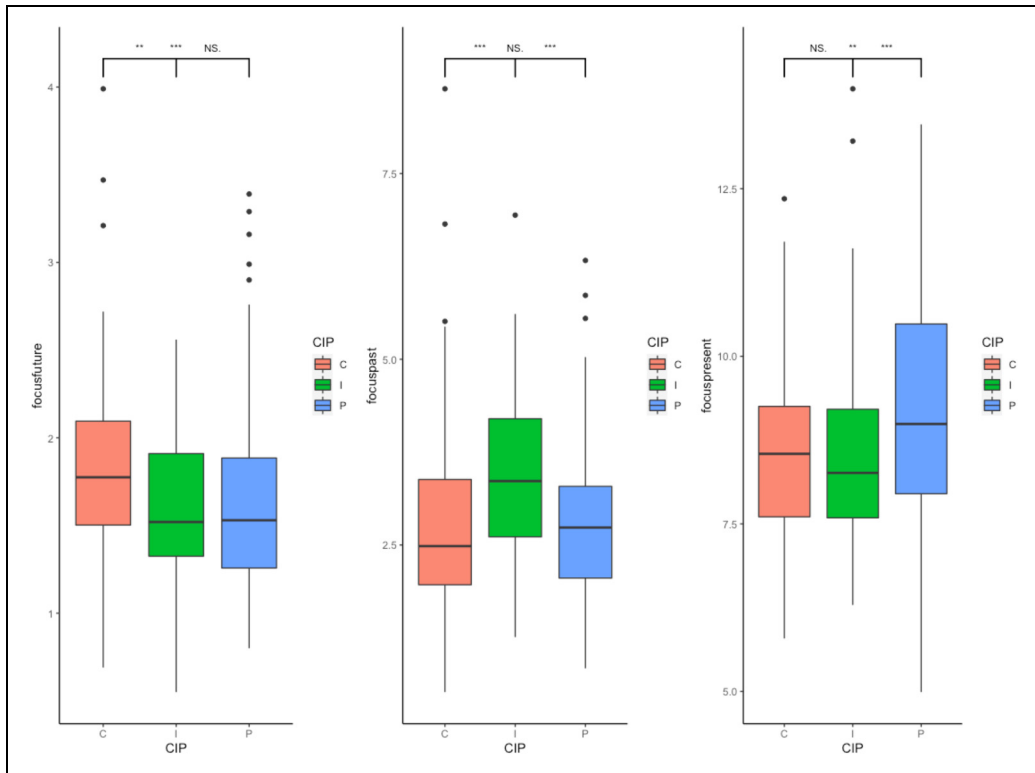
Supervised machine learning uses labeled input and output data (i.e., ground truth data) to build a model capable of predicting outcomes (e.g., classifications, labels) in unlabeled data (Janasik et al., 2009). In other words, with the labeled data, the model learns how predictor variables, referred to as *features* in supervised machine learning, map onto the known outcomes, such that those features can be used to predict unknown classifications in new data. Although the algorithms may vary based on the outcome variable of interest, supervised machine learning can be used to predict continuous

**Table 2.** Summary of Features, Respective LIWC Dictionary Categories, and Sample Words.

Feature Name (case sensitive)	Dictionary Category	Sample Words	Total Words in Category
achieve	Achievement	Goal, able, proud	213
adverb	Common adverbs	Very, really	140
affect	Affective processes	Best, loss, success	1393
article	Articles	A, an, the	3
assent	Assent	Agree, OK, cool	36
auxverb	Auxiliary verbs	Am, will, have	141
cogproc	Cognitive processes	Wiseest, doubtful, mistaken	797
Dash	Dashes	NA (punctuation)	NA
differ	Differentiation	Hasn't, but, else	81
filler	Fillers	I mean, you know	14
focusfuture	Future focus	Hopefully, ahead, someday	97
focuspast	Past focus	Accepted, knew, remember	341
focuspresent	Present focus	Present, depends, currently	424
hear	Hear	Listen, hearing	93
home	Home	Roommates, chores, families	100
informal	Informal language	NA	380
insight	Insight	Think, know	259
leisure	Leisure	Travel, play, parties	296
negate	Negations	No, not, never	62
nonflu	Non-fluencies	Er, hm, umm	19
number	Numbers	Second, thousand	36
prep	Prepositions	To, with, above	74
quant	Quantifiers	Few, many, much	77
risk	Risk	Hesitant, doubt, unsure	103
see	See	View, saw, seen	126
sexual	Sexual	Passion, sex, lover	131
social	Social processes	Party, danced, team	756
space	Space	Open, locals, border	360
swear	Swear words	Hell, damn, prick	131
time	Time	Date, today, current	310
Tone	Emotional tone	NA (summary category)	NA
verb	Common verbs	Eat, come, carry	1000
WC	Word count	NA (summary category)	NA
we	1 <sup>st</sup> person plural	We, us, our	12
work	Work	Company, startup, sector	444
WPS	Words per Sentence	NA	NA
you	2 <sup>nd</sup> person pronoun	You, your, ya'll	30

Note. Dictionary categories, sample words, and totals were identified in Pennebaker et al. (2015).

and categorical outcome variables. There are many dedicated packages in *R* that allow for the prediction of specific types of outcome variables (e.g., *nnet*, *e1071*, *kernlab*). For this study, however, we used the *caret* package (Kuhn, 2021) as it provides a highly flexible interface that allows for the execution of numerous algorithms, within the package, to predict whichever type of outcome variable is relevant to the study of interest. In other words, although we used *caret* to predict a categorical outcome variable (CIP), the same package can be used to predict continuous outcome variables, as well, by simply changing the “method” within the *caret* code. As such, our



**Figure 2.** Boxplots with Tukey post hoc of time orientation by CIP style.

Note. C, I, and P stand for charismatic, ideological, and pragmatic, respectively.

example (and R code) is generalizable to nearly any supervised machine learning model. The *caret* package is essentially a one-stop shop for supervised machine learning in R, as it is possible to complete each phase of the machine learning process within the package, including data pre-processing, prediction algorithms, feature selection, and cross-validation.

**Data Pre-Processing.** The overarching objective of supervised machine learning is to train a model capable of accurately predicting outcomes in unseen data. To minimize noise in the data, and maximize the accuracy of the model, there are certain pre-processing steps that can be taken. For example, centering is a pre-processing technique that subtracts the mean score from each feature (i.e., predictor variable). Scaling is another pre-processing technique that normalizes the feature data, often using either decimal scaling, min-max normalization, log transformation, or *z-score* normalization (Nwanganga & Chapple, 2020). In the *caret* package, the *scale* function simply divides each feature by its standard deviation, therefore, in the present study, we used both the *center* and *scale* options in the *caret* package to effectively implement a *z-score* normalization.

Another pre-processing consideration, for classification problems, such as predicting CIP, is to determine if the sample is balanced among the classes (e.g., CIP). In other words, is the sample relatively equally distributed among the classes? This is important because using supervised machine learning on a dataset with a large imbalance in the classes may result in high accuracy scores, however, the results can be misleading (Mohammed, Rawashdeh, & Abdullah, 2020). For instance, a poor performing model might assign every observation to the most represented category, and thereby look accurate, but the accuracy would purely be a byproduct of the data imbalance, not

model performance. Hence, if there is substantial imbalance, researchers often use data manipulation techniques, such as over- or under-sampling (referred to as *upsampling* or *downsampling* in the *caret* package) to balance the data among the classes (Nguyen, Cooper, & Kamei, 2012). Over-sampling is a process of randomly sampling instances from a minority class and then duplicating them until the minority class has the same number of instances as the majority class. Under-sampling is a process of selecting a number of random instances from the majority class that is equal to the number of instances in the minority class. For the present study, our sample had some imbalance among the CIP classes with 74, 83, and 220 instances per class, respectively; however, after comparing model performance metrics with and without over- and/or under-sampling, we determined that the data manipulation techniques did not improve the model. Thus, we did not use either over- or under-sampling. We did, however, ensure that the *k-folds* were stratified (see *Cross-Validation Methods* below for a description of *k-folds*), such that each fold had a relatively equal number of instances for each category (Kohavi, 1995).

**Prediction Models/Algorithms.** There are dozens of prediction algorithms used in supervised machine learning. Some of the algorithms are specifically used for continuous outcome variables, such as linear regression, lasso regression, and ridge regression; whereas others are used specifically for categorical outcomes, such as linear discriminant analysis, logistic model trees, naïve Bayes, and multinomial logistic regression; and others can predict both continuous and categorical outcomes. As CIP is a categorical outcome, we tested the predictive accuracy of eight algorithms that can classify categorical outcomes: 1) Linear Discriminant Analysis, 2) Logistic Model Trees, 3) Multinomial Logistic Regression, 4) Regularized Discriminant Analysis, 5) Neural Networks, 6) k-Nearest Neighbors, 7) Random Forests, and 8) Naïve Bayes. A summary of these eight algorithms is provided in Table 3 and a comparison of model performance is provided in Table 4 and Figure 3.

**Feature Selection.** When it comes to predictor variables or *features*, as they are referred to in supervised machine learning, having more is not always better. The decision on how many features to include in the model is often driven by sample size guidelines and model performance (e.g., accuracy). To reduce the number of features, researchers use a process referred to as *feature selection*, which is a process of identifying relevant features (i.e., predictor variables) and removing those that may be redundant or unimportant to the prediction model (Yu & Liu, 2004). Feature selection can be driven by theory, data, or, as is often the case, some combination of the two. For example, based on CIP theory, all time orientation variables (*focusfuture*, *focuspast*, *focuspresent*) should be important for the differentiation among CIP styles. Based on theory alone, this would suggest that these three features should be included in the model. However, including only these three features results in a poor performing model (mean accuracy = .58; mean kappa<sup>1</sup> = .10), so we needed to determine which other features to include.

As we chose to use multinomial logistic regression (MLR), we followed feature selection guidelines for MLR, which recommend having a minimum of ten instances per feature (Starkweather & Moske, 2011). Accordingly, with a sample size of 377 instances, we were only able to select a maximum of 37 features to include in our model. To identify which 37 features to include, we used the *varImp* function, of the *caret* package in *R*, to rank the features from most important to least important for classifying the CIP outcomes. As shown in Figure 4, our theoretically relevant time-orientation features are the 2<sup>nd</sup> (*focuspresent*), 8<sup>th</sup> (*focuspast*), and 60<sup>th</sup> (*focusfuture*) most important features in our model. Although *focusfuture* falls outside of the top 37, rather than dismiss a theoretically important feature, we instead chose the top 36 features and added *focusfuture* as the 37<sup>th</sup> feature in our final model.

**Cross-Validation Methods.** There are various approaches to cross-validation that can be used in supervised machine learning. For example, *Leave One Out Cross-Validation (LOOCV)* is the simplest cross-validation approach in which the data is divided into two non-overlapping parts and these two parts are used for training and validating the model, respectively (Yadav & Shukla,

**Table 3.** Classification Algorithm Descriptions.

Algorithm	Description
Linear Discriminant Analysis (LDA)	“The LDA technique is developed to transform the features into a lower dimensional space, which maximizes the ratio of the between-class variance to the within-class variance, thereby guaranteeing maximum separability” (Tharwat, Gaber, Ibrahim, & Hassanien, 2017, p. 170).
Logistic Model Trees (LMT)	“...consists of a standard decision tree with logistic regression at the leaves” (Landwehr, Hall, & Frank, 2005, p.173).
Multinomial Logistic Regression (MLR)	MLR is similar to the more oft-used logistic regression, with a primary difference being that it is used when there are more than two categories in the dependent variable. The goal of MLR is to model the odds of a particular categorical outcome, as a function of the model covariates (Hosmer & Lemeshow, 2015).
Regularized Discriminant Analysis (RDA)	RDA is an extension of LDA that is particularly useful for smaller samples with a large number of features, as it incorporates regularization parameters to decrease larger eigenvalues and increase smaller eigenvalues, which counteracts the bias inherent in sample-based estimation (Friedman, 1989).
Neural Networks (NNET)	“Neural networks are a wide class of flexible nonlinear regression and discriminant models, data reduction models, and nonlinear dynamical systems. They consist of an often large number of ‘neurons,’ i.e. simple linear or nonlinear computing elements, interconnected in often complex ways and often organized into layers” (Sarle, 1994, p. 1).
k-Nearest Neighbors (kNN)	“kNN classifier is to classify unlabeled observations by assigning them to the class of the most similar labeled examples” (Zhang, 2016, p. 2).
Random Forests (RF)	“Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest” (Breiman, 2001, p. 5).
Naïve Bayes (NB)	Bayesian classifiers assign the most likely class to a given instance, as it is described by its feature vector. Naïve bayes classifiers simplify this process by assuming that features are independent given class (Rish, 2001).

2016). The strength of this approach is that it reduces the likelihood of overfitting a model, however, it may not provide optimum model accuracy, as the training data is limited to one subset of the data. *Bootstrap Cross-Validation*, or *Repeated Hold Out Cross-Validation*, is another form of *LOOCV* in which the process of sub-setting the data into training and validation sets is repeated several times, with replacement (Kim, 2009). In other words, the training and validation data may be recycled as the data is randomly subset with each iteration. The strength of this approach is that it provides more training data, which can improve the accuracy of the model, however, there is an increased risk of overfitting the model to the idiosyncrasies of dataset. Similarly, *Bootstrap .632 Cross-Validation* randomly samples and resamples the data, with replacement, however the procedure adjusts for upward bias in the error rate (Efron & Tibshirani, 1997). Finally, *k-fold Cross-Validation (k-fold CV)* is a process in which the data is split into  $k$  equal parts from which the model is trained on  $k - 1$  parts and validated on the remaining one part (Kohavi, 1995). This process is then repeated  $k$  number of times, such that each fold is used exactly one time as the validation subset. With *k-fold CV*, the number of folds is arbitrary, however, 10 folds is the most popular choice (Kim, 2009). For this study, we chose to use *k-fold CV* with 10 folds and repeated the process three times, for a total of 30 iterations of training and validation. With each repeat of the *k-fold CV*, the decay rate, a hyper-parameter used to minimize overfitting (Ismail, Ahmad, Soh, Hassan, & Harith, 2019), was

**Table 4.** Algorithm Accuracy and Kappa Performance.

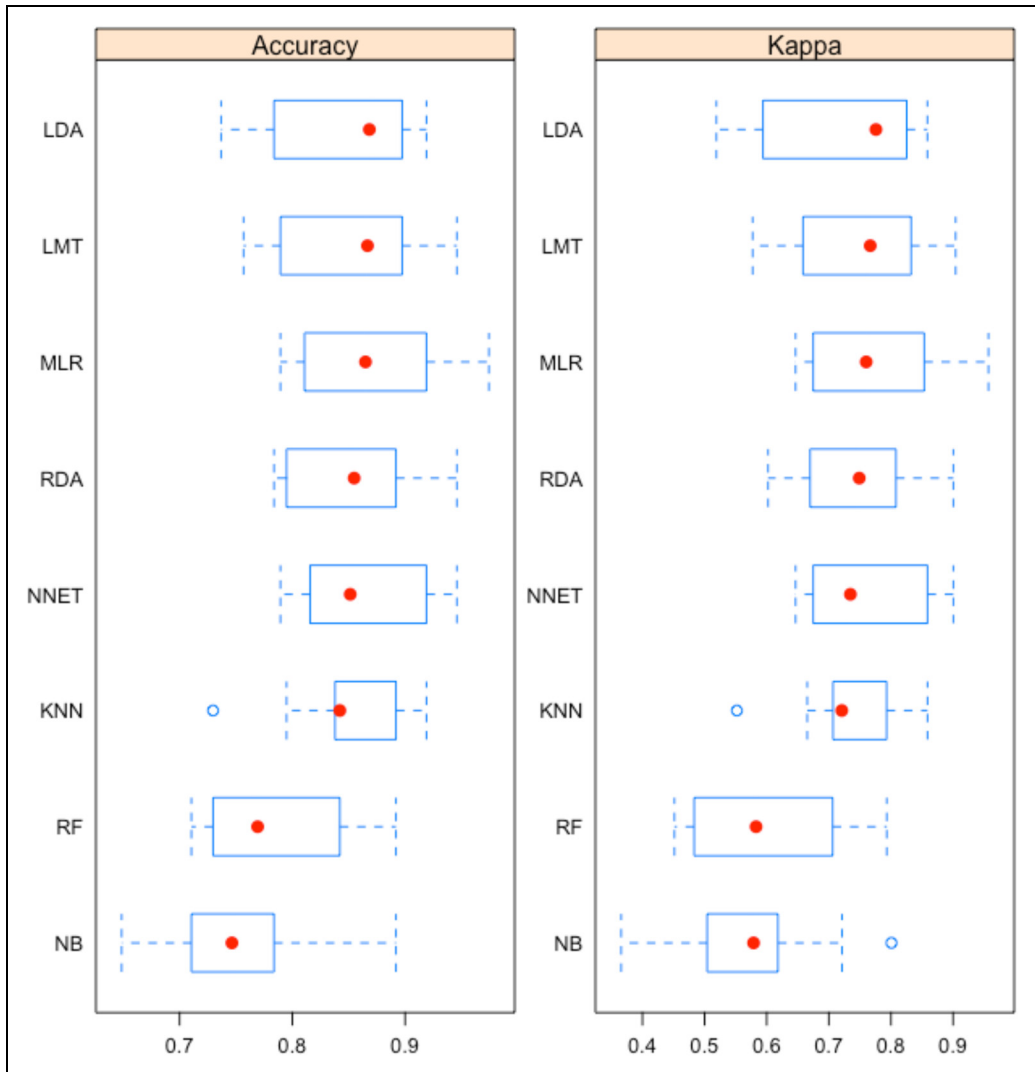
<b>Accuracy</b>						
	Min	Median	Mean	Max	SD	
MLR	0.789	0.865	0.867	0.974	0.061	
NNET	0.789	0.851	0.862	0.946	0.055	
LMT	0.757	0.867	0.857	0.946	0.061	
RDA	0.784	0.855	0.852	0.946	0.055	
LDA	0.737	0.868	0.846	0.919	0.070	
KNN	0.730	0.842	0.843	0.919	0.054	
RF	0.711	0.769	0.785	0.892	0.065	
NB	0.649	0.746	0.753	0.892	0.074	
<b>Kappa</b>						
	Min	Median	Mean	Max	SD	
MLR	0.647	0.760	0.768	0.957	0.105	
NNET	0.647	0.735	0.762	0.900	0.094	
LMT	0.577	0.767	0.750	0.904	0.105	
RDA	0.602	0.749	0.738	0.900	0.096	
LDA	0.519	0.776	0.731	0.859	0.125	
KNN	0.552	0.721	0.727	0.859	0.085	
RF	0.451	0.583	0.597	0.793	0.126	
NB	0.366	0.579	0.572	0.801	0.129	

Note. Each algorithm was run using all available features (i.e., all LIWC dictionaries). MLR = multinomial logistic regression, NNET = neural network, LMT = logistic model trees, RDA = regularized discriminant analysis, LDA = linear discriminant analysis, KNN = *k*-nearest neighbors, RF = random forests, NB = naïve Bayes.

automatically manipulated within the *caret* package (0, .0001, and .1, respectively) to identify an optimum decay rate for model accuracy. As our decision was purely data-driven, based on the model performance, an optimum decay rate of .1 was selected for our model. The results of all 30 training-validation iterations are presented in Table 5. To better illustrate the *k-fold* CV process, in Figure 5, we provide a visual depiction of the process with 10 folds.

### **Multinomial Logistic Regression (MLR)**

After comparing the performance of eight different machine learning algorithms, we settled on using multinomial logistic regression (MLR) for our model, as it had the highest accuracy and kappa values, on average. Commonly referred to as a *discrete choice model*, MLR is similar to the more oft-used logistic regression, with a primary difference being that it is used when there are more than two categories in the dependent variable. In general, the goal of MLR is to model the odds of a particular categorical outcome, as a function of the covariates (i.e., features) in the model (Hosmer & Lemeshow, 2015). Like binary logistic regression, MLR uses maximum likelihood to estimate the probability of category membership (Starkweather & Moske, 2011). More precisely, we used MLR, in this study, to construct a model capable of explaining the relationship between the features (LIWC categories) and the labeled outcomes (CIP styles), so that the model could accurately predict CIP in the governors' COVID-19 press briefings, for which we did not have a CIP classification. As previously noted, although there are multiple *R* packages available to conduct MLR (e.g., *mnlogit*, *VGAM*, *maxent*, *nnet*), we used *caret*



**Figure 3.** Algorithm accuracy and kappa comparisons.

*Note.* Accuracy and Kappa were calculated using *k-fold cross validation* with 10 folds and 3 repeats, resulting in 30 iterations. The red dots represent the median results for each algorithm. LDA = Linear Discriminant Analysis, LMT = Logistic Model Trees, MLR = Multinomial Logistic Regression, RDA = Regularized Discriminant Analysis, NNET = Neural Network, KNN = *k*-Nearest Neighbors, RF = Random Forest, and NB = Naïve Bayes.

(Kuhn 2021) for this study. For reference, our script is provided in *R* markdown format in Appendix D.

**Model Specification.** Prior to running an MLR, the researcher has to establish which variables to include in the model. There are multiple ways to select variables and, ultimately, the researcher uses a combination of theory, statistics, and discretion to choose the final model. As described above, we used the *varImp* function, of the *caret* package in *R*, to identify the most important features for classifying the CIP outcomes, ultimately choosing 37 features to include in the final model.

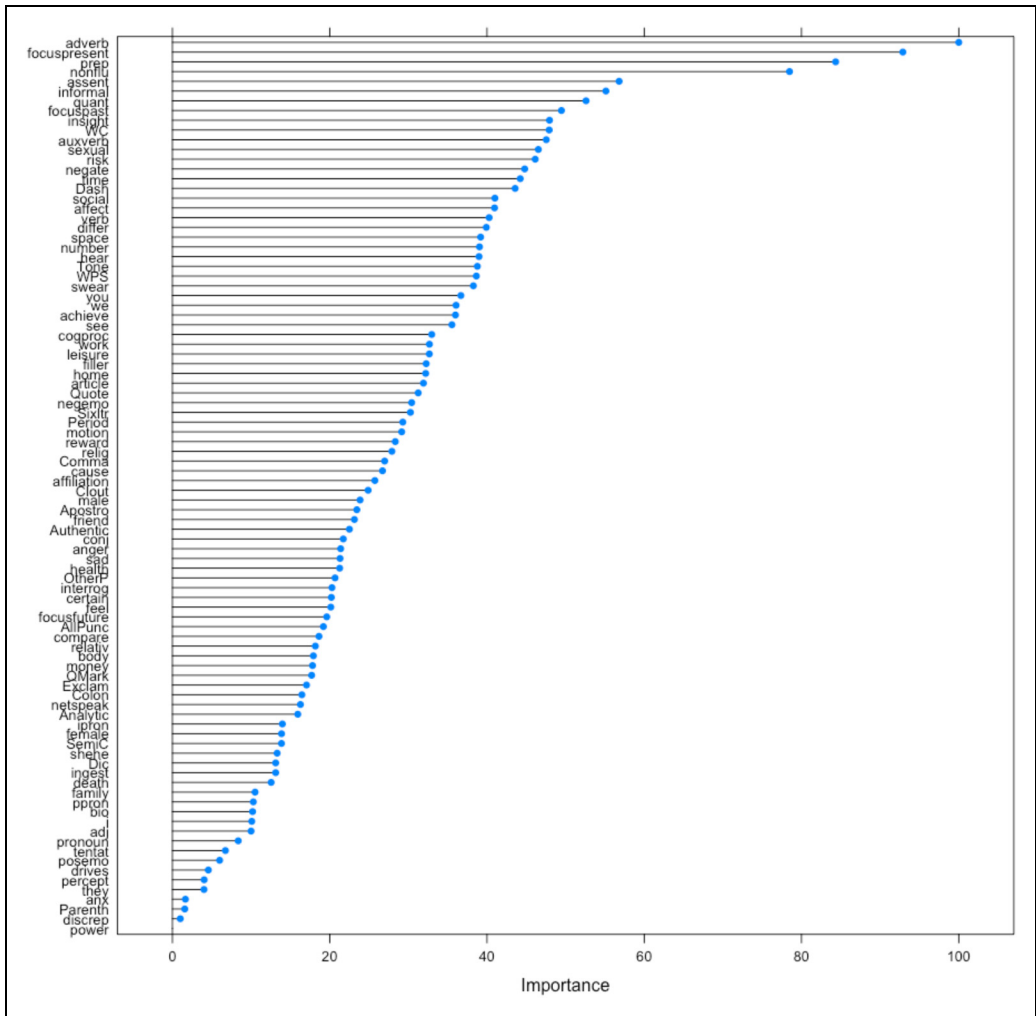


Figure 4. Model features ranked in order of importance for CIP prediction.

Following notation from Hastie and colleagues (2009), the resulting model can be expressed in the following two equations:

$$\log\left(\frac{\Pr(CIP = I)}{\Pr(CIP = C)}\right) = \beta_{10} + \beta_{11}(achieve) + \beta_{12}(adverb) \dots \beta_{137}(you) + \epsilon \quad (1)$$

$$\log\left(\frac{\Pr(CIP = P)}{\Pr(CIP = C)}\right) = \beta_{20} + \beta_{21}(achieve) + \beta_{22}(adverb) \dots \beta_{237}(you) + \epsilon \quad (2)$$

Equation 1 represents the logit, where a one-unit increase (or decrease) in the independent variable(s) is associated with an increase (or decrease) of the log odds for the model selecting *I* versus *C*. Similarly, Equation 2 represents the logit, where a one-unit increase (or decrease) in the independent variable(s) is associated with an increase (or decrease) of the log odds for the model selecting *P* versus *C*. The final model had high mean accuracy (87.7%) and kappa (78.3%) values with relatively stable performance



**Table 5.** Fit and Accuracy Summary of Final Multinomial Logistic Regression Model.

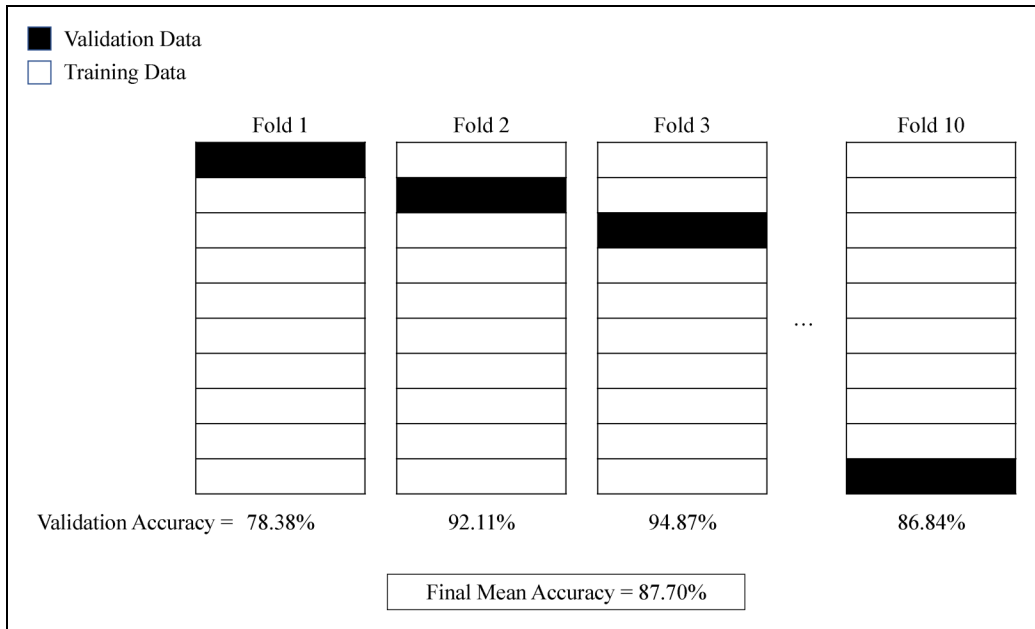
Accuracy	Kappa	decay	Resample
0.730	0.545	0	Fold01
0.784	0.623	0.1	Fold01
0.757	0.583	0.0001	Fold01
0.895	0.814	0	Fold02
0.921	0.856	0.1	Fold02
0.921	0.861	0.0001	Fold02
0.872	0.785	0	Fold03
0.949	0.914	0.1	Fold03
0.897	0.827	0.0001	Fold03
0.757	0.560	0	Fold04
0.784	0.601	0.1	Fold04
0.757	0.560	0.0001	Fold04
0.811	0.670	0	Fold05
0.865	0.765	0.1	Fold05
0.838	0.712	0.0001	Fold05
0.892	0.815	0	Fold06
0.838	0.713	0.1	Fold06
0.892	0.815	0.0001	Fold06
0.919	0.859	0	Fold07
0.892	0.808	0.1	Fold07
0.865	0.765	0.0001	Fold07
0.974	0.957	0	Fold08
0.949	0.913	0.1	Fold08
0.974	0.957	0.0001	Fold08
0.842	0.725	0	Fold09
0.921	0.860	0.1	Fold09
0.816	0.684	0.0001	Fold09
0.789	0.628	0	Fold10
0.868	0.775	0.1	Fold10
0.816	0.686	0.0001	Fold10

Note. Accuracy is the percentage of predictions that matched with the known outcomes. Kappa is the percentage of predictions that matched with known outcomes, controlling for the expected accuracy, based on class distribution. Decay is a regularization parameter used to avoid overfitting.

across cross-validation samples (accuracy  $SD = .06$ , kappa  $SD = .11$ ). To view the full model, including the coefficients, standard errors, and Wald statistics for each independent variable, see Table 6.

**Application to New Data.** After systematically building, training, and validating the model, we finally applied the model to the governors' COVID-19 press briefing dataset. Unlike the Presidential dataset, there were no known CIP outcomes in the governor dataset. Hence, there were no accuracy scores to be calculated. Instead, we relied upon the training/validating procedure in the Presidential address dataset to build our confidence in the model's ability to accurately assess CIP in the governor dataset. Ultimately, the model classified the press briefings based on the CIP category with the highest probability of being correct. A subset of the modeled probabilities is provided in Table 7.

**Model Validation.** After applying the model to the governors' press briefings, we consulted with three subject matter experts (SME), who were also familiar with some of the governors' press briefings, to



**Figure 5.** Visual illustration of k-fold cross validation.

Note. The final mean accuracy is based on 10 folds with a decay value of .10.

establish face validity of the measure of CIP in governors. Each SME we consulted has published multiple scholarly journal articles and book chapters on CIP. Each SME conversation was separate, and we did not define any criteria for agreement. After showing the CIP results to the SMEs, and having a discussion, there were two points of contention: 1) lack of charismatic representation in the governors' press briefings and 2) lack of stability in the CIP style over time. For example, one scholar stated, "I remember 'learning' that charismatic leadership tends to emerge during times of crisis – when people look for inspiration and direction, so I was surprised to not see as much charisma, given COVID." However, after discussing the problem solving strengths of different CIP styles (Griffith et al., 2018; Lovelace et al., 2019) and the fact that COVID-19 is not a typical crisis event (i.e., it is ongoing and people are dying every day), the scholar reconsidered her initial reaction and acknowledged, "That makes sense and is interesting". Regarding stability of CIP overtime, the initial push back (only from one SME) was alleviated after we discussed the limitations of the historiometric approach, used in prior empirical studies, to monitor the fine-grained detail of CIP overtime. Further, we discussed the anecdotal evidence of a situational aspect of CIP, referring to Barack Obama's shift from a charismatic campaigner to a pragmatic President (Yammarino, Sotak, & Serban, 2020), which helped to validate the findings for that SME. Another SME stated, "I always thought CIP stuff was stable," but added, "validation evidence [is] good." Ultimately, after comparing our results with their expected results, the SMEs indicated that the measure was assessing CIP in the governors with relative accuracy. As such, this step provided further support for content validity (Short et al., 2010).

As a reminder, the CIP styles of the Presidents were established in a human-coded historiometric study (Yammarino et al., 2013). Thus, in addition to establishing support for content validity, these results also suggest convergent validity (Campbell & Fiske, 1959). Regarding external validity, we chose Presidential addresses to develop our text-based model because of the similarity in context (i.e., political leadership) and content (addresses and briefings) to our test sample of U.S. state governors' COVID-19 press briefings.

**Table 6.** Multinomial Logistic Regression Coefficients, Standard Errors, and Wald Statistics.

Variable Name	I			P		
	Coeff.	Std. Err.	z	Coeff.	Std. Err.	z
Intercept	0.56	1.21	0.47	4.87	0.93	5.24
achieve	-1.53	0.80	-1.92	-0.49	0.62	-0.79
adverb	2.56	1.36	1.88	1.81	1.14	1.58
affect	-0.11	1.10	-0.10	-1.77	0.77	-2.29
article	0.30	0.88	0.35	-1.32	0.74	-1.79
assent	0.99	1.05	0.94	1.45	0.94	1.54
auxverb	0.86	1.20	0.71	-0.88	0.89	-0.99
cogproc	0.28	1.56	0.18	0.02	1.23	0.01
Dash	1.11	1.30	0.86	2.29	1.18	1.94
differ	0.38	1.16	0.33	0.55	0.94	0.58
filler	-1.66	0.96	-1.73	-0.16	0.53	-0.30
focusfuture	0.52	0.74	0.71	-0.29	0.60	-0.48
focuspast	1.08	1.36	0.79	-0.09	1.17	-0.08
focuspresent	-2.12	2.33	-0.91	1.16	1.94	0.60
hear	1.37	0.99	1.38	0.68	0.79	0.86
home	-0.45	0.79	-0.57	0.20	0.58	0.34
informal	1.45	1.41	1.03	0.45	1.08	0.41
insight	1.03	1.08	0.96	-0.61	0.84	-0.72
leisure	0.81	0.87	0.93	-0.62	0.64	-0.96
negate	1.56	1.05	1.48	-0.08	0.78	-0.10
nonflu	3.48	1.37	2.54	-0.15	1.09	-0.14
number	1.70	0.87	1.96	-0.16	0.71	-0.22
prep	2.46	0.86	2.85	1.08	0.70	1.53
quant	-1.24	0.80	-1.55	-0.88	0.66	-1.34
risk	-0.64	0.93	-0.69	-1.16	0.67	-1.74
see	-1.22	0.83	-1.47	-0.60	0.65	-0.92
sexual	-2.72	1.28	-2.13	0.03	0.64	0.05
social	-0.91	1.19	-0.76	-1.55	0.95	-1.62
space	-0.53	0.79	-0.67	-0.74	0.63	-1.17
swear	-0.24	1.87	-0.13	-1.23	1.38	-0.89
time	1.25	0.91	1.37	1.50	0.76	1.99
Tone	0.23	0.93	0.25	-0.53	0.66	-0.81
verb	-1.00	2.69	-0.37	-0.70	2.29	-0.31
WC	1.09	0.77	1.42	-0.99	0.58	-1.70
we	2.25	1.12	2.01	1.94	0.95	2.04
work	0.30	0.82	0.37	-0.53	0.65	-0.81
WPS	0.55	0.92	0.60	1.25	0.82	1.52
you	-1.53	1.07	-1.43	0.14	0.69	0.20

Note. Coeff. = coefficient, Std. Err. = standard error, and z = Wald statistic.

## Results for Illustration

A summary of the results is provided in Table 8. In brief, the breakdown of the press briefing CIP categorization was as follows: 3 charismatic (< 0.01%), 174 ideological (20.84%), and 658 pragmatic (78.80%). This means that less than 1% of the press briefings used language associated with a charismatic approach, such as a future-oriented and positive emotional expression of the governor's vision.

**Table 7.** Sample Predictions with Probabilities for Each CIP Category.

Press Briefing	Model Prediction	Probability C	Probability I	Probability P
1	P	0.00%	0.00%	100.00%
2	I	0.00%	99.40%	0.60%
3	P	0.00%	8.80%	91.20%
4	I	0.00%	81.70%	18.30%
5	P	0.00%	2.40%	97.60%
6	I	0.00%	99.20%	0.80%
7	I	0.00%	68.00%	32.00%
8	I	0.00%	61.30%	38.70%
9	P	0.00%	0.00%	100.00%
10	P	0.00%	0.00%	100.00%
11	P	0.00%	0.10%	99.90%
12	P	0.00%	8.30%	91.70%
13	P	0.00%	0.20%	99.80%
14	P	0.00%	0.00%	100.00%
15	P	0.00%	0.10%	99.90%
16	P	0.00%	30.00%	70.00%
17	I	0.00%	99.10%	0.90%
18	P	0.00%	0.00%	100.00%
19	P	0.00%	15.00%	85.00%
20	P	0.00%	0.20%	99.80%
21	P	0.00%	0.00%	100.00%
22	P	0.00%	0.10%	99.90%
23	P	0.00%	16.20%	83.80%
24	I	0.00%	75.50%	24.50%
25	I	0.00%	99.60%	0.40%
26	P	0.00%	0.80%	99.20%
27	P	0.00%	0.40%	99.60%
28	P	0.00%	0.00%	100.00%
29	P	0.00%	0.20%	99.80%
30	P	0.00%	0.20%	99.80%

Note. C, I, and P stand for *charismatic*, *ideological*, and *pragmatic*, respectively. Percentages represent the likelihood that a particular observation (i.e., press briefing) falls into each CIP category, based on the application of a trained multinomial logistic regression model.

With thousands of people dying every day, a charismatic approach, as defined by the CIP model (Mumford, 2006), would have likely been socially inappropriate. With nearly 80% of the briefings classified as *pragmatic*, the data suggests that governors, in general, communicated a data-driven approach to solving the present issues, without dwelling on the past or focusing on the future, in response to the COVID-19 pandemic. This makes sense as there have been, and continue to be, ample data collected and analyzed throughout the pandemic. Finally, 20% of the briefings were classified as *ideological*, which suggests that some governors occasionally used a communication strategy focused on the past and filled with negative emotion. Again, with thousands of people dying every day and a longing for life as we knew it (i.e., the past), it makes sense that some governors used an ideological style.

An important aspect of these results is the within-person variability of the CIP classifications over time, which is shown in Figure 6. Prior theorizing and empirical work on the CIP model have found and/or assumed CIP to be a trait-like, fixed quality in individuals. Perhaps the lack of granularity in the historiometric methodology, which has dominated the study of CIP to date

**Table 8.** Total Governor Press Briefings Classified by CIP Category.

Governor Name	State	C	I	P
Governor Abbott	TX	0	8	18
Governor Baker	MA	0	7	19
Governor Beshear	KY	0	0	12
Governor Brown	OR	0	0	9
Governor Carney	DE	0	0	1
Governor Cooper	NC	0	0	25
Governor Cuomo	NY	3	36	121
Governor DeSantis	FL	0	12	50
Governor DeWine	OH	0	24	60
Governor Ducey	AZ	0	0	6
Governor Edwards	LA	0	2	16
Governor Evers	WI	0	0	2
Governor Gordon	WY	0	0	1
Governor Grisham	NM	0	0	1
Governor Herbert	UT	0	0	2
Governor Hogan	MD	0	5	11
Governor Holcomb	IN	0	1	3
Governor Hutchinson	AR	0	0	4
Governor Inslee	WA	0	1	17
Governor Ivey	AL	0	3	4
Governor Justice	WV	0	0	1
Governor Kelly	KS	0	1	3
Governor Kemp	GA	0	3	5
Governor Lamont	CT	0	2	8
Governor Lee	TN	0	0	7
Governor Little	ID	0	1	1
Governor McMaster	SC	0	1	6
Governor Murphy	NJ	0	15	33
Governor Newsom	CA	0	47	44
Governor Noem	SD	0	0	3
Governor Northam	VA	0	0	37
Governor Parson	MO	0	0	2
Governor Polis	CO	0	0	6
Governor Pritzker	IL	0	2	39
Governor Raimondo	RI	0	0	3
Governor Reeves	MS	0	1	2
Governor Reynolds	IA	0	1	13
Governor Sisolak	NV	0	0	2
Governor Stitt	OK	0	0	1
Governor Walz	MN	0	0	6
Governor Whitmer	MI	0	1	46
Governor Wolf	PA	0	0	8
TOTAL		3	174	658

Note. C, I, and P stand for *charismatic*, *ideological*, and *pragmatic*, respectively. Predictions were made using the Multinomial Logistic Regression algorithm.

(Lovelace et al., 2019), led researchers to perceive and theorize that CIP style is relatively stable in individuals. With our CATA and machine learning operationalization, however, we were able to

capture the within-person change over time that has previously remained untested. Although there was variability, most governors tended to have a dominant style. For example, Governor Cooper (North Carolina) had all 25 briefings classified as *P*, Governor Beshear (Kentucky) had all 12 briefings classified as *P*, and Governor Northam (Virginia) had all 37 briefings classified as *P*. Others were more balanced between two styles. For example, Governor Newsom (California) had 44 briefings classified as *P* and 47 classified as *I*, and Governor DeWine (Ohio) had 60 briefings classified as *P* and 24 classified as *I*. We attribute the relative lack of charismatic press briefings to the fact that these briefings all took place during a global pandemic. The positive emotional imagery and focus on the future that characterizes the charismatic style of leadership may have been difficult to portray, and likely ill-advised, when thousands of individuals were dying on a daily basis.



**Figure 6.** Governors’ press briefings classified as C, I, or P over time. Note. C, I, and P stand for *charismatic*, *ideological*, and *pragmatic*, respectively. Predictions were made using a multinomial logistic regression algorithm.

## Discussion

We introduced a method, that combines web scraping, CATA, and supervised machine learning techniques to advance new streams of research. Leveraging labeled ground truth data, from a historic study (Yammarino et al., 2013), we illustrated how this method and operationalization can be used to predict categorical outcomes. In addition to classifying categorical outcomes, a similar approach could be used to assess continuous variables, such as personality (Harrison, Thurgood, Boivie, & Pfarrer, 2019) or implicit motives (Schultheiss, 2013), as well. Hence, this methodological combination could be applied in a variety of research disciplines and contexts. The process, as we have described, may be especially useful to researchers who study populations that are difficult to assess via traditional research methods (e.g., upper echelons). Not all aspects of the method, however, need to be used in conjunction, as we have done. Instead, we envision each aspect of our method (web scraping, CATA, machine learning) as a stand-alone module that can aide micro-, meso-, and macro-oriented researchers as they expand their research tool kits. In brief, not only does this text-based approach open doors to new data sources, but it may also spur new streams of theoretical inquiries.

### *Method Limitations and Future Research*

There are certainly limitations to our study and this method. First, although 87% seems to be high accuracy percentage for predicting CIP, based on running text, there is still a 13% chance that the model predicted the CIP style incorrectly. The second limitation to this method is that it purely relies upon words and dictionary categories to classify individuals, without any other behavioral or contextual information factored into the model. Although this method appears to explain a large portion of the variance among CIP styles, it may be missing some valuable information that cannot be fully captured from the words alone. A third limitation, to this particular study, is the fact that all of the press briefings took place during a global pandemic. The gravity of the situation may have overpowered an individual leader's natural style and necessitated a certain style of communication. Although this provides additional theoretical considerations for the study of CIP, it may place a constraint on our method's ability to generalize to less situationally constrained times (i.e., normal, non-crisis life). On the other hand, this limitation could also be viewed as a strength, as restricting our sample to a consistent context may minimize situational confounds and allow for CIP variance to be attributed to the individuals, rather than the context.

We put forth this method in hopes of stimulating and generating new streams of future research. By combining web scraping, CATA, and supervised machine learning techniques there are many possibilities. For example, researchers can use other ground truth datasets in which data are classified by certain categorical variables and apply the approach demonstrated in this study. Similarly, researchers could identify ground truth datasets in which individuals are scored on continuous scales of other psychological constructs (e.g., personality, goal orientation) and apply a similar, modified approach. With a continuous variable as an outcome, rather than using multinomial logistic regression, the researcher would simply train a linear model using one of the various algorithms available in the *caret* package. With access to paid textual datasets, such as LexisNexis Full Disclosure Wire or Wharton Research Data Service (WRDS), one could skip the web scraping portion of the method and jump right into the CATA and machine learning. Beyond the use of ground truth datasets and archival textual data, new online communication platforms (e.g., Zoom) provide another avenue for using the methods we put forth in this manuscript. For example, researchers could collect psychometric data from participants and have the participants engage in some sort of online Zoom activity (or perhaps use an already recurring Zoom activity, such as a college course), save the transcripts from the Zoom session, and proceed with the above-mentioned methods. This

method could also be applied in mixed methods studies in which researchers collect both survey data and qualitative data from participants. Put simply, this method is broadly applicable to a variety of research topics and disciplines.

### *Theoretical Implications for CIP*

While the primary focus of this manuscript was to advance a research method for using textual data to classify outcomes, there are some theoretical implications as well. Similar to the way in which experience sampling methods (ESM; Gabriel et al., 2019) allowed researchers to capture the within-person variation of personality over time (Fleeson, 2001, 2004), our text-based operationalization of CIP will allow researchers to further examine the stability assumptions of CIP that were previously untestable using historiometric methods (Lovelace et al., 2019). Stated differently, our method of assessing CIP from running text allows researchers to examine CIP from multiple levels of analysis (e.g., within and between individuals), over time, and across situations, rather than being limited to merely aggregating CIP across time and situations, assuming any within-person fluctuations to be error.

## **Conclusion**

There is a blue ocean of text available on the internet. There is also a plethora of research studies conducted on public figures. We put forth a method that combines web scraping, CATA, and supervised machine learning techniques to leverage these data sources to classify categorical outcomes. To illustrate the implementation of this methodology, we classified U.S. governors' COVID-19 press briefings according to their categorical CIP leadership styles. Additionally, as this method is highly flexible, we opened the door for other researchers to build their research streams as well.

## **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## **ORCID iD**

Jason D. Marshall  <https://orcid.org/0000-0001-9899-9307>

## **Supplemental Material**

Supplemental material for this article is available online.

## **Note**

1. Kappa is the percentage of predictions that matched with known outcomes, controlling for the expected accuracy, based on class distribution. In other words, kappa is a weighted accuracy metric.

## **References**

Aguinis, H., & Bakker, R. M. (2021). Time is of the essence: Improving the conceptualization and measurement of time. *Human Resource Management Review*, 31(2). <https://doi.org/10.1016/j.hrmr.2020.100763>



- Aguinis, H., Suarez-Gonzalez, I., Lannelongue, G., & Joo, H. (2021). Scholarly impact revisited. *Academy of Management Perspectives*, 26(2), 105–132. <https://doi.org/10.5465/amp.2011.0088>
- Antonakis, J., Bastardo, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), 293–319. <https://doi.org/10.1146/annurev-orgpsych-041015-062305>
- Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, 33(4), 445–459. <https://doi.org/10.1007/s10869-017-9528-3>
- Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *Leadership Quarterly*, 15(2), 211–239. <https://doi.org/10.1016/j.leaqua.2004.02.005>
- Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods*, 21(3), 633–659. <https://doi.org/10.1177/1094428117690235>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- DeTienne, K. B., DeTienne, D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, 6(2), 236–265. <https://doi.org/10.1177/1094428103251907>
- Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The .632+ bootstrap method. *Journal of American Statistical Association*, 92(438), 548–560. Retrieved from <http://www.jstor.org/stable/2965703>.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6), 1011–1027. <https://doi.org/10.1037/0022-3514.80.6.1011>
- Fleeson, W. (2004). Moving personality beyond the person-situation debate. *Current Directions in Psychological Science*, 13(2), 83–87. <https://doi.org/10.1111/j.0963-7214.2004.00280.x>
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405), 165–175. <https://doi.org/10.1080/01621459.1989.10478752>
- Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods*, 22(4), 969–1006. <https://doi.org/10.1177/1094428118802626>
- Griffith, J. A., Gibson, C., Medeiros, K., MacDougall, A., Hardy, J., & Mumford, M. D. (2018). Are you thinking what i'm thinking?: The influence of leader style, distance, and leader–follower mental model congruence on creative performance. *Journal of Leadership and Organizational Studies*, 25(2), 153–170. <https://doi.org/10.1177/1548051817750537>
- Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review*, 32(2), 334–343. <https://doi.org/10.5465/AMR.2007.24345254>
- Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2), 193–206. <https://doi.org/10.2307/258434>
- Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, 40(8), 1316–1330. <https://doi.org/10.1002/smj.3023>
- Hart, R. P., & Carroll, C. E. (2012). *DICTION: The text analysis program: Help manual*. Digitext, Inc.
- Hastie, T., Tibshirani, R. J., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hosmer, D. W., & Lemeshow, S. (2015). *Applied logistic regression*. (N. A. C. Cressie, N. I. Fisher, I. M. Johnstone, J. B. Kadane, D. W. Scott, B. W. Silverman, ... D. G. Kendall, eds.) (2nd ed.). John Wiley & Sons, Inc.

- House, R. J., & Howell, J. M. (1992). Personality and charismatic leadership. *The Leadership Quarterly*, 3(2), 81–108. [https://doi.org/10.1016/1048-9843\(92\)90028-E](https://doi.org/10.1016/1048-9843(92)90028-E)
- Hunter, S. T., Cushenbery, L., Thoroughgood, C., Johnson, J. E., & Ligon, G. S. (2011). First and ten leadership: A historiometric investigation of the CIP leadership model. *Leadership Quarterly*, 22(1), 70–91. <https://doi.org/10.1016/j.leaqua.2010.12.008>
- Ismail, A., Ahmad, S. A., Soh, A. C., Hassan, K., & Harith, H. H. (2019). Improving convolutional neural network (CNN) architecture (miniVGGNet) with batch normalization and learning rate decay factor for image classification. *International Journal of Integrated Engineering*, 11(4), 51–59. <https://doi.org/10.30880/ijie.2019.11.04.006>
- Janasik, N., Honkela, T., & Bruun, H. (2009). Text mining in qualitative research. *Organizational Research Methods*, 12(3), 436–460. <https://doi.org/10.1177/1094428108317202>
- Kim, J. H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics and Data Analysis*, 53(11), 3735–3745. <https://doi.org/10.1016/j.csda.2009.04.009>
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733–765. <https://doi.org/10.1177/1094428117722619>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 1137–1145).
- Kuhn, M. (2021). Package ‘caret’.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, 59, 161–205. <https://doi.org/10.1007/s10994-005-0466-3>
- Lawson, C. (2020). Personal communication. Rev.com.
- Ligon, G. S., Harris, D. J., & Hunter, S. T. (2012). Quantifying leader lives: What historiometric approaches can tell us. *Leadership Quarterly*, 23(6), 1104–1133. <https://doi.org/10.1016/j.leaqua.2012.10.004>
- Lovelace, J. B., Neely, B. H., Allen, J. B., & Hunter, S. T. (2019). Charismatic, ideological, & pragmatic (CIP) model of leadership: A critical review and agenda for future research. *Leadership Quarterly*, 30(1), 96–110. <https://doi.org/10.1016/j.leaqua.2018.08.001>
- McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: improving the accuracy of computer-aided text analysis. *Journal of Management*, 44(7), 2909–2933. <https://doi.org/10.1177/0149206316657594>
- McKenny, A. F., Short, J. C., & Newman, S. M. (2012). CAT Scanner. Retrieved from <http://www.catscanner.net/>.
- McKenny, A. F., Short, J. C., & Payne, G. T. (2012). Using computer-aided text analysis to elevate constructs: an illustration using psychological capital. *Organizational Research Methods*, 16(1), 152–184. <https://doi.org/10.1177/1094428112459910>
- Minbashian, A., Bright, J. E. H., & Bird, K. D. (2010). A comparison of artificial neural networks and multiple regression in the context of research on personality and work performance. *Organizational Research Methods*, 13(3), 540–561. <https://doi.org/10.1177/1094428109335658>
- Mitchell, R. (2018). *Web scraping with python. (A. MacDonald & J. Billing, Eds.) (2nd ed.)*. O'Reilly Media, Inc.
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine learning with oversampling and undersampling techniques: Overview study and experimental results. 2020 11th International Conference on Information and Communication Systems, ICICS 2020, 243–248. <https://doi.org/10.1109/ICICS49469.2020.239556>.
- Mumford, M. D. (2006). *Pathways to outstanding leadership: A comparative analysis of charismatic, ideological, and pragmatic leaders*. Lawrence Erlbaum Associates.
- Nguyen, H. M., Cooper, E. W., & Kamei, K. (2012). A comparative study on sampling techniques for handling class imbalance in streaming data. 6th International Conference on Soft Computing and Intelligent Systems, and 13th International Symposium on Advanced Intelligence Systems, SCIS/ISIS 2012, 1762–1767. <https://doi.org/10.1109/SCIS-ISIS.2012.6505291>.

- Nwanganga, F., & Chapple, M. (2020). *Practical machine learning in R*. Wiley & Sons, Inc. <https://doi.org/10.1002/9781119591542>.
- O’Kane, P., Smith, A., & Lerman, M. P. (2021). Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *Organizational Research Methods*, 24(1), 104–139. <https://doi.org/10.1177/1094428119865016>
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, 22(3), 765–797. <https://doi.org/10.1177/1094428117745648>
- Pang, J. S., & Ring, H. (2020). Automated coding of implicit motives: A machine-learning approach. *Motivation and Emotion*, 44(4), 549–566. <https://doi.org/10.1007/s11031-020-09832-8>
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015 operator’s manual*. Pennebaker Conglomerates. Retrieved from [www.LIWC.net](http://www.LIWC.net)
- Pollach, I. (2012). Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. *Organizational Research Methods*, 15(2), 263–287. <https://doi.org/10.1177/1094428111417451>
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rev. (2020). Terms of service. Retrieved from <https://www.rev.com/about/terms>.
- Richardson, L. (2020). Beautiful soup documentation.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence* (pp. 41–46).
- Sarle, W. S. (1994). Neural networks and statistical models. In *Nineteenth Annual SAS Users Group International Conference, April, 1994* (pp. 1–13). <https://doi.org/10.5023/jappstat.24.77>.
- Schultheiss, O. C. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in Psychology*, 4(OCT), 1–20. <https://doi.org/10.3389/fpsyg.2013.00748>
- Serban, A., Yammarino, F. J., Sotak, K. L., Banoeng-Yakubo, J., Mushore, A. B. R., Hao, C., & Mumford, M. D. (2018). Assassination of political leaders: The role of social conflict. *Leadership Quarterly*, 29(4), 457–475. <https://doi.org/10.1016/j.leaqua.2018.01.003>
- Sergent, K., & Stajkovic, A. D. (2020). Women’s leadership is associated with fewer deaths during the COVID-19 crisis: Quantitative and qualitative analyses of United States governors. *Journal of Applied Psychology*, 105(8), 771–783. <https://doi.org/10.1037/apl0000577>
- Shamir, B., Arthur, M. B., & House, R. J. (1994). The rhetoric of charismatic leadership: A theoretical extension, a case study, and implications for research. *The Leadership Quarterly*, 5(1), 25–42. <https://doi.org/10.1108/S1479-35712018000009016>
- Short, J. C., Broberg, J. C., Coglisier, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA). *Organizational Research Methods*, 13(2), 320–347. <https://doi.org/10.1177/1094428109335949>
- Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 415–435. <https://doi.org/10.1146/annurev-orgpsych-032117-104622>
- Starkweather, J., & Moske, A. K. (2011). Multinomial logistic regression. Retrieved January 2, 2021, from [http://www.unt.edu/rss/class/Jon/Benchmarks/MLR\\_JDS\\_Aug2011.pdf](http://www.unt.edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011.pdf).
- Tharwat, A., Gaber, T., Ibrahim, A., & Hassaniien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 30(2), 169–190. <https://doi.org/10.3233/AIC-170729>
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.
- Weber, M. (1924). *The theory of social and economic organizations*. Free Press.
- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold cross-validation over hold-out validation on colossal datasets for quality classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, (Cv), 78–83. <https://doi.org/10.1109/IACC.2016.25>.

- Yammarino, F. J., Mumford, M. D., Serban, A., & Shirreffs, K. (2013). Assassination and leadership: Traditional approaches and historiometric methods. *Leadership Quarterly*, 24(6), 822–841. <https://doi.org/10.1016/j.leaqua.2013.08.004>
- Yammarino, F. J., Sotak, K. L., & Serban, A. (2020). Charismatic, ideological, and pragmatic model with shared and collective leadership: A multi-level integration. In S. T. Hunter, & J. B. Lovelace (Eds.), *Multiple pathways to success: Extending the charismatic, ideological, and pragmatic approach to leadership* (pp. 116–142). Routledge (Taylor & Francis Group).
- Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5, 1205–1224.
- Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, 4(11), 1–7. <https://doi.org/10.21037/atm.2016.03.37>
- Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, 5(1), 44–53. <https://doi.org/10.1093/nsr/nwx106>

### Author Biographies

**Jason D. Marshall** is an incoming Assistant Professor of Management in the Heider College of Business at Creighton University. He earned his PhD in Management, with a concentration in Leadership & Organizational Science, from Binghamton University, State University of New York. With a broad interest in strategic leadership, Marshall's research largely centers on using computer-aided text analysis (CATA) and machine-learning techniques to assess psychological constructs in new ways and in new contexts. His work has been published in *The Leadership Quarterly* and *American Journal of Health Education*.

**Francis J. Yammarino** is SUNY Distinguished Professor of Management and Director, Bernard M. & Ruth R. Bass Center for Leadership Studies, at Binghamton University, State University of New York. He was a senior editor of *The Leadership Quarterly* and a co-editor of *Research in Multi-Level Issues*, has served on 10 scholarly journal editorial review boards, and is a Fellow of the SIOP, APA, APS, and SOB. His research focuses on leadership and levels of analysis issues and has been funded by several federal agencies. In 2010, Dr. Yammarino received the Distinguished Leadership Scholar Career Award from *The Leadership Quarterly* and *Academy of Management Network of Leadership Scholars*, and in 2021 he received the Sage Publications/RMD Distinguished Career Award from the *Academy of Management Research Methods Division*.

**Srikanth Parameswaran** is an Assistant Professor of Management Information Systems in the School of Management at Binghamton University. His research interests are in the areas of IT innovation, user-generated and web content, and technology-mediated health outcomes. His research has been published in *Information and Management (I&M)*, *Journal of the American Medical Informatics Association (JAMIA)*, *Information Systems Frontiers (ISF)*, and other outlets. He earned his MS in management information systems and PhD in Management Science and Systems from the State University of New York at Buffalo. He received his bachelor's degree in Computer Science and Engineering from Anna University. He was featured on the Poets & Quants list of the "Top 50 Undergraduate Business Professors of 2021."

**Minyoung Cheong** is an Assistant Professor of Management and Organization in the School of Graduate Professional Studies at Great Valley, Pennsylvania State University. He earned his PhD in Management from Binghamton University, State University of New York. His research interests relate to leadership with a primary interest in empowering leadership, organizational paradox, and levels of analysis and multilevel issues in leadership. His work has been published in *The Leadership Quarterly*, *Journal of Organizational Behavior*, and *Group & Organization Management*.