IST 736

Final Project

Submitted by:
Connor Gendron - cgendron@syr.edu
Jeremiah Graff - jgraff@syr.edu
Jean Paul Uwimana - juwimana@syr.edu
David Yamin - dyamin@syr.edu

6/9/2020

CONTENTS

I. INTRODUCTION

Dating back to the mid-19th century, the Democratic and Republican parties have held national political conventions every four years (each, a "Convention" and together, the "Conventions"). The primary purpose of the Conventions is to select the party's nominee for president. [1]

Figure 1. Barry Goldwater supporters at the 1964 Republican Convention.

In the modern era, the Conventions have lost virtually all of the drama and unpredictability they once held. This is mostly due to the fact that, beginning in the 1960s, both parties' presidential candidates have amassed the delegates necessary for first-ballot nomination through a months-long series of state primaries and caucuses. Although the parties' presidential selection is a *fait accompli* long before the Conventions are gaveled to order, most of their other traditions have remained intact.

Examples of these traditions include the opening night keynote address by a party rising star, the debate over, and formal adoption of, the party platform, and the state-by-state roll call of delegate votes.

Far and away the most significant Convention feature that continues to thrive, however, is the presidential nomination acceptance speech by the party's candidate (each, an "Acceptance Speech" and, collectively, the "Acceptance Speeches"). Traditionally delivered on the Convention's final night and nationally televised in prime time, the presidential candidate's Speech typically represents his or her best opportunity to capture the attention and imagination of the electorate. William Galston, a Senior Fellow at the Brookings Institute, has summarized the purpose of an Acceptance Speech as follows:

> Successful acceptance speeches at national party conventions accomplish two tasks: they lay out what is at stake in the forthcoming election in clear, unmistakable terms; and they do what is necessary in the circumstances of the times to build majority support, not only for the nominee personally, but also for the nominee's strategy for surmounting the most urgent and important problems of the day. Although not every presidential candidate enters the general election contest with realistic prospects of victory, each has the opportunity to play the hand they've been dealt either well or badly, and, typically, we can judge which is which.[2]

This analysis focuses on Acceptance Speeches by candidates from both parties during the period from 1900 through 2016. This subset of Acceptance Speeches captures many of the most famous and enduring Acceptance Speeches ever delivered.

---

[1] Several other U.S. political parties also hold or have held quadrennial presidential nominating conventions, but this analysis is devoted exclusively to the Conventions.

[2] Galston, W.A. (2016, July 21). What's the purpose of an acceptance speech? *Brookings*. Retrieved from: https://brook.gs/2AOdrzm.

At the outset of the 20th century, three-time Democratic presidential nominee William Jennings Bryan's acceptance speeches established him as the preeminent orator of this generation. Bryan's legendary "Cross of Gold Speech" propelled him from relative obscurity to the Democratic nomination in 1896. At age 36, he was, and remains, the youngest presidential nominee in American history.

Figure 2. William Jennings Bryan delivers Acceptance Speech at the 1900 Democratic Convention.

Although he lost the 1896 election to Republican William McKinley, Bryan's oratorical skills were so compelling that the Democrats nominated him to face McKinley again in 1900 (another loss) and yet again in 1908 against Republican William Howard Taft (a record third loss).

While he never won the ultimate political prize, Bryan's skills defined political communication in his era. Two decades before the Age of Radio, and even longer before the Age of Television, Bryan stood alone on a stage before entire Conventions and mesmerized the crowds with his pure speech making and charismatic gifts.

Figure 3. Franklin D. Roosevelt delivers Acceptance Speech at the 1932 Democratic Convention.

Three decades later, another great presidential communicator, Franklin D. Roosevelt, would use the power of radio ─ which by then was a fixture in nearly every American home ─ to bring his record four Acceptance Speeches beyond the confines of the convention halls and directly to the American electorate.

Delivering his Acceptance Speeches (and countless other presidential addresses) over the radio allowed Roosevelt to focus his listener's attention on his message as delivered in his preternaturally optimistic and buoyant patrician voice ─ and to help mask the fact that he had been confined to a wheelchair as a result of contracting polio in his early-30s.

Although, as discussed above, by the 1960s Conventions had begun to lose much of their drama and unpredictability, the relatively new medium of television was entering its prime. As **Figure 4** at right reflects, color television brought a new immediacy to the Convention experience for home viewers. Among other things, it captured the color and pageantry of the events for viewing public that has always been drawn to an American spectacle.

Figure 4. Richard Nixon concludes his Acceptance Speech at the 1968 Republican Convention.

Figure 5. Ronald Reagan delivers his Acceptance Speech at the 1980 Republican Convention.

Just as Franklin Roosevelt mastered the medium of radio in an earlier era, Ronald Reagan was the undisputed master of the now-ubiquitous and dominant television medium. Initially, the conventional wisdom about Reagan and his quest for the presidency held that his right-wing views on the Cold War, the appropriate government response to protests and demonstrations, and other issues of the day risked frightening away moderate voters. Fairly or not, Reagan's mental acuity was also an issue of concern to the public, given that he was the oldest person to win a major party presidential nomination in 1980.

Due in large part to his past career as a screen actor and his affable, easy-going demeanor on television, Reagan successfully used his Acceptance Speeches in 1980 and 1984 to reassure Americans that he was a safe and capable choice to lead the country as president during the 1980s.

Television's expansion of Acceptance Speeches' national reach brought with it a concomitant downside risk. Nationwide Convention audiences heard and remembered what candidates proclaimed and promised in their Acceptance Speeches as never before. Two candidates in particular learned this lesson the hard way.

In 1964, Republican Barry Goldwater was nominated by the Republicans. Goldwater, whose politics were more to the right than any presidential candidate up to that time, alarmed the voting public (including many Republicans) when he stated in his Acceptance Speech that "extremism in defense of liberty is no vice" and "moderation in pursuit of justice is no virtue." Voters who were already uneasy about Goldwater's potential to recklessly escalate the Vietnam war were repelled by this declaration in significant numbers.

George H.W. Bush, the Republican's 1988 nominee, learned a different lesson about making expedient campaign promises in a widely-viewed Acceptance Speech. In his speech, Bush attempted to shore up lingering concerns about the depth of his fiscal conservatism by promising voters that, if Congress decided to expand federal spending, he would respond by telling it to "Read my lips … no new taxes!" When Bush signed a bi-partisan Congressional tax increase into law during in the early-1990s, voters took note of his broken pledge. He did not win re-election in 1992.

Figure 6. Barry Goldwater's Acceptance Speech at the 1964 Republican Convention in which he infamously said: *"Extremism in defense of liberty is no vice. Moderation in pursuit of justice is no virtue."*

Figure 7. George H.W. Bush's Acceptance Speech at the 1988 Republican Convention in which he promised to tell Congress: *"Read my lips. No new taxes!"*

Finally, here in the 21st century, Acceptance Speeches briefly returned to the charismatic oratorical tradition personified by William Jennings Bryan's Acceptance Speeches a century earlier. Barack Obama first garnered national attention with his rousing keynote address at the 2004 Democratic Convention in Boston. Obama parlayed his new fame into a U.S. Senate seat and, in 2008, he rode his stirring oratorical gifts to the presidency. Due in large part to his powerful communication skills, Obama won a second term in 2012.

Figure 8. President Barack Obama delivers Acceptance Speech at the 2012 Democratic Convention.

For over 100 years, Convention Acceptance Speeches have been a key component of presidential campaign discourse and ─ despite the avalanche of technological and societal changes that have characterized the 21st century ─ their importance shows no signs of abating.

The primacy and special weight given to Acceptance Speeches inevitably begs many questions. For example, a modern presidential campaign may well ask: "What separates strong Acceptance Speeches from the less effective variety? What can a candidate do to align their Acceptance Speech with past successful Acceptance Speeches?" Likewise, those curious about the interplay between Acceptance Speeches and what they reflect about then-current social and economic events in American society at large will have a host of questions of their own.

Is it even possible to answer that question without sitting down and studying dozens of Acceptance Speeches? Even if one had the time to undertake such a reading project, how would one identify subtle patterns and nuances that, in their totality, may have much to say about the candidate's likelihood of electoral success and the extant economic key performance indicators of the day?

The instant analysis endeavors to answer these questions through the application of machine learning and text mining technology to the text of 55 Acceptance Speeches from 1900 to 2016 (the "Acceptance Speech Corpus" or the "Corpus") ─ approximately 250,000 words in all.

Figure 9. What may be learned at the intersection of campaign rhetoric and modern technology?

"Machine learning" (ML) is a term that has taken root in the popular consciousness in recent years. Very generally, ML typically involves designing and applying algorithms that enable computers to learn from past cases, identify patterns in data for classification and other purposes, and accurately predict future outcomes.

"Text mining," while not as widely understood, describes a concept that augments ML and expands its capabilities. Text mining typically involves the process of structuring a text collection or "corpus" ─ usually by parsing, vectorizing and tokenizing processes ─ so that it may be inserted into a database. Once this process is complete, ML algorithms may be applied for the purpose of deriving and evaluating patterns in the text.

This combination of digital text mining and processing tools and accessible ML packages has opened a world of possibilities ─ especially for those seeking to identified previously inscrutable connections between American political rhetoric (embodied here by the Acceptance Speeches) and broader societal conditions.

Accordingly, the analysis below applies ML predictive models to the Acceptance Speech Corpus. After mining and processing the text, multiple ML algorithms are asked to identify multiple factors concerning the candidate's party affiliation, the ultimate success of their campaign and multiple then-extant economic indicator levels.

II. ANALYSIS AND MODELS

The 55 Acceptance Speeches comprising the Corpus were copied from the American Presidency Project website and pasted into individual text files. The historical polling and economic data (collectively, the "Categorical Data") utilized in the analysis were collected from a variety of online sources as described in greater detail below.

A. About the Data

The pre-preprocessing steps applied to the Corpus and the Categorical Data is discussed below.

1. Text pre-processing steps.

The Acceptance Speech Corpus was pre-processed and prepared for analysis pursuant to the following steps:

Step One: The 55 Acceptance Speeches comprising the Corpus were copied from the American Presidency Project website and pasted into individual text files. Specific references to "Democrat(s)", "Democratic" or "Republican(s)" were removed.

Figure 10 below depicts a representative portion of one such text file:

Figure 10. Excerpt from JFK's 1960 Acceptance Speech

Governor Stevenson, Senator Johnson, Mr. Butler, Senator Symington, Senator Humphrey, Speaker Rayburn, Fellow partys, I want to express my thanks to Governor Stevenson

It was my great honor to place his name in nomination at the 1956 party Convention, and I am delighted to have his support and his counsel and his advice in the coming mon

With a deep sense of duty and high resolve, I accept your nomination.

I accept it with a full and grateful heart--without reservation--and with only one obligation--the obligation to devote every effort of body, mind and spirit to lead our Party back

I am grateful, too, that you have provided me with such an eloquent statement of our Party's platform. Pledges which are made so eloquently are made to be kept. "The Right

And I am grateful, finally, that I can rely in the coming months on so many others--on a distinguished running-mate who brings unity to our ticket and strength to our Platform,

I feel a lot safer now that they are on my side again. And I am proud of the contrast with our party competitors. For their ranks are apparently so thin that not one challenger h

Step Two: The Corpus and the Categorical Data (see below) were combined in a .xlsx file.

Step Three: The combined Corpus and Categorical Data were read into Python in the form of a Pandas data base.

Step Four: Separate and separable Python lists were then created for all fifteen Categorical Data variables.

Step Five: SKLearn's CountVectorizer function was applied to initially vectorize and tokenize the Acceptance Speech Corpus on three bases: (i) total frequency ("TF");

(ii) Boolean; and (iii) term frequency – inverse document frequency ("TF-IDF"). CountVectorizer's English stop words list was utilized.

Step Six: The vectorized/tokenized corpus was converted into a new Pandas data frame with the individual words in the corpus as column headers. **Figure 11** depicts an excerpt from the vectorized/tokenized Acceptance Speech Corpus.

Figure 11. Excerpt from vectorized/tokenized Corpus.

	aback	abandon	abandoned	abandoning	...	zip	zone	zones	zyje
3	1	1	4	0	...	0	0	0	0
43	0	0	1	0	...	1	0	0	0
54	0	0	0	0	...	0	2	0	0
51	0	0	0	0	...	0	0	0	0
48	0	0	1	0	...	0	0	0	0
14	0	0	0	0	...	0	0	0	0
47	0	2	0	0	...	0	0	1	0
53	0	0	0	0	...	0	0	0	0
17	0	0	0	0	...	0	0	0	0
37	0	0	1	0	...	0	0	0	0
4	0	0	0	0	...	0	0	0	0
19	0	0	0	0	...	0	0	0	0
18	0	0	0	0	...	0	0	0	0
42	0	0	0	0	...	0	0	0	0
21	0	1	0	0	...	0	0	0	0
30	0	0	0	0	...	0	0	0	0
29	0	0	0	0	...	0	0	0	0
9	0	0	0	0	...	0	0	0	0
50	0	1	0	1	...	0	0	0	0
35	0	0	1	0	...	0	0	0	0
22	0	0	0	0	...	0	0	0	0
44	0	0	0	0	...	0	0	0	0
34	0	0	0	0	...	0	0	0	0
28	0	0	0	0	...	0	0	0	0
13	0	0	0	0	...	0	0	0	0
26	0	0	0	0	...	0	0	0	1
46	0	0	0	0	...	0	0	2	0
38	0	1	1	0	...	0	0	0	0
23	0	0	1	0	...	0	0	0	0
6	0	1	0	0	...	0	0	0	0
20	0	0	0	0	...	0	0	0	0
12	0	1	0	0	...	0	0	0	0
7	0	0	0	0	...	0	0	0	0
52	0	1	1	0	...	0	0	0	0
36	0	0	0	0	...	0	0	0	0
49	0	0	0	0	...	0	0	1	0

2. Categorical Data pre-processing steps.

The Categorical Data was gathered and converted to categorical variables as set forth in Table 1 below:

Table 1: Categorical Variables

Variable	Categorical Conversion	Source
Candidate's political party	0 = Democrat; 1 = Republican	Am. Pres. Project
Did candidate win the election?	0 = No; 1 = Yes	Am. Pres. Project
Was candidate incumbent president?	0 = No; 1 = Yes	Am. Pres. Project
Incumbent president's party same as the candidate's?	0 = No; 1 = Yes	Am. Pres. Project
Acceptance Speech sentiment score	0= < 0; 1= ≥ 0	Daniel Loper
Unemployment rate	0= > 6%; 1= < 6%	Fraser; the Balance
GDP growth	0= < 3%; 1= > 3%	Fraser; the Balance
Inflation Rate	0 = > .1% or < 3%; 1 = <.1% or >3%	Official Data; the Balance
Real household income growth (8-year look back)	0 = negative; 1 = positive	Fed. Reserve. (St. Louis)
Prevailing income tax rate	0 = ≥ 0.1%; 1 = ≤ 0 %	Tax Policy Center
Dow Jones Industrial Average Volume (4-year look back)	0 = ≤ 29% growth; 1 = > 29% growth	Stooq
Consumer Confidence Index	0 = ≤ 100; 1 = > 100	Fed. Reserve (St. Louis)
Poll: "Are you dissatisfied with country's direction?"	0 = > 75%; 1 = ≤ 75%	Gallup
Incumbent president's approval rating	0 = < 48%; 1 = ≥ 48%	Am. Pres. Project
Consumer Price Index	0 = < 3; 1 = ≥ 3	Inflation Data

Figure 12 depicts a representative excerpt of historical Consumer Price Index data in its raw form:

Figure 12. Excerpt from raw historical Consumer Price Index data.

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	AVG	PERC CHANGE
1913	9.8	9.8	9.8	9.8	9.7	9.8	9.9	9.9	10	10	10.1	10	9.9	
1914	10	9.9	9.9	9.8	9.9	9.9	10	10.2	10.2	10.1	10.2	10.1	10	
1915	10.1	10	9.9	10	10.1	10.1	10.1	10.1	10.1	10.2	10.3	10.3	10.1	
1916	10.4	10.4	10.5	10.6	10.7	10.8	10.8	10.9	11.1	11.3	11.5	11.6	10.9	9.00
1917	11.7	12	12	12.6	12.8	13	12.8	13	13.3	13.5	13.5	13.7	12.8	23.87
1918	14	14.1	14	14.2	14.5	14.7	15.1	15.4	15.7	16	16.3	16.5	15.1	34.02
1919	16.5	16.2	16.4	16.7	16.9	16.9	17.4	17.7	17.8	18.1	18.5	18.9	17.3	33.76
1920	19.3	19.5	19.7	20.3	20.6	20.9	20.8	20.3	20	19.9	19.8	19.4	20	32.74
1921	19	18.4	18.3	18.1	17.7	17.6	17.7	17.7	17.5	17.5	17.4	17.3	17.9	2.48
1922	16.9	16.9	16.7	16.7	16.7	16.7	16.8	16.6	16.6	16.7	16.8	16.9	16.8	-8.70
1923	16.8	16.8	16.8	16.9	16.9	17	17.2	17.1	17.2	17.3	17.3	17.3	17.1	-6.22
1924	17.3	17.2	17.1	17	17	17	17.1	17	17.1	17.2	17.2	17.3	17.1	-0.97
1925	17.3	17.2	17.3	17.2	17.3	17.5	17.7	17.7	17.7	17.7	18	17.9	17.5	2.94
1926	17.9	17.9	17.8	17.9	17.8	17.7	17.5	17.4	17.5	17.6	17.7	17.7	17.7	2.71
1927	17.5	17.4	17.3	17.3	17.4	17.6	17.3	17.2	17.3	17.4	17.3	17.3	17.4	-0.19
1928	17.3	17.1	17.1	17.1	17.2	17.1	17.1	17.1	17.3	17.2	17.2	17.1	17.1	-2.47
1929	17.1	17.1	17	16.9	17	17.1	17.3	17.3	17.3	17.3	17.3	17.2	17.1	-1.72
1930	17.1	17	16.9	17	16.9	16.8	16.6	16.5	16.6	16.5	16.4	16.1	16.7	-2.91
1931	15.9	15.7	15.6	15.5	15.3	15.1	15.1	15.1	15	14.9	14.7	14.6	15.2	-10.41
1932	14.3	14.1	14	13.9	13.7	13.6	13.6	13.5	13.4	13.3	13.2	13.1	13.7	-16.12
1933	12.9	12.7	12.6	12.6	12.6	12.7	13.1	13.2	13.2	13.2	13.2	13.2	13	-14.47
1934	13.2	13.3	13.3	13.3	13.3	13.4	13.4	13.4	13.6	13.5	13.5	13.4	13.4	-4.06
1935	13.6	13.7	13.7	13.8	13.8	13.7	13.7	13.7	13.7	13.7	13.8	13.8	13.7	2.49
1936	13.8	13.8	13.7	13.7	13.7	13.8	13.9	14	14	14	14	14	13.9	3.99
1937	14.1	14.1	14.2	14.3	14.4	14.4	14.5	14.5	14.6	14.6	14.5	14.4	14.4	5.37
1938	14.2	14.1	14.1	14.2	14.1	14.1	14.1	14.1	14.1	14	14	14	14.1	0.71
1939	14	13.9	13.9	13.8	13.8	13.8	13.8	13.8	14.1	14	14	14	13.9	-1.65
1940	13.9	14	14	14	14	14.1	14	14	14	14	14	14.1	14	-0.94
1941	14.1	14.1	14.2	14.3	14.4	14.7	14.7	14.9	15.1	15.3	15.4	15.5	14.7	5.00
1942	15.7	15.8	16	16.1	16.3	16.3	16.4	16.5	16.5	16.7	16.8	16.9	16.3	14.79
1943	16.9	16.9	17.2	17.4	17.5	17.5	17.4	17.3	17.4	17.4	17.4	17.4	17.3	15.33
1944	17.4	17.4	17.4	17.5	17.5	17.6	17.7	17.7	17.7	17.7	17.7	17.8	17.6	9.32
1945	17.8	17.8	17.8	17.8	17.9	18.1	18.1	18.1	18.1	18.1	18.1	18.2	18	5.47
1946	18.2	18.1	18.3	18.4	18.5	18.7	19.8	20.2	20.4	20.8	21.3	21.5	19.5	10.59
1947	21.5	21.5	21.9	21.9	21.9	22	22.2	22.5	23	23	23.1	23.4	22.3	21.42

Finally, **Figure 13** depicts a representative excerpt of Categorical Data in its processed form:

Figure 13. Excerpt of Categorical Data in its processed form.

id	candidate	party	year	sentiment	party_num	incumbent_candidate	incumbent_party	winner	unemployment	GDP	Inflation
2	mckinley_william	r	1900	0	1	1	1	1	1	1	1
7	bryan_william	d	1900	0	0	0	0	0	1	1	1
18	parker_alton	d	1904	0	0	0	0	0	1	1	1
54	taft_william08	r	1908	1	1	0	1	1	0	1	0
4	wilson_woodrow12	r	1912	0	1	0	0	1	1	1	1
13	taft_william12	r	1912	0	1	1	1	0	1	1	0
19	hughes_charles	r	1916	1	1	0	0	0	1	1	0
20	wilson_woodrow16	d	1916	0	0	1	1	1	1	1	0
53	harding_warren	r	1920	0	1	0	0	1	1	1	0
8	hoover_herbert28	r	1928	0	1	0	1	1	1	0	0
23	smith_al	d	1928	1	0	0	0	0	1	0	0
24	hoover_herbert32	r	1932	0	1	1	1	0	0	0	0
38	roosevelt_franklin32	d	1932	0	0	0	0	1	0	0	0
40	landon_alf	r	1936	0	1	0	0	0	0	1	1
41	roosevelt_franklin36	d	1936	0	0	1	1	1	0	1	1
22	willkie_wendell	r	1940	0	1	0	0	0	0	1	0
47	roosevelt_franklin40	d	1940	0	0	1	1	1	0	1	0
25	dewey_thomas44	r	1944	1	1	0	0	0	1	1	1
33	roosevelt_franklin44	d	1944	1	0	1	1	1	1	1	1
27	truman_harry	d	1948	0	0	1	1	1	1	1	1
42	dewey_thomas48	r	1948	0	1	0	0	0	1	1	1
5	eisenhower_dwight52	r	1952	1	1	0	0	1	1	1	0
51	stevenson_adlai52	d	1952	0	0	0	1	0	1	1	0
9	eisenhower_dwight56	r	1956	0	1	1	1	1	1	0	1
36	stevenson_adlai56	d	1956	0	0	0	0	1	1	0	1
26	nixon_richard60	r	1960	0	1	0	1	0	0	0	1
29	kennedy_john	d	1960	1	0	0	0	1	0	0	1
17	johnson_lyndon	d	1964	0	0	1	1	1	1	1	1
31	goldwater_barry	r	1964	0	1	0	0	0	1	1	1
16	nixon_richard68	r	1968	0	1	0	0	1	1	1	0
32	humphrey_hubert	d	1968	0	0	0	1	0	1	1	0
21	nixon_richard72	r	1972	0	1	1	1	1	1	1	1

With both the Acceptance Speech Corpus and the Categorical Data in their final form (the "Processed Data"), the pre-processing stage of this analysis was complete.

B. Models

As set forth in greater detail below, several of Sklearn's ML algorithms ─ specifically, Multinomial Naïve Bayes ("MNB") TF, Bernoulli and TF-IDF, Support Vector Machine ("SVM") Linear Kernel, SVM RBF and SBF Poly, K-Means Clustering and Latent Derelicht Analysis probabilistic model ("LDA") ─ were applied to the pre-processed data as follows:

Step One: Both data frames were split into "training" and "test" portions using either: (i) a hold-out methodology; or (ii) a cross-validation methodology.

Step Two: Labels were removed from all component parts of the data frame.

Step Three: The algorithm was applied to "fit" or "train" the training portion of the text data.

Step Four: After the training process was complete, the algorithm's prediction function was applied to the "test" portion of the text data.

Step Five: A confusion matrix was generated.

Step Six: The process is repeated after tuning the model hyper-parameters.

III. RESULTS

Multiple ML algorithms were applied to the Processed Data. The results of each such algorithm application are summarized in this section.

A. Multinomial Naïve Bayes

The methodology utilized in the application of the MNB algorithm to the Processed Data and the results achieved thereby were as follows:

1. Methodology

In the analysis described below, three variations of the SKLearn MNB ML algorithm were applied to the Processed Data (including all categorical variables) : (i) MNB-TF; (ii) MNB-Bernoulli; and (iii) MNB-TF-IDF.

- ***K-Fold Cross Validation.*** With respect to each algorithm variation, the training and testing data subsets were developed utilizing SKLearn's K-Fold Cross Validation function. The initial k-fold setting was 3.

- ***Tuning the K-Fold Parameter.*** After an initial total accuracy value was obtained using 3 folds, the k-fold value was (i) decreased downward to 2; then (ii) increased upward until a maximal total accuracy value was achieved.

- ***Tuning the Alpha Parameter.*** The alpha parameter was set to zero while the initial k-fold parameter tuning was underway. Thereafter, the initial k-fold value that achieved the maximal accuracy value was maintained while the alpha parameter was decreased and increased until a maximal total accuracy value was achieved.

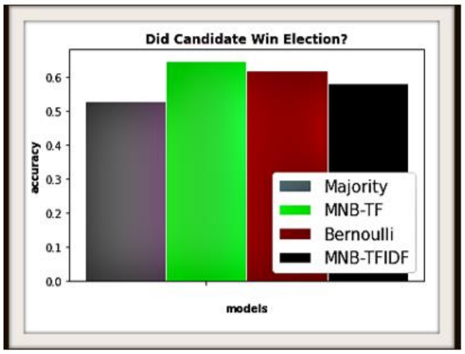
- ***Re-tuning the K-Fold Parameter.*** The k-fold parameter was again adjusted upward and downward while the alpha value that, together with the initial k-fold value, achieved the maximal total accuracy value was held constant. If such further k-fold adjustment achieved a new maximal total accuracy value, the new k-fold value replaced the initial k-fold accuracy value.

- ***Re-tuning the Alpha Parameter.*** The alpha parameter was re-tuned using a process that paralleled the k-fold re-tuning described above. In no case did a re-tuned alpha parameter emerge that yielded a new maximal total accuracy value.

- ***Confusion Matrix and other Metrics.*** With parameter tuning complete, the model was run again and a confusion matrix, precision and recall scores and an F-1 value were generated.

2. Performance result for "Did candidate win the election?" categorical variable

Figure 14. All MNB models outperform majority rule baseline for the "Did candidate win election?" variable.



Application of the three MNB algorithm variations produced a particularly noteworthy result with respect to the categorical variable identifying whether each Acceptance Speech was given by a candidate who either won or lost the general election.

In assessing the performance of a classification model, one particularly useful method involves comparing the model's results to the so-called "majority rules" baseline ("<u>Majority Rules</u>" or "<u>Majority</u>"). This baseline is calculated by calculating the total accuracy score that would result if the model identified the categorical variable that appears most frequently in the training set and used that variable in classifying every item in the test set.

With respect to the "Did candidate win election?" categorical variable, the **Majority Rules approach** would generate a **total accuracy score of .523**. As **Figure 14** reflects, all three MNB models generated total accuracy scores greater than the Majority Rules baseline.

The MNB-TF algorithm was the leading performer vis-à-vis this categorical variable. **Figures 15** and **16** below set forth the confusion matrix and related metrics quantifying the MNB-TF model's performance:

Figure 15. MNB-TF confusion matrix for the "Did candidate win election?" variable.

	Predict False	Predict True
Actual False	11	15
Actual True	4	25

Figure 16. Key metrics MNB-TF's classification results for the "Did candidate win election?" variable.

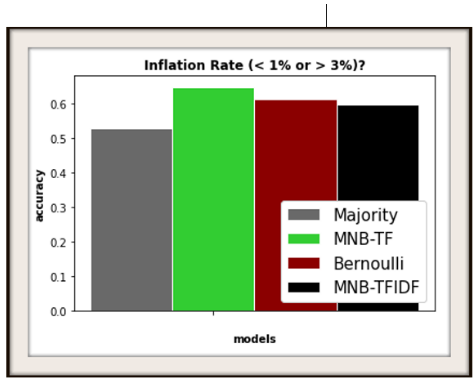
Metric	Value
MNB-TF Accuracy	**.647**
Majority Rule	**.523**
Percent Change	**+123.7%**
K-Folds	8
Alpha	.028
Precision (TP)	.625
Recall (TP)	.862
Precision (TN)	.688
Recall (TN)	.423
F-1 Score	.725

3. Performance Result for "Inflation Rate" categorical variable.

Application of the three MNB algorithm variations similarly produced a noteworthy result with respect to the categorical variable identifying whether the national rate of inflation was .1% or less or 3% or higher. While the significance of high inflation rates are generally known, an inflation rate that is too low ─ such as .1%, for example ─ are equally as undesirable. When unusually low rates occur, the phenomenon is called "deflation". Deflation was most notably prevalent in the United States and worldwide during the Great Depression in the 1930s.

With respect to the "inflation rate categorical variable, the **Majority Rules approach** would generate a **total accuracy score of .527**. As **Figure 17** reflects, all three MNB models once again generated total accuracy scores greater than the Majority Rules baseline.

Figure 17. All MNB models outperform majority rule baseline for the "Inflation Rate" variable.



Likewise, the MNB-TF algorithm was once again the leading performer vis-à-vis this categorical variable. **Figures 18** and **19** below set forth the confusion matrix and related metrics quantifying the MNB-TF model's performance:

Figure 18. MNB-TF confusion matrix for the "Inflation Rate" variable.

	Predict False	Predict True
Actual False	**8**	**18**
Actual True	**1**	**28**

Figure 19. Key metrics MNB-TF's classification results for the "Inflation Rate" variable.

Metric	Value
MNB-TF Accuracy	.647
Majority Rule	.527
Percent Change	+122.77%
K-Folds	4
Alpha	5.5
Precision (TP)	.609
Recall (TP)	.966
Precision (TN)	.889
Recall (TN)	.308
F-1 Score	.747

Summary of MNB Results

All three iterations of the MNB algorithm (TF, Bernoulli and TF-IDF) performed well in classification of the categorical variables. Two results, in particular, stood out:

- ***"Did candidate win election?" variable.*** All MNB models out-performed the Majority Rules baseline of .523. The MNB-TF model achieved the highest total accuracy (.647).
- ***"Inflation Rate" variable.*** All MNB models out-performed the Majority Rules baseline of .527. Again, the MNB-TF model achieved the highest total accuracy (.647).

B. SVM

The methodology utilized in the application of the SVM algorithm to the Processed Data and the results achieved thereby were as follows:

1. Methodology

In the analysis described below, three variations of the SKLearn SVM algorithm were applied to the Processed Data (including all categorical variables) : (i) SVM-DF; (ii) SVM-Boolean; and (iii) SVM-TF-IDF.

- ***SVM Kernels*** (**Table 2** below sets forth the associated costs for each kernel):
  - **Linear Kernel.** With respect to each algorithm variation, nine predictive algorithms were developed utilizing SKLearn's Linear SVC Kernel with three different costs for each of the three variations (i.e., three Linear kernels x three variations each, as set forth above).

- **Sigmoid Kernel.** Three predictive algorithm models were built for each variable using the Sigmoid Kernel on the Boolean vectorized data frame. Each of the different models was run using a different cost.

- **Polynomial Kernel.** Three predictive algorithm models were built for each variable using the Polynomial Kernel on the Boolean vectorized data frame. Each of the different models was run using a different cost, as seen below.

Table 2. Costs associated with each SVM kernel.

SVM Kernel	Data Frame Vectorization Method	Cost
Linear	Term/Document Frequency	0.01
Linear	Term/Document Frequency	1
Linear	Term/Document Frequency	100
Linear	Boolean	100
Linear	Boolean	1
Linear	Boolean	0.01
Linear	TF-IDF	0.001
Linear	TF-IDF	1
Linear	TF-IDF	100
Sigmoid	Boolean	10
Sigmoid	Boolean	1000
Sigmoid	Boolean	0.001
Polynomial	Boolean	1000
Polynomial	Boolean	10
Polynomial	Boolean	0.001

- ***Variables Tested.*** All fifteen variables were run against the fifteen different Kernel-Cost models described above for a total of 225 different SVM models.

- ***Confusion Matrix and other Metrics.*** With parameter tuning complete, the model was run again and a confusion matrix, precision and recall scores and an F-1 value were generated.

- ***Compilation of overall performance of the models.*** All fifteen variables were ranked with a score that measured the average performance of all fifteen models run on each of the variables. Additionally, the top twenty highest performing were identified.

2. Average Model Prediction Performance Across the 15 different variables

Figure 20. Average SVM accuracy performance across all variables.

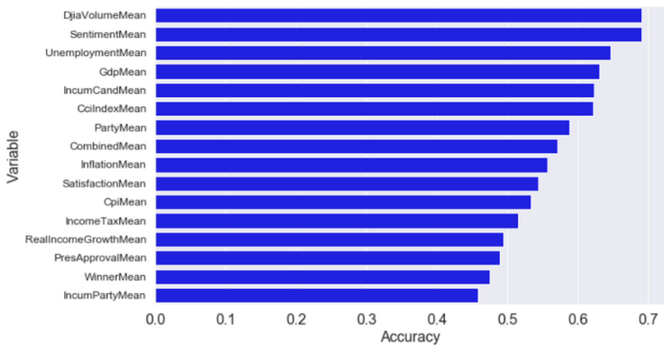


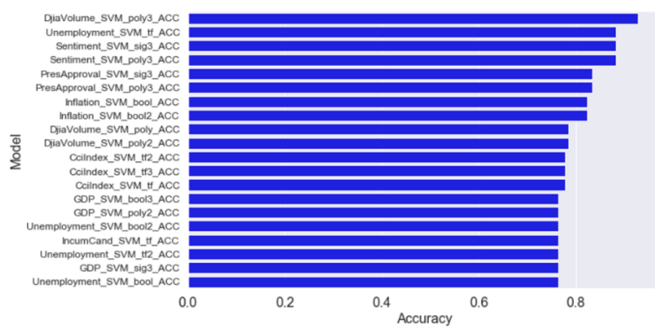
Figure 20 at left indicates that the two variables with the greatest predictive capabilities were: (i) the Dow Jones Industrial Average Volume; and (ii) Sentiment. These two variables produced an average 69% accuracy across all fifteen models run on each of them. The average accuracy across all 225 models run was 57% as seen in the bar for "Combined Mean" at left.

What this means, practically, is that the text from the nominees' Acceptance Speeches can be used to determine whether the Dow Jones Volume in the year the speech was given represented less than or equal to 29% growth from the volume recorded four years earlier or whether such growth was greater than 29% over the same period. *See* **Table 1** above. Likewise, on average, the models classified the overall Sentiment of the Acceptance Speeches based upon patterns in the candidates' word usage with the same high degree of accuracy (69%).

3. Most accurate prediction models across the 15 different variables

Figure 21 at left ranks the top 20 highest performing individual SVM models of the 225 that were run on the data set. The lowest of these 20 models performed at 76.5% accuracy, with the highest coming in at 94% accuracy. The seven models that came in at 76.5% accuracy still performed 19.5 percentage points better than the overall model average of 57%, while the top model came in with an accuracy that was 37 percentage points higher.

Figure 21. Top 20 highest performing individual SVM models.



The top performing model overall was from the polynomial kernel on the DJIA Volume. This is unsurprising insofar as the DJIA Volume variable was tied with the Sentiment variable for the highest overall prediction accuracy when averaged across all 15 models. Similarly, the Sentiment variable had the third and fourth highest performing models of the group of 225 models.

Summary of SVM Results

Of the 225 models that were run, two overall results, in particular, stood out:

- ***DJIA Volume and Sentiment are top performing variables.*** Models classifying these variables had an average score (across all 15 model permutations) of 69% accuracy. The average accuracy of the combined 225 models was 57%, so these two variables scored twelve percentage points higher than the overall average.

- ***DJIA Volume is the top model-producing variable.*** The DJIA Volume variable produced a model with 94% accuracy when run against the polynomial kernel with a Boolean vectorized data frame and a cost of 0.001. This means that, using this model, the Acceptance Speeches can be used to predict whether the Dow Jones Volume in the year the speech was given represented less than or equal to 29% growth from the volume recorded four years earlier or whether such growth was greater than 29% over the same period with 94% accuracy.

C. Clustering

The methodology utilized in the application of K-means and hierarchical clustering to the Processed Data and the results achieved thereby were as follows:

1. Methodology

Two distinct forms of clustering were run against the data to explore any unseen relationships, and whether the clusters align with the any of the pre-existing labels. The two clustering techniques used were SKlearn's k-means and scipy's hierarchy. Three different vectors were made for clustering: Boolean, Term Frequency, and TFIDF normalized. The different vectors did not appear to impact clusters, however.

- ***Tuning K-means Clusters:*** Originally the plan was to generate two clusters that could be compared against the binary labels that were derived (i.e. win or lose).

- ***Tuning K-means n_int:*** The n_init parameter determines the number of different centroid seeds that the K-means algorithm will run through, this was originally left at its default level of 10.

- ***Other K-means parameters:*** The K-means model was run using the 'auto' algorithm, a random state of 0 and verbose set to false.

- ***Tuning Hierarchical Linkage:*** The 'average' linkage was the first setting chosen for the hierarchical model. This setting uses the average distance between points to generate clusters.

- ***Re-tuning K-means Clusters:*** It was found that three clusters provided more illuminating groupings of presidential speeches.

- ***Re-tuning K-means n_init:*** Changing the n_init value from 10 to 20 also improved the clusters.

- ***Re-tuning Hierarchical Linkage:*** The clusters generated by changing the linkage from 'average' to 'complete' more closely resembled the k-means clusters. 'Complete' linkage determines clusters by the maximum distance between points.

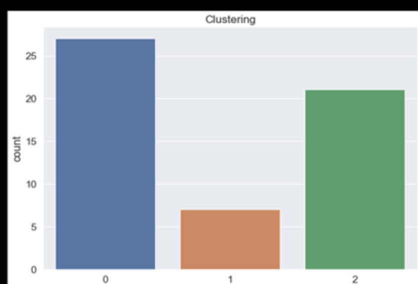
- ***Confusion Matrices:*** Confusion matrices were developed to show how the K-means clusters compared to the pre-existing labels, but did not reveal useful information. Instead the decision was made to focus on presenting the clusters as is and searching for intuitive reasons behind the clustering.

- ***Hierarchical Dendrogram:*** A dendrogram was developed to help visualize the decision-making process behind the hierarchical clusters.

2. <u>What did k-means clustering reveal</u>?

Figure 22. K-means clustering results.

	cluster	candidate
0	0	mckinley_william
2	0	parker_alton
6	0	hughes_charles
7	0	wilson_woodrow16
11	0	hoover_herbert32
12	0	roosevelt_franklin32
13	0	landon_alf
14	0	roosevelt_franklin36
15	0	willkie_wendell
16	0	roosevelt_franklin40
17	0	dewey_thomas44
18	0	roosevelt_franklin44
19	0	truman_harry
20	0	dewey_thomas48
21	0	eisenhower_dwight52
22	0	stevenson_adlai52
23	0	eisenhower_dwight56
24	0	stevenson_adlai56
26	0	kennedy_john
27	0	johnson_lyndon
28	0	goldwater_barry
30	0	humphrey_hubert
32	0	mcgovern_george
33	0	carter_jimmy76
34	0	ford_gerald
35	0	reagan_ronald80
37	0	mondale_walter



	cluster	candidate
1	1	bryan_william
3	1	taft_william08
4	1	wilson_woodrow12
5	1	taft_william12
8	1	harding_warren
9	1	hoover_herbert28
10	1	smith_al

	cluster	candidate
25	2	nixon_richard60
29	2	nixon_richard68
31	2	nixon_richard72
36	2	carter_jimmy80
38	2	reagan_ronald84
39	2	bush_georgehw88
40	2	dukakis_michael
41	2	clinton_bill92
42	2	bush_georgehw92
43	2	dole_robert
44	2	clinton_bill96
45	2	gore_al
46	2	bush_georgew00
47	2	kerry_john
48	2	bush_georgew04
49	2	mccain_john
50	2	obama_barack08
51	2	romney_mitt
52	2	obama_barack12
53	2	clinton_hillary
54	2	trump_donald

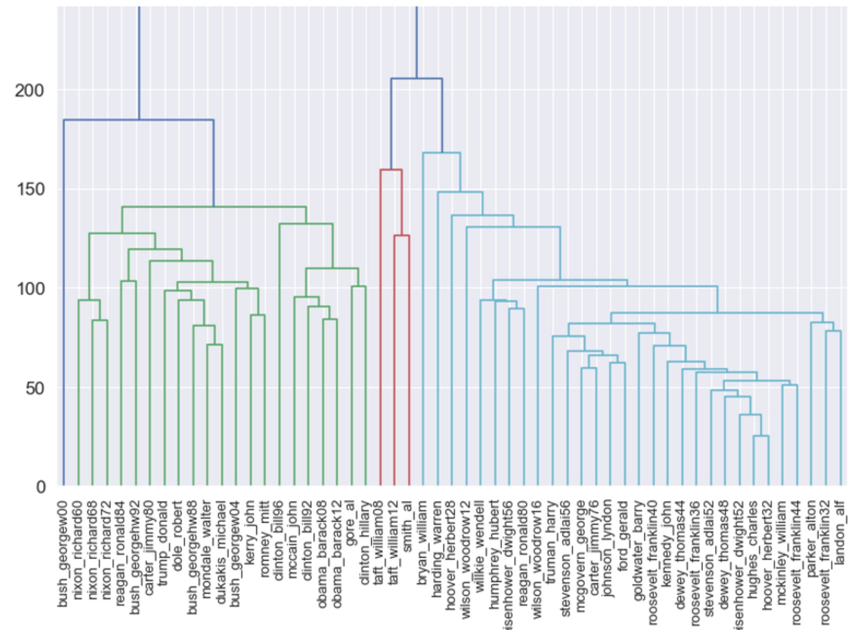
The three clusters that the final k-means parameters generated are set forth in **Figure 22** at left. The primary takeaways are that speech clustering is more dependent on temporal factors than on political affiliation, winners and losers, and economic status. Intuitively this makes sense, over the span of 116 years speech and language have evolved in this country. With this in mind, the different clusters can be viewed as indicative of linguistic changes over time. In addition to the linguistic choices, another confounding factor is that the most important political problems also change over time. Which side of the debate any given speaker is on seems to be outweighed by the presence or absence of that debate in the first place.

With these factors in mind the clusters can be viewed in terms of political eras. When applied to the three distinct clusters generated, this suggests that 1900-1930 constitutes one era, possibly the end of a longer era that can't be explored for lack of data. The next era seems to begin with FDR in the 1930s and continues on until around 1980, followed the modern era of politics from 1980 through the present.

3. Hierarchical dendrogram

Most of the hierarchical results reinforce the same concepts as the k-means cluster. One difference worth exploring is that the dendrogram actually generates four clusters, instead of the three generated by k-means. This fourth cluster contains only one speech, George W. Bush's Acceptance Speech during the 2000 campaign.

Figure 23. Hierarchical clustering



This is worth some examination – while other presidents have managed to win the electoral college without the popular vote, George W. Bush had the smallest margin of victory in the electoral college, only five electoral votes. In addition, George W. has been known for malapropisms and his 'down to earth' style of speech which may have stood out amongst the rest of the speeches.

Finally, it is worth mentioning that William Taft and Al Smith were grouped together in both clustering methods. These clusters were also by far the smallest, suggesting that the pair are particularly unique. However, when this is taken into the context of political eras, it is worth noting that their era has missing speeches and likely spans further back in time than 1900.

Summary of Clustering Results

The following clustering results were particularly noteworthy:

- ***K-means clustering.*** Acceptance Speech clustering was more dependent on temporal factors than on political affiliation, winners and losers, and economic status.

- ***Hierarchical dendrogram.*** Most of the hierarchical results reinforced the same concepts as the k-means cluster. One difference worth exploring is that the dendrogram actually generates four clusters, instead of the three generated by k-means. This fourth cluster contains only one speech, George W. Bush's Acceptance Speech during the 2000 campaign.

D. Latent Dirichlet Allocation (LDA) Topic Modeling

The methodology utilized in the application of the LDA algorithm to the Processed Data and the results achieved thereby were as follows:

1. Methodology

In the analysis described below, one model of the Sklearn's Latent Dirichlet Allocation was applied to the Preprocessed Data (text corpus only) to identify the most common topics that were discussed by the presidential nominees. Given the way LDA algorithm works which is based on raw word frequency, only one vector of non-normalized frequency count was utilized with LDA taking the following parameters:

- ***N_component***. After several trial and errors, a value of 4 seemed to produce the most realistic topics.

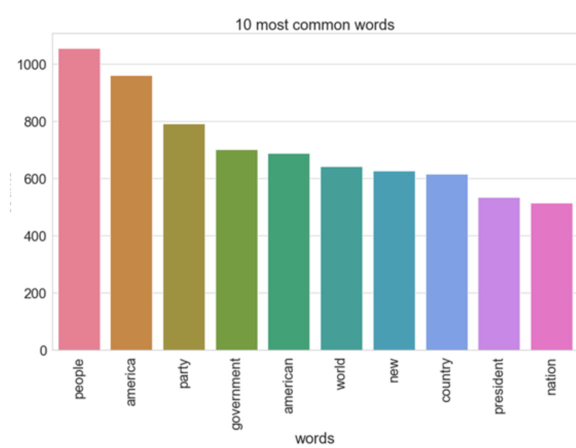
- ***Max_iter***. Several different values ranging from 200 to 600 were tested. There was no significant difference observed, therefore the parameter was set to 300.

- ***Learning_method***. This parameter was set to ***online*** and no other value was tested.

2. Most frequent words

Application of the CountVectorizer algorithm to the entire Acceptance Speech Corpus produced the following results whereby the most frequent words are ranked by their frequency counts. **Figure 24** below shows the top 10 most frequent words from the Corpus. The words ***people*** and ***America*** were used over 1000 and 900 times, respectively.

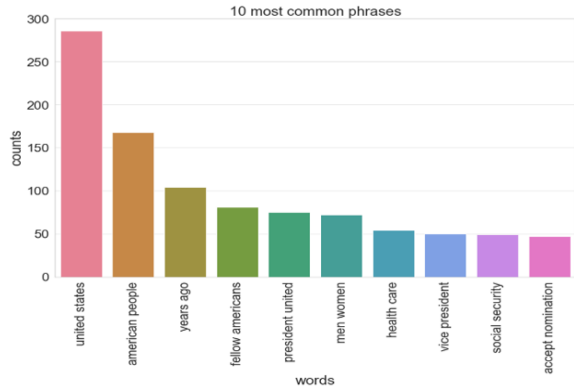
Figure 24. Top 10 words by frequency.



3. Most frequent phrases

Similarly, the application of CountVectorizer algorithm to the entire text corpus produced the following bigram results ranked by their frequency count in **Figure 25** at right.

Figure 25. Top 10 bigrams by frequency.

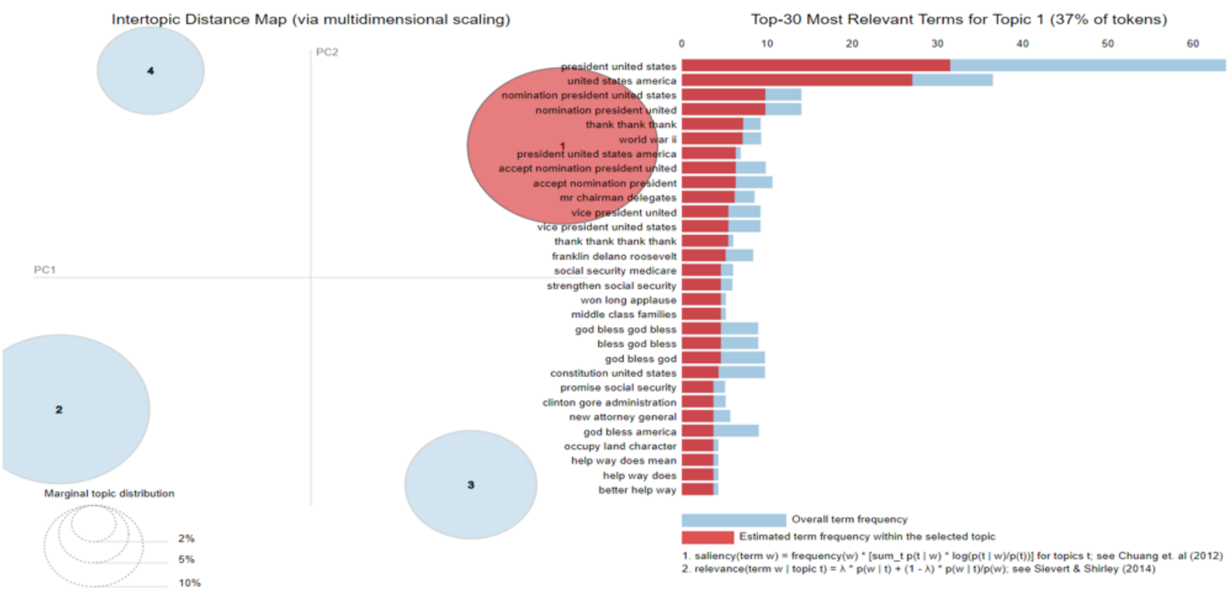


The phrases ***United States***, ***American people***, ***health care*** and ***social security*** appeared more than 250, 150, 50, and 50 times, respectively. ***Health care*** and ***social security*** are particularly interesting because the entire text corpus was composed of 55 speeches, therefore having both phrases being repeated over 50 times each suggests they were used in the vast majority of Acceptance Speeches.

4. Topic identification

Application of the LDA algorithm using unigrams (single words) was not particularly useful for identifying topics underlying the Corpus as the words overlapped across different topics produced by LDA. After several iterations of unigram modeling, it made sense to use to use multi-word grams. When bigrams and trigrams were used, topics began to be revealed. **Figure 26** below shows a graphical representation of all four topics that were identified from the Corpus with topic 1 (shaded in red) shown with the 30 most relevant terms/phrases in that topic.

Figure 26. Topic identification.



As mentioned in the preceding section, single words were not very indicative of discrete topics, hence the adoption of trigrams. In **Figure 27** at right, the top 30 words from each of the four topics are presented. And it can be seen that the topics discussed in the Corpus were ***economy***, ***heath care/social security***, ***tax***, and ***trade***. Some of the trigrams indicating the topics of discussion include ***millions new jobs***, ***affordable health care,*** ***strengthen social security***, for the first three topics, respectively. With respect to ***trade,*** the word ***trade*** appeared frequently, however, there were no prominent bigrams or trigrams associated with it.

Figure 27. Top 30 phrases in each topic.

Topic #0:

president united states, united states america, nomination president united states, nomination president united, thank thank thank, world war ii, president united states america, accept nomination president united, accept nomination president, mr chairman delegates, vice president united, vice president united states, thank thank thank thank, franklin delano roosevelt, social security medicare, strengthen social security, god bless god bless, bless god bless, won long applause, middle class families, god bless god, constitution united states, promise social security, clinton gore administration, new attorney general, god bless america, bless god bless america, occupy land character, help way does mean, better help way

Topic #1:

president united states, new cove america, presidency united states, let cause child, constitution united states, nomination president united, nomination president united states, leadership party leadership, president harry truman, united states america, small business man, millions new jobs, accept nomination accept, confidence american people, rights life liberty pursuit, rights life liberty, executive branch government, today years ago, god bless america, fi cial stability, office years ago, fi cial power, created equal endowed, bless god bless, god bless god bless, new ideas new, bless god bless america, affordable health care, god bless god, thank god bless

Topic #2:

president united states, bridge st century, people united states, build bridge st, build bridge st century, anti trust law, government united states, want build bridge st, want build bridge, audience boo president, interstate commerce commission, start decade want, years years president, audience years years, years years years president, audience years years years, years years years, mr vice president, united states america, latin american countries, state new york, think american people, candidate vice president, million new jobs, tax increase history, let look record, believe american people, right track st, right track st century, track st century

Topic #3:

president united states, says say yes, united states america, opponent says say yes, opponent says say, party party future, going let away, presidency united states, hit em harder harder, em harder harder, hit em harder, em hit em harder, em hit em, hit em hit em, hit em hit, viva bush viva, bush viva bush, viva bush viva bush, candidate conservative values, party future party, party party future party, years ago stood, vice president united, vice president united states, million new jobs, god bless god, accept nomination president, governor clinton congress, audience members boo president, audience members boo

Three of the four topics generated by the LDA algorithm could only be identified using trigrams, while the fourth topic was identified using unigram frequency words. It was nearly impossible to tell what the topic would be by just using unigrams for all the topics as there were no generated unique words that could help identify what a particular topic of conversation was across all four topics. Individual words overlapped from topic to topic, and lack of rare or outstanding words was very remarkable. This could be attributed to the fact that all nomination speeches are fairly similar in structure and are all geared toward one goal which is to set out a party narrative that voters can relate to and to thank the party for the nomination.

<u>Summary of LDA Topic Modeling Results</u>

The primary takeaway from this analysis is that while LDA is powerful, the nature of the Corpus makes it very difficult to draw topics from text as most nominees may touch on different topics in the same sentence and different keywords and phrases may be used over and over again as the nominee attempts to emphasize the focus of the issue being talked about.

As a result, it becomes harder to pinpoint specific unigrams and multi-grams that would otherwise help identify topics as opposed to other types of text collections such as news articles where a particular paragraph may be exclusively geared toward a specific.

IV. CONCLUSION

Overall, the application of Text Mining and ML methods to the Acceptance Speeches may be considered a successful endeavor, particularly insofar as classification modeling results were concerned. Clustering and LDA topic modeling results were also enlightening and generally underscored the reality that political language ─ just like the everyday language used by the larger society itself ─ evolves and changes with the passage of time.

While these results are interesting and thought provoking as an academic matter, they may potentially be put to pragmatic uses as well.

As illustrated by the Multinomial Naïve Bayes modeling described above, a moderately strong positive correlation between Acceptance Speeches and presidential winners and losers. The MNB-Total Frequency algorithm classified Acceptance Speeches with election outcomes with a **total accuracy score** of **.647** ─ this represented a **123.7% improvement** over the **Majority Rules baseline** of **.523**.

These results strongly suggest that, with additional model improvement, presidential candidates could use ML learning to craft their Convention Acceptance Speech to be in alignment with the patterns and language usage that have historical characterized Acceptance Speeches delivered by past winning candidates.

Figure 24. Application of Text Mining and ML methods to the 2016 U.S. presidential campaign has drawn substantial criticism.

Of course, it is possible that the 2020 presidential campaigns are already ─ secretly ─ doing this for their candidates. The 2016 presidential campaign was the first in which ML learning models were extensively utilized by candidates. These methods subsequently received considerable scrutiny and criticism insofar as they involved tailoring the candidates messaging based on extensive mining of prospective voters' personal information from their social media accounts. One wonders whether such tactics could be combined with the prediction capabilities of the models presented in this analysis to create a synergy that would produce even more powerfully results.

Finally, the SVM models described herein had even greater success classifying shifts in the Dow Jones' volume and other key economic indicators based on the Acceptance Speech Corpus. For example, the SVM polynomial kernel classified **four-year Dow Jones Industrial Average growth** with an impressive **94% accuracy**.

Given that success, the logical next step for Acceptance Speech classification modeling would seem to be experiments design to test the predictive accuracy of the Acceptance Speech Corpus vis-à-vis ***future*** Dow Jones and other economic indicator performance.

Indeed, the combination of Text Mining and ML with stock market performance already has a fairly robust history. One commentator surveyed the literature and summarized developments along these lines as follows:

> Already in 1998, Wuthrich et al. attempted to predict stock markets based on online news articles such as The Wall Street Journal. The idea was

straightforward: count occurrences of manually defined keywords in articles and correlate their presence with the stock values using machine learning techniques. Despite a low accuracy, this idea spawned interest and a large number of approaches were attempted to tackle the problem: genetic algorithms (Thomas and Sycara, 2000), naive Bayes (Lavrenko et al. 2000), support vector machine (Mittermayer et al. 2004), etc.

The core idea is always the same: first retrieve relevant documents, then correlate the documents content with the stock prices. While first approaches focused on financial news (Schumaker and Chen 2009), some authors also investigated financial reports (Loughran and McDonald, 2011). An important trend started when researchers considered documents directly produced by users on financial forums rather than expert journalists such as Antweiler and Frank, 2004.

Parallel to these works, the natural language processing community was interested in extracting sentiments from text with the seminal papers from Pang et al. 2002; Turney 2002. Instead of directly correlating the text contents to stock prices, researchers then correlated with success the sentiment (positive or negative) of financial forums posts to the stock prices (Das and Chen, 2007).[3]

Accordingly, shifting the focus to future predictions ─ including some form of stock price prediction ─ presents an exciting new avenue of exploration for this unlikely combination of American political tradition in the form of Acceptances Speeches and 21st century Text Mining and ML technology.

C.G.
J.G.
J.P.U.
D.Y.

[3] Denis, A. (n.d.). Twenty years of research in stock market prediction from text mining. *Societe Generale Tech Magazine*. Retrieved from: https://bit.ly/3h6teKr