ANALYSIS AND USAGE OF NATURAL LANGUAGE FEATURES IN SUCCESS PREDICTION OF LEGISLATIVE TESTIMONIES

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Marine Cossoul

March 2023

COMMITTEE MEMBERSHIP

TITLE: Analysis and Usage of Natural Language Features in Success Prediction of Legislative Testimonies

AUTHOR: Marine Cossoul

DATE SUBMITTED: March 2023

COMMITTEE CHAIR: Foaad Khosmood, Ph.D.
Professor
Computer Science Department

COMMITTEE MEMBER: Alex Dekhtyar, Ph.D.
Professor
Computer Science Department

COMMITTEE MEMBER: Cameron Jones, Ph.D
Professor
History Department

ABSTRACT

Analysis and Usage of Natural Language Features in Success Prediction of Legislative Testimonies

Marine Cossoul

Committee meetings are a fundamental part of the legislative process in which constituents, lobbyists, and legislators alike can speak on proposed bills at the local and state level. Oftentimes, unspoken "rules" or standards are at play in political processes that can influence the trajectory of a bill, leaving constituents without a political background at an inherent disadvantage when engaging with the legislative process. The work done in this thesis aims to explore the extent to which the language and phraseology of a general public testimony can influence a vote, and examine how this information can be used to promote civic engagement.

The Digital Democracy database contains digital records for over 40,000 real testimonies by non-legislator public persons presented at California Legislature committee meetings 2015-2018, along with the speakers' desired vote outcome and individual legislator votes in that discussion. With this data, we conduct a linguistic analysis that is then leveraged by the Constituent phraseology Analysis Tool (CPAT) to generate a user-based intelligent statistical comparison between a proposed testimony and language patterns that have previously been successful.

The following questions are at the core of this research: Which (if any) language features are correlated with persuasive success in a legislative context? Does the committee's topic of discussion impact the language features that can lend to a testimony's success? Can mirroring a legislator's speech patterns change the probability of the vote going your way? How can this information be used to level the playing field for constituents who want their voices heard?

Given the 33 linguistic features developed in this research, supervised classification models were able to predict testimonial success with up to 85.1% accuracy, indicating that the new features had a significant impact on the prediction of success. Adding these features to the 16 baseline linguistic features developed in Gundala's [18] research improved the prediction accuracy by up to 2.6%. We also found that balancing the dataset of testimonies drastically impacted the prediction performance metrics, with 93% accuracy achieved for the imbalanced dataset and 60% accuracy after balancing. The Constituent Phraseology Analysis Tool showed promise in the generation of linguistic analysis based on previously successful language patterns, but requires further development before achieving true usability. Additionally, predicting success based on linguistic similarity to a legislator on the committee produced contradictory results. Experiments yielded a 4% increase in predictive accuracy when adding comparative language features to the feature set, but further experimentation with weight distributions revealed only marginal impacts from comparative features.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Foaad Khosmood, for your invaluable guidance and tireless patience. I would also like to thank my committee members, Alex Dekhtyar and Cameron Jones, as well as Christine Robertson for your valuable time and feedback.

Thank you to the Cal Poly Computer Science department for providing me with the knowledge, tools, and resources to complete this research, and to the Institute for Advanced Technology and Public Policy and the Digital Democracy team for providing the necessary data.

And finally, my endless gratitude to my mom, dad, family, and friends for your tireless support; the completion of this research would not have been possible without you. Mille mercis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The purpose of this research is to further our understanding of the role of general public phraseology in legislative contexts. As we know, many factors are at play in the political sphere; relationships between legislators, corporate lobbyist interests, and hidden biases are only some of the veiled influences that can shape the outcome of legislative votes. Because of the sheer volume of potential factors that influence these votes, we cannot say with certainty that constituent phraseology alone can sway a committee's vote. However, constituents do have a part to play in civic processes that govern them, and this research aims to uncover some of the linguistic strategies that could help communicate more effectively.

The work done in this thesis builds off of Sanjana Gundala's Legislative Language of Success [18] which used Natural Language Processing and Machine Learning techniques to analyze testimonies from the Digital Democracy Database [19]. The process of determining features linked to persuasive success involved data filtration, feature extraction, implementation of classification models, and feature analysis. A set of 16 natural language features was extracted from the set of testimonies and used by a collection of classification models to predict the success of each testimony. A testimony is considered successful if the speaker's intended bill outcome matches the true bill outcome. The added value of this research can be quantified in several ways.

First, the set of natural language features was extended from 16 to 49. New features included named entity counts, ambiguity scores, quantifiers for passive and active language, sentiment analysis, and more. The full set of added features along with

their descriptions can be referenced in Table 4.1 and are discussed in more detail in Section 4.2.1. The persuasiveness of a testimony is measured using machine learning models to predict whether the result of a committee's vote matches the speaker's desired outcome. If the speaker is arguing against the passing of a bill and the bill is ultimately not passed, the testimony is considered successful. The expansion of the feature set provides a broader understanding of linguistic techniques that contribute to a testimony's persuasiveness.

Second, the context surrounding the presentation of a testimony was explored further. This part of the work stems from the hypothesis that the persuasive strength of different language strategies can be dependent on the context of the speech. The topic of discussion (i.e. education vs natural resources) may influence which strategic linguistic decisions can contribute to success. For example, a discussion on gun control may be more persuasive when using more facts-based language, while a discussion on early education could be more easily influenced with emotional arguments. Additionally, the target audience (in this case, legislators) can influence which language will me more or less successful. It could be argued that a legislator would feel a stronger connection to an argument that is presented using language that mirrors their own speech patterns.

Finally, this work looks into the usability of this data analysis. Constituents could be advantaged by providing a look into how their language compares to what has been successful in the past. To this end, the statistical analysis of the speech patterns correlated with persuasive success was aggregated and presented in a user-readable interface. This interface is not yet consumer-ready, but could be a start at providing additional information to aid the general public in drafting of testimonies.

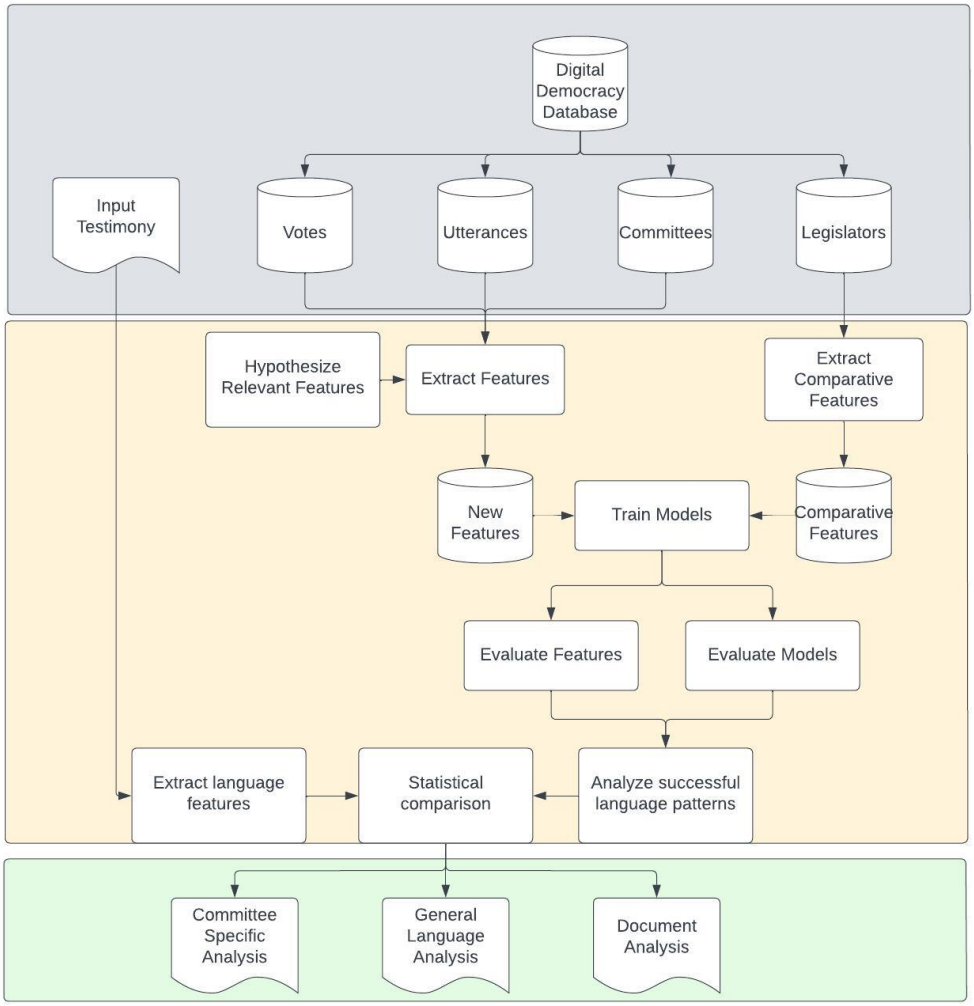

Figure 1.1: System Design

Chapter 2

BACKGROUND

2.1 The Structure of Committee Meetings

In order to accurately analyze the testimonies presented in legislative committee meetings and their resulting outcomes, we must understand the manner in which these meetings are structured. As the name suggests, each committee meeting is held by a unique committee, which discusses the particularities of the bill and its potential impacts. Public Health, Higher Education, and Transportation and Housing are all examples of separate committees at which a bill may be discussed. The committee members will then vote to pass the bill - with our without amendment - or to table it, essentially preventing it from moving forward [1]. This process prepares the bills to be read and discussed by state's the senate or assembly members. Throughout a bill's life cycle, it can be discussed at many different committees (i.e. Budget, Appropriations, Environment, Health, etc). This process is illustrated in Figure 2.1, the National Conference of State Legislature (NCSL)'s government infographic on the Legislative Process [25].

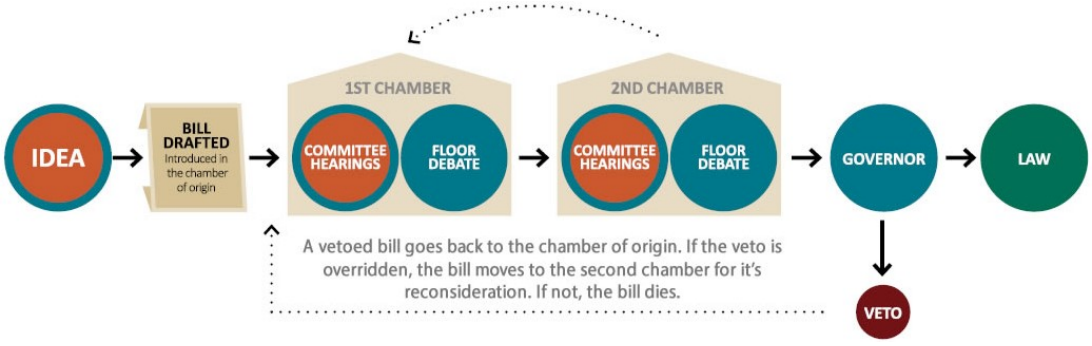

Figure 2.1: Legislative Process Diagram [25]

2.2 Digital Democracy

Digital Democracy is an online platform that provides the general public with access to past legislative committee hearings. Historically, committee meetings were only recorded through video or audio format, making it extremely difficult to conduct research and educate oneself on democratic proceedings. The resulting lack of a searchable database created a significant barrier in civic engagement. With Digital Democracy, each meeting is transcribed and labeled with the bill, topic, speaker, date, and much more. This added transparency and accessibility allows constituents to inform themselves on the issues at hand, speak up on important legislation, and keep their legislators accountable for their decisions [19],

One important characteristic of Digital Democracy database [19] that is important to consider is the imbalance of the data. The dataset includes twice as many successful testimonies as unsuccessful testimonies, and two times more testimonies arguing *for* the passing of a bill than against. The complete breakdown of the imbalanced dataset can be found in Table 5.1. This imbalance is taken into consideration through the structuring and setup of this research.

2.3 Datasets

The Digital Democracy database is an extensive, accessible, and easily searchable database with the goal of providing transparency into the legislative process [19]. To reduce the back-end overhead involved in parsing through the complex and extensive database, a flattened set of relevant data was extracted for the purposes of this research. To reduce the overall size of the data and resulting execution time,

the extracted data was organized into six datasets - Utterances, Committee Votes, Committee Details, Legislator Votes, and Legislator Utterances.

2.3.1 Utterances Data

The *Utterances* dataset is the main dataset containing information about the general public speaker, their testimony, their alignment, and more. The Digital Democracy Database stores testimonies as collections of utterances, or about 30 second chunks of speech, which make up the body of natural language that is analyzed in this research. Each row of the Utterances dataset represents a collection of information about single utterance, including the speaker's name and alignment, the date of the utterance, and identification numbers for the discussion, hearing, and bill. The complete list of fields and their descriptions can be found in Table 2.1.

2.3.1.1 Grouping Utterances

Because this research aims to analyze the structure and linguistic choices of successful testimonies as a whole, the utterances are combined into full speeches. This is done by grouping rows with matching hearing, discussion, bill, video, and committee IDs, speaker names, and dates. The utterances from these rows are concatenated to form a single speech. To access the Vote counts for a given bill discussion, the Votes table is simply referenced with the corresponding Discussion ID. As a result, any discussion in the *Utterances* dataset that does not have a corresponding Discussion ID in the *Votes* dataset is dropped. The Votes dataset is further discussed below.

Table 2.1: Complete List of Utterances Fields and Descriptions

Field Name	Description
Utterance ID	Unique ID for the utterance
First Name	First Name of Speaker
Last Name	Last Name of Speaker
Person ID	Unique ID for each person
Person Type	Type of Person Speaking (Lobbyist, Legislative Member, General Public, etc.)
Hearing ID	Unique ID for each hearing
Bill ID	Unique ID for each bill
Discussion ID	Unique ID for each bill discussed at a specific date and committee
Committee ID	Unique ID for the committee
Speech Type	Purpose of the speech
Alignment	For, Against, Neutral
Video ID	Unique ID for each video
Date	Date of the Testimony
Session Year	Year of the Legal Session
Text	Utterance within Testimony

2.3.2 Votes Data

The *Votes* dataset acts as a lookup table to reference the committee-wide outcome of a vote. The number of ayes, noes, and abstains are represented here as well as any motions that have been filed. The fields and their descriptions can be referenced in Table 2.2. This dataset was primarily utilized to determine the outcome of each testimony. A higher percentage of ayes than noes illustrates that the bill was ultimately passed through the committee. If the speaker's alignment matches the result of the vote, the testimony's outcome is considered successful.

2.3.2.1 Filtering Votes

The *Votes* dataset was cleaned in a couple ways. The first was to remove any instances of unanimous votes (remove rows where the number of Ayes or Noes was reported to

Table 2.2: Complete List of Votes Fields and Descriptions

Field Name	Description
Discussion ID	Unique ID for each bill discussed at a specific date and committee
Bill ID	Unique ID for each bill
Hearing ID	Unique ID for each hearing
Vote ID	Unique ID for each Vote (helps distinguish votes if multiple have occurred in one discussion)
Ayes	Number of votes in favor of the bill in question
Noes	Number of votes against the bill in question
Abstains	Number of votes abstaining from the bill in question
Motion ID	Unique ID of the motion
Motion Text	Actual motion to be voted on

be zero). A unanimous vote can indicate a strong decisiveness in the voting members, which can mean that not much persuasion was necessary to achieve a vote. Because the goal of this research is to identify methods of persuasion, testimonies in discussions with ultimately unanimous votes are less likely to have had an impact on the outcome and are therefore removed from the analysis.

The next step in cleaning the dataset was to look at cases in which multiple votes were conducted within a single discussion. To establish the ultimate outcome of the committee's decision, only the final vote count is necessary. As a result, any earlier votes counts conducted in such a discussion is removed from consideration.

2.3.3 Committees Data

The fields and descriptions of the *Committees* dataset can be found in Table 2.3. A Discussion ID represents a unique ID at which a single bill is discussed. Each committee has several hearings in which multiple discussions are held. A bill can be (and often is) discussed at several committees, as they must undergo approval from a variety of legislative standpoints.

2.3.4 Legislator Data

A part of this research concerns the analysis of legislator language patterns. As a result, the 10 most active legislators were selected from the database following the findings in Grace and Khosmood's "Feature Engineering for US State Legislative Hearings: Stance, Affiliation, Engagement and Absentees" [17].

2.3.4.1 Utterances

The spoken contributions of each of the selected legislators across all discussions in the database were extracted and logged in a pandas dataframe containing their Person ID, Discussion ID, and the transcribed utterance (Table 2.4. Because we are comparing this language to the speech patterns of the general public, the utterances are similarly grouped into speeches containing all spoken language by a single legislator withing a single discussion.

2.3.4.2 Votes

Every recorded vote by the selected legislators was extracted from the database as well, along with the legislator's Person ID, the Discussion ID, and the Vote ID (Table 2.5. Because some discussions required several rounds of voting, only the last vote for each discussion (defined as the largest Vote ID) is kept.

2.4 Machine Learning Models

A large portion of this research involves evaluating the performance of classification models on predicting testimony success. The Machine Learning models used to this

Table 2.3: Complete List of Committee Fields and Descriptions

Field Name	Description
Bill ID	Unique ID for each bill
Discussion ID	Unique ID for each bill discussed at a specific date and committee
Committee ID	Unique ID for the committee
Hearing ID	Unique ID for each hearing
House	Senate, Assembly, Joint
Name	Full Committee Name
Type	Standing, Floor, Joint, Budget Subcommittee, Extraordinary

Table 2.4: Complete List of Legislator Utterances Fields and Descriptions

Field Name	Description
First Name	Legislator's First Name
Last Name	Legislator's Last Name
PID	Person ID
DID	Unique ID for each bill discussed at a specific date and committee
Text	Utterance

Table 2.5: Complete List of Legislator Votes Fields and Descriptions

Field Name	Description
PID	Person ID
DID	Unique ID for each bill discussed at a specific date and committee
VID	Vote ID
Vote	"Aye", "Nae", or "Abstain"
Date	Date of the vote

end are Support Vector Machines, Gaussian Naive Bayes, Multinomial Naive Bayes, and Random Forest models, all of which are further discussed below.

2.4.1 Support Vector Machines (SVM)

Support Vector Machines are a set of supervised learning methods used for classification, regression, and outlier detection [29]. SVM operates by creating an n-dimensional separating hyperplane to separate points belonging to a categorization from points outside the categorization. The closest points to the hyperplane are called support vectors, and the distance between the support vectors and the plane is the margin. To determine the best separating hyperplane, the algorithm finds a line that is the most "in the middle" of the two groups, or in other words, the maximum-margin hyperplane. A soft margin allows for a specified number of errors in the data classification to avoid overfitting and skewing the representativeness of the model [26]. The benefits of SVMs include versatility as kernel functions in the decision process can be optimized to best fit the data, and effective learning in high dimensional spaces [29].

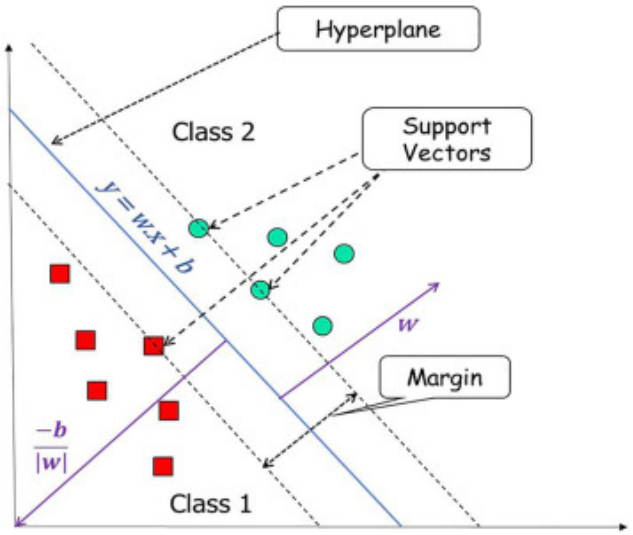

Figure 2.2: Support Vectors, Hyperplane, and Maximum Margin [28]

2.4.2 Random Forest

A random forest is a collection of decision trees that operates on random subsamples of the data and averages results to maximize predictive accuracy [20]. In order to reduce the correlation between the decision trees, the algorithm generates a random subset of features. Comparatively to individual decision trees, this method of averaging from a random subset of features ultimately reduces overfitting and can provide insight into feature importance [20].

2.4.3 Naive Bayes

The Naive Bayes Classifier is a probabilistic machine learning model that is very popular. It is based on Bayes Theorem which describes the conditional probability of two events. In the equation below the probability of A given B has already occurred is described [15]. For classification, A is a label and B is the "evidence."

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.1}$$

When there are many events (features) that make up the evidence, we naively assume they are mutually independent which makes it possible to find the parameters and apply the calculation much easier. The classifier makes a decision by finding the label for which the conditional probability $P(A|B)$ is highest.

This research utilizes two Naive Bayes classifiers: Gaussian and Multinomial Naive Bayes. Gaussian Naive Bayes assumes that that the continuous values of the feature set follow a normal distribution curve. Multinomial distribution is useful to model feature vectors where each value represents, for example, the number of occurrences

of a term or its relative frequency [32]. Since this is exactly the feature structure we will be using for our success predictions, this model is a good tool to leverage.

2.5 Statistics

Much of the evaluation phase of this research relies on statistical evaluation methods. Here, we establish a background on these methods.

2.5.1 Point Biserial Correlation Coefficient

Correlations are a statistical measure of relationships between two variables that indicate a variable's change in a specific direction as the other variable changes in value. While many correlation coefficients identify linear relationships, the Point Biserial Correlation coefficient compares dichotomous data to continuous data [13]. Assuming that the dichotomous variable is separated into two groups - group 0 and group 1 - and y is the continuous data (the language feature), the formula for Point Biserial Correlation is as follows [13]:

$$r_{pb} = \frac{M_0 - M_1}{s_y} \sqrt{\frac{n_0}{n} \frac{n_1}{n}} \tag{2.2}$$

Where:

M0 = the mean of the data from group 0.

M1 = the mean of the data from group 1.

Sy = the standard deviation of the continuous data.

n0 = the number of items in group 0.

n1 = the number of items in group 1.

n = the number of items in both groups together (aka the total rows in the data set).

2.5.2 ML Performance Metrics

Evaluating the performance of a machine learning model involves the use of statistical metrics. The confusion matrix allows us to observe how well the model was able to classify the provided data given a set of features.

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 2.3: Confusion Matrix [31]

In this research, we are classifying testimonies as 0 (unsuccessful) or 1 (successful). Therefore a True Positive represents a successful testimony that the model correctly classified as successful, a True Negative is an unsuccessful testimony that was correctly classified as unsuccessful, a False Negative is a successful testimony incorrectly classified as unsuccessful, and a False Positive is an unsuccessful testimony incorrectly classified as successful. Given this confusion matrix, we can calculate different performance measures to illustrate the strengths and weaknesses of the classification models.

2.5.2.1 Accuracy

Accuracy gives us an idea of how often the classifier was correct. Given the total number of classifications, how many were correctly labeled as positive (successful) or negative (unsuccessful)?

$$\text{accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{total}} \tag{2.3}$$

2.5.2.2 Precision

Precision is a measure of Predicted Positives. Out of all the testimonies the model classified as successful, how many were actually successful? This is calculated using the following equation.

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \tag{2.4}$$

2.5.2.3 Recall

Recall calculates how many of the Actual Positives were correctly captured by the model. In other words: out of all the successful testimonies, how many were correctly identified as successful?

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \tag{2.5}$$

2.5.2.4 F1 Score

The F1 score captures a balance between the precision (correctly classified instances) and recall (not missing a significant number of instances). In some cases, accuracy does not capture both the precision and robustness of a classifier, as high precision and low recall can still yield a high accuracy. The following equation is used for F1 calculation:

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{2.6}$$

2.6 Software Tools

2.6.1 pandas

pandas [23] is an efficient and flexible an open source data analysis library built on top of python. The pandas DataFrame structure is leveraged for almost all data manipulation and representation done in this research. Data was filtered, transformed, aggregated, indexed, and analyzed using pandas DataFrames. Important DataFrames were saved as CSV files and read back into the program by the pandas library.

2.6.2 SciPy

SciPy is an open-source scientific computation library with modules for statistics, optimization, and more. In this research, SciPy is used for the extraction of feature correlations - specifically the Point Biserial Correlation coefficient and corresponding p-values [34].

2.6.3 Scikit-Learn

Scikit-learn (or sklearn) is a machine learning python library built on SciPy (among other things). The Support Vector Machine, Naive Bayes, and Random Forest classifiers as well as feature normalization processes were implemented using this package [27].

2.6.4 NLTK

This program revolves around Natural Language Processing (NLP) techniques. NLP is the branch of computer science that links machines to human language and strives to allow machines to understand spoken and written language as humans do [3]. There are many different techniques that have been developed in the field of NLP to this end, some of which were central to the operations of the CPAT.

The Natural Language Toolkit (NLTK) [8] is an open source set of Python modules created by Steven Bird and Edward Looper in 2001. Many of the modules are central to the feature analysis involved in this research.

2.6.4.1 WordNet

WordNet is and large lexical database of English in which words are grouped into sets of synonyms (synsets) representing concepts, much like a thesaurus. The relationships between words are labeled hierarchically in "is-a" formats, which is particularly useful for measuring similarity [14].

2.6.4.2 NLTK Tokenizer

The NLTK tokenizer splits a given string into a list of "tokens". Any point in this research that involves identifying a certain word or phrase withing a string utilizes the NLTK tokenizer to do so. This is to avoid the unintentional matching of a given word *within* another word. For example, in trying to count the instances of the word "organ" within a phrase, searching a given string without tokenization would identify "**organ**ization" as a match. As this would clearly generate incorrect results, tokenization is applied to most instances of string analysis.

2.6.4.3 NLTK N-grams

N-grams are collections of n consecutive words within a given text. For example "esteemed committee members" is a trigram or (3-gram) and "thank you for your time" is a 5-gram. The NLTK n-gram module extracts all n-grams of a specified n value from a given text. For example, the phrase "I am a concerned constituent" tokenized into 2-grams would be: "I am", "am a", "a concerned", and "concerned constituent".

2.6.4.4 NLTK POS Tagging

The NLTK POS-tagger (Part of Speech tagger) takes in a sequence of words and attaches a part of speech tag to each word [9]. The parts of speech include verbs, nouns, adjectives, prepositions, and determinants. For example, given the phrase "I like to swim", the POS-tagger returns: ('I', 'PRP'), ('like', 'VBP'), ('to', 'TO'), ('swim', 'VB').

2.6.4.5 NLTK VADER

VADER (Valence Aware Dictionary for sEntiment Reasoning) is a model used for sentiment analysis that calculates the sentiment polarity (positive or negative) as well as intensity (strength) of a given text. Calculating this sentiment involves several steps, as there are many factors involved in the correct gauging of sentiment within a text. First, a lexical dictionary was created by having human testers assign sentiment scores to individual words and taking average human-assigned score for each word. This dictionary is referenced for each word in the input document; the scores are then compounded and evaluated with 5 additional heuristics. Punctuation, capitalization, degree modifiers, shift in polarity due to "but", and examining the three previous words to the phrase at hand to catch a polarity negation [11].

This sentiment analysis is outputted as a set of scores in four categories: negative, neutral, positive, and compound. The compound score is computed by normalizing the other three categories and serves as the sentiment metric for this research.

2.6.5 spaCy

spaCy is another open source python library for natural language processing [4]. spaCy uses convolutional neural network models to parse natural language documents and perform POS tagging, dependency parsing, named entity recognition, and more.

2.6.5.1 Named Entity Recognition

A named entity is a real world object that is assigned a name [2]. spaCy recognizes the named entities in a given document with pre-trained model predictions. Some examples of entities are companies, monetary values, and cities.

Chapter 3

RELATED WORK

3.1 Policy Influences

As we know, there are many factors that contribute to legislative policy decisions. The phraseology of testimonies presented at legislative committee meetings is just one many potential influences that committee members must take into consideration when voting for or against a bill. Lobbyists, private interests, and political affiliations are just some examples of additional influences in political decision making. Research into these influences has yielded some pertinent information.

3.1.1 Quantifying Politician Relationships to Interest Groups

Kim and Kunisky's "Mapping Political Communities: A Statistical Analysis of Lobbying Networks in Legislative Politics" [22] proposes a new methodology for inferring relationships between politicians in the 113th U.S Congress and special interest groups. This research first constructs a database linking politicians who sponsor bills to special interest groups lobbying those bills. With textual lobbying data, overlapping mentions of specific bills can indicate shared interests between these two parties, with the number of mentions quantifying the strength of the relationship.

The Bipartite Link Community Model (biLCM) is then applied to this data to model the relationships between political actors and special interest groups with a focus on the acknowledgement that both groups can be a part of several communities simultaneously. As a result, the biLCM models interactions between groups as "the

sum of independent interactions in all possible communities rather than a mixture of possible interactions in different communities" [22].

The research found that the biCLM [22] successfully illustrated many of the community memberships of political actors and how interactions occurred in those communities. The methodology allows for a quantitative analysis of the extent of participation in these communities for each political actor, and provides insight into the complex relationships that drive legislative lobbying [22].

3.1.2 How Lobbying Affects Public Policy

Baumgartner et al.'s "Money, Priorities, and Stalemate: How Lobbying Affects Public Policy" observes how money impacts policy changes through a large scale interview-based study of nearly 100 randomly selected lobbying cases in Washington. The research found that there was virtually no impact of money on outcomes, most likely because corporate bias is already woven into the structure of legislative processes. There is a distinction between changing policy and establishing policy. The status quo has already been established to protect corporate interests with money, and the vast majority of lobbying is about changing the status quo. The biggest impact of money was found in the disparity between public interests and lobbyist agendas. Corporate entities with the means to organize grass roots campaigns to protect their interests have the loudest voices when it comes to speaking for their interests. [7]

3.2 Linguistic Analysis of Real Testimonies

The analysis of actual discourse from political debates has been the basis of much research over the past few decades. Using real data can allow researchers to make observations and inferences pertinent to real world situations.

Several research projects have stemmed from the Digital Democracy database [19] in the analysis of language in legislative proceedings.

3.2.1 Predicting the Vote Using Legislative Language

Aditya Budhwar's "Predicting the Vote Using Legislative Language" [10] explores the analysis of utterances made by legislators in these committee hearings to predict their final vote. To do this, Budhwar extracts features such as speech length, number of interruptions, sentiment features, number of questions, and custom dictionary features from the legislator speeches. Each of these features can indicate an inclination for or against the bill, for example a higher number of interruptions made by an individual in discussing a specific bill can indicate they have issues with the points being made. Budhwar then explores several supervised learning prediction models, ultimately focusing on Support Vector Machines (SVM), Random Forest, and TensorFlow Neural Networks. Budhwar was ultimately able to achieve accuracies up to 83% in predicting votes given speeches [10].

3.2.2 Legislative Language of Success

Sanjana Gundala's "Legislative Language of Success" [18] follows a similar research flow with the aim of predicting the success of a non-legislator speech on the outcome of a vote. In Gundala's research, "success" was defined in the same way as in this

research: the speaker's alignment matches the bill's outcome. Gundala implemented two types of classification models in her work: the first implemented manual feature extraction prior to feeding into the model, the second made us of the models' preprocessing feature extraction capabilities. The models used with manual feature extraction were Naive Bayes, SVM, and FCNN, and the models that conducted feature extraction through preprocessing were TF-IDF and BERT. Gundala found the most successful models to be SVM and Multinomial Naive Bayes, and was able to predict the success of a testimony with an accuracy close to 90% [18]. Gundala's research is foundational to this thesis as we will be using similar techniques to analyze the success of previous speeches in comparison to the input speech. [18]

3.2.3 Learning Alignments from Legislative Discourse

Kauffman et. al. also researched linguistic patterns in legislative processes in their paper "Learning Alignments from Legislative Discourse" [21]. The research aims to identify legislator alignments with organizations or entities based on their language employed in committee meetings. The authors use the labeled organization alignments along with the legislator language and voting history to predict alignment scores. Then, they employed machine learning techniques to determine the accuracy of these predictions using the discussion text, utterance frequency, utterance duration, bias corpus hit rate, sentiment score, donations, and political party as their features. The authors decided to run their experiments on the each feature individually and then every possible combination of features, otherwise known as the "power set" of features. They were ultimately able to achieve a 78% accuracy of alignment prediction with the combination of all features, although this high accuracy level compared to previous experiments could be attributed to the binary nature of their prediction value (rather than ternary) [21].

3.3 Persuasive Language

A large part of this research involved the identification and extraction of language features that could contribute to a testimony's success. What linguistic choices can be made to maximize the strength of a persuasive argument in the scope of legislative decision making? For example, it can be hypothesized that the total length of a testimony indicates a higher or lower chance of success. While crafting a short and concise speech could theoretically help get one's message across more clearly, a longer speech could allow for more in-depth evidence and arguments. Which is the better strategy? By analyzing the success of past testimonies with different lengths, we can observe the presence (or absence) of patterns in testimony success based on individual language features.

3.3.1 Learning to Classify Documents According to Formal and Informal Style

Sheikha et al. explores the classification of documents as formal or informal by extracting features indicative of formality from bodies of text [30]. Their research found up to 98.5% accuracy in the classification for general-domain texts at the document level. Some of the features outlined as formal are:

1. The use of impersonal style and passive voice.

2. The lesser use of contractions.

3. The use of complex words and sentences.

4. The use of objective style, citing facts and references.

These features inspired some of the features extracted for this research, following the hypothesis that formal speech in a legislative testimony may improve its efficacy. [30]

3.3.2 Analysis and Detection of Persuasive Discourse

The work of Khazaei et al. shows promising results in the machine learning identification of persuasive language using linguistic dimensions. Some of the features that have been shown to be the most defining in persuasiveness are argument structure [12], argument content [24], and comprehensibility of text[5]. Frameworks for the study of argument structures have been studied for years [33]. Khazaei et al utilize Reddit's Change My View sub-reddit to conduct their research. Within the Change My View sub-reddit, a user posts an original viewpoint and instructs other users to persuade them otherwise, then he most convincing arguments are tagged by the user with a delta. This pre-tagged dataset allows for the analysis of the language that is deemed persuasive.

This analysis was conducted using the Linguistic Inquiry and Word Count (LIWC) tool which consists of categories belonging to four main processes: lingustic processes, psychological processes, personal concerns, and spoken categories. The research found 21 features with statistically significant t-test results between persuasive and non-persuasive groups. These features were used for three supervised learning algorithms using 10 fold cross validations: Ligistic, Sequential Minimum Optimization, and Regression. The average resulting F-score was 0.75, suggesting that these language features can greatly contribute to the persuasiveness of a message.

Chapter 4

SYSTEM

4.1 Defining Success

When speaking at a legislative committee, one's goal is to either argue for or against a proposed bill. The success of a speech is determined by comparing the speaker's alignment (*for* or *against*) to the bill's outcome. If the speaker is arguing for a bill and the bill passes or if the speaker is arguing against the bill and it does not pass, then the speech is considered successful. Conversely, if the alignment and the outcome do not match, the speech is considered unsuccessful.

4.2 Natural Language Features

4.2.1 Feature Extraction

The first step in the research process was to compile a list of linguistic features that were hypothesized to be correlated with success. This research builds off of the work done in Gundala's Legislative language of success [18], whose complete list of features can be found in Figure 4.1. In addition the Gundala's 16 language features, 33 more linguistic features were extracted from the dataset. The primary objective of this feature engineering is to identify speech patterns that may be structurally and linguistically advantageous in the pursuit of legislative persuasion. The list of new features and their descriptions can be found in Table 4.1.

After defining a concrete list of features to explore, the features are extracted from the database of testimonies. The program iterates through each testimony in the dataset using Pandas built in iterrows() function and extracts each linguistic feature from the row.

Features	Description
person_type	General Public or Lobbyist
alignment	For or Against or Indeterminate
word_count	Number of Words in Testimony
sentence_count	Number of Sentences in Testimony
avg_sentence_length	Average number of words in sentences
flesch	Flesch Readability Score
smog	Simple Meaure of Goobledygook (SMOG) Readability Score
successful_trigrams	Number of successful 3 word phrases
successful_quadgrams	Number of successful 4 word phrases
successful_pentagrams	Number of successful 5 word phrases
avg_connections_word	Average number of connections per word
avg_connections_sent	Average number of connections per sentence
max_connections	Max number of connections for one word in a testimony
NN	Proportion of nouns in testimony
VB	Proportion of verbs in testimony
DT	Proportion of determinates in testimony
PRP	Proportion of Proper nouns in testimony
JJ	Proportion of Adjectives in testimony
outcome	Outcome of the Bill (0 or 1)

Figure 4.1: Gundala Original Features List with Descriptions [18]

4.2.2 Feature Selection

Now that the features have been extracted from the testimonies database, we want to select the features that are the most highly correlated with success. This was done using SciPy's Spearman R function to evaluate the correlation between each individual column from the features dataset and the outcome column, as well as to compute the p value for each correlation. The top 15 correlations can be referenced in Table 6.1 and the full set of results can be referenced in Appendix A.2. As we can see in Table 6.1, Trigrams, Quadgrams, Pentagrams, Vader Sentiment, Word

Table 4.1: List of New Linguistic Features with Descriptions

Features	Descriptions
Question Count	Number of questions in testimony
Vader Sentiment Score	The Composite Sentiment Score for the entire testimony
Contraction Count	Number of contractions in testimony
Expansion Count	Number of possible contractions in expanded form
Cardinal Reference Count	Number of numerals referenced that do not fall under another entity type
Date Reference Count	Number of absolute or relative dates or periods referenced
FAC Reference Count	Number of buildings, airports, highways, bridges, etc. referenced
Law Reference Count	Number of references to named documents made into law
Money Reference Count	Number of references to monetary values
NORP Reference Count	Number of references to nationalities or religious or political groups
Percentages Reference Count	Number of references to percentages
Person Reference Count	Number of references to people, including fictional
Quantity Reference Count	Number of references to measurements, as of weight or distance
Current Bill Reference Count	Number of references to the current bill being discussed
Other Bill Reference Count	Number of references to another bill
Judgements Pro and Con Phrases	Number of phrases from this Word Menu category
Reasoning and Informing Phrases	Number of phrases from this Word Menu category
Reason and Rationale Phrases	Number of phrases from this Word Menu category
Order, Hierarchy, and Systems Phrases	Number of phrases from this Word Menu category
Judgements and Critiques Phrases	Number of phrases from this Word Menu category
Approval, Respect, and Recognition Phrases	Number of phrases from this Word Menu category
Support, Encouragement, and Agreement Phrases	Number of phrases from this Word Menu category
Disapproval, Disrespect, and Denial Phrases	Number of phrases from this Word Menu category
Opposition, Disagreement, and Attack Phrases	Number of phrases from this Word Menu category
Passive Percent	Percentage of sentences that are structured in passive voice
Very Short Sentence Count	Number of sentences under 5 words
Short Sentence Count	Number of sentences between 5 and 10 words
Medium Length Sentence Count	Number of sentences between 10 and 20 words
Long Sentence Count	Number of sentences between 20 and 30 words
Very Long Sentence Count	Number of Sentences over 30 words
Ambiguity	Average Number of Word Net Synsets per word
Contrast	Number of contrast words such as “but”, “yet”, “however”

Menu Categorizations, Contractions, and Parts of Speech Ratios all demonstrate statistically significant correlations with success and are therefore valuable to evaluate as a part of our statistical analysis.

The selected features are discussed in depth below.

4.2.2.1 Successful Phrases

This feature was first developed by Gundala in "Legislative Language of Success"[18]. This feature allows us to observe and count the phrases that statistically appear most in successful testimonies. Given the nature of persuasive testimonials, it is important to distinguish between language arguing for the passing of a bill and language arguing against the passing of a bill. As a result, the testimonies were first separated according to alignment. Then, the testimonies were tokenized into 3-grams, 4-grams, and 5-grams using NLTK's n-grams module and the occurrences of each n-gram were counted. To find the success rates of each n-gram according to the speaker's alignment, 4 pieces of information were recorded.

1. N-gram occurrences in successful testimonies arguing "For"

2. Total n-gram occurrences in testimonies arguing "For"

3. N-gram occurrences in successful testimonies arguing "Against"

4. Total n-gram occurrences in testimonies arguing "Against"

Each n-gram's success rate was computed as its number of successful occurrences in testimonies of matching alignments divided by the total number of occurrences of the n-gram in all testimonies of matching alignment. In other words, of all the times this n-gram was used with an alignment for/against a bill, what percentage of

time was the testimony successful? N-grams that appeared under 10 times across all testimonies were filtered out as well as n-grams with success rates under 50%.

Now given the list of successful n-grams, each testimony was searched for occurrences of n-grams in the list. The number of successful n-grams detected in the testimony was recorded for trigrams (n = 3), quadgrams (n = 4), and pentagrams (n = 5) [18].

4.2.2.2 Vader Sentiment Score

The overall sentiment score was computed from the each testimony using NLTK's VADER module. The compound sentiment score was a value between -1 and 1, with a score approaching 1 indicating a positive overall sentiment and a score approaching -1 as a negative overall sentiment.

4.2.2.3 Contraction and Expansion Count

As observed by Sheikha et al. [30], the use of contractions can be a sign of informal style . To explore whether this has any impact on a testimony's success, the number of contractions was counted as well as the use of their expanded forms. This ultimately gives us an idea of the number of possible contractions that were used in their contracted and expanded forms. To correctly identify contractions, we used a list of contractions and their expanded forms that was extracted from Wikipedia and shared on stack overflow [6].

4.2.2.4 Word Menu Phrases

Word Menu by Stephen Glazier is a reference book that organizes language by subject matter [16]. Although this book was published in the early 1990s and may not contain

a comprehensive vocabulary for the modern English language, it is an incredibly valuable resource to identify the use of words and phrases from a variety of different categorizations. Word Menu is structured into broad classifications of words, such as "Science and Technology", "Arts and Leisure", and "The Human Condition". Within these classifications are various categories which are further broken down into final groupings of words or phrases and their corresponding definitions. Nine of these groupings were selected from the "Human Condition" classification in the "Action and Sense" and "Cognition" categories based on their relevance to legislative persuasion tactics. The selected language classifications were:

1. Judgements Pro and Con
2. Reasoning and Informing
3. Reason and Rationale
4. Order, Hierarchy and Systems
5. Judgements and Critiques
6. Approval, Respect and Recognition
7. Support, Encouragement, and Agreement
8. Disapproval, Disrespect, and Denial
9. Opposition, Disagreement, and Attack

The complete list of phrases for each category extracted from the Word Menu can be found in Appendix A.1.

4.2.2.5 Parts of Speech

The parts of speech of each testimony were tagged using NLTK's POS Tagger module. The number of tags for verbs, nouns, adjectives, prepositions, and determinants were counted and recorded as separate features.

4.2.3 Normalization

To maintain uniformity in feature analysis, all features were extracted as ratios with respect to the number of sentences rather than raw counts. For example, the "Question Count" feature counts the number of sentences posed as questions and divides the results by the total number of sentences, giving us the percentage of sentences in question form. In addition to this uniform extraction, all numerical features were normalized using the MinMaxScaler technique from sklearn's preprocessing module.

4.3 CPAT for a Proposed New Testimony

Now that we have extracted features from the database and selected the feature set we want to work with, we can observe language patterns that correlate with a testimony's success. The next step is to use this information to analyze a new proposed testimony and compare its language patterns to successful ones.

The features that were most highly correlated with success were determined to be the most relevant to include in the suggestion tool. The input speech is broken down and analyzed by the aforementioned list of natural language features (such as passive/active voice, sentence complexity, sentiment analysis, etc). The breakdown of these features is then compared to previous speeches in a similar legislative con-

Table 4.2: Committee Topic Extraction Example

Committee Name	Extracted Keywords
Assembly Standing Committee on Water, Parks, and Wildlife	[Water, Parks, Wildlife]
Senate Standing Committee on Natural Resources and Water	[Natural Resources, Water]
Senate Standing Committee on Labor and Industrial Relations	[Labor, Industrial Relations]
Assembly Standing Committee on Labor and Employment	[Labor, Employment]
Senate Standing Committee on Elections and Constitutional Amendments	[Elections, Constitutional Amendments]

text. Finally, a statistical analysis of the language is generated based on statistical probabilities of success.

4.3.1 Committee Topic Based Analysis

Each committee meeting is held by a specific committee, such as the Senate Standing Committee on Health or the Assembly Standing Committee on Privacy and Consumer Protection. It is important to note that some committees may have overlapping topics. As can be seen in Table 4.2, the *Assembly Standing Committee on Water, Parks, and Wildlife* and the *Senate Standing Committee on Natural Resources and Water* are both concerned with the topic of Water. We hypothesized that topic of the committee at which the testimony is being presented may have an impact on the linguistic features involved in successful speeches. For example, a testimony at a committee on Public Health may need to be structured differently than a testimony at a committee on Higher Education.

The topics are extracted from the committee name using NLTK parsing tools. Some resulting topic keywords can be seen in Table 4.2.

Committees containing "Appropriations" or "Finance" keywords are removed, as these are committees that all bills must pass through and are ultimately not topical committees.

4.3.2 Analysis Output

Each of the selected features were discussed in detail in section 4.2.2. Now, these features are extracted from the provided input testimony and interpreted to generate human-readable statistical analysis. This section describes the analysis and resulting output of each of these features.

4.3.2.1 Word Menu

The Word Menu contains words that have been classified into specific language categories. The goal is to analyze the use of language categorized as "Pro and Cons", "Support, "Opposition", etc. that may strengthen a persuasive argument.

To this end, we first calculate the historical success rate of each phrase within the Word Menu categories. We do this by iterating through each testimony in the database and counting the occurrences of each phrase. We calculate the number of successful occurrences / the total number of occurrences of the given phrase and log the resulting category, phrase, and success rate into a new pandas DataFrame.

Now that we know the success rates for each phrase in the word menu, we can take a closer look at the input testimony. Positive feedback is outputted for any successful word menu phrases found in the input testimony. Then, for each phrase in the input testimony, the set of synonyms is extracted using NLTK's Word Net module and is stored into a dictionary of lists. The dictionary key is the input phrase and the value is its the list of synonyms. If a successful word menu phrase appears as a synonym of an input phrase, it is logged as a valid potential replacement.

The potential replacement is verified a final time by comparing the success rate of the replacement to the success rate of the actual phrase extracted from the input

document. If the synonym has a higher success rate, the Tool outputs a line describing the difference in success rates between the two phrases. It is important to note that there is no sense disambiguation in this process. As a result, the analysis is outputted as a statistical observation rather than a replacement suggestion.

4.3.2.2 Vader Sentiment

The input testimony's sentiment score is extracted using NLTK's VADER module as a numerical representation of the testimony's positive/negative language. This score is compared to the average extracted score from the successful testimonies in our database. A few statistics are calculated and outputted to provide a more detailed illustration of the comparison.

If the input testimony has a higher sentiment score than the average successful score, the percentage of successful testimonies with a lower score than the input's score is calculated. If the input testimony has a lower sentiment score, the percentage of testimonies with a higher score is calculated. Then, the actual difference between the input's score and the average successful is outputted. Because a scale of -1 to 1 may seem arbitrary to the user, this is outputted as a percent using Equation 4.1.

$$\frac{(\text{avg successful score} - \text{input score})}{2} * 100 \tag{4.1}$$

Finally, to offer an idea of which words in the input testimony are impacting the sentiment score the most, the sentiment scores are calculated for each individual word in the input testimony. If the input testimony's score is too high, the individual words with the highest sentiment are outputted and vice versa.

4.3.2.3 Parts of Speech

The parts of speech are extracted from the input document using NLTK's [8] POS tagger and are compared to the POS features extracted from the successful general public testimonies matching the input speaker's alignment. The POS features in question are the number of verbs, nouns, adjectives, prepositions, and determinants in the testimony. To ensure a size agnostic comparison, the ratio of each feature to the total number of POS-tagged words is calculated.

For each of these features, the average value across the aforementioned successful testimonies is calculated and compared to the input's extracted value. If the input testimony's ratio of nouns to total words is over 10% greater or smaller than the average successful ratio of nouns to total words, the statistical difference is outputted to the user. As an added evaluation of the testimony, the sentence that most exemplifies the fault is outputted as well. For example, the input testimony contains 20% more adjectives than the average successful testimony. "My house burned in the fire" has a lower ratio of adjectives than the other sentences.

4.3.2.4 N-grams

Given the pre-extracted successful phrases across all testimonies in the dataset, the input testimony is examined for exact matches or similar phrases. The successful phrases are extracted from the general public testimonies dataset prior to the feature extraction phase; a detailed explanation of this process can be referenced in Section 4.2.2.

The pre-extracted successful phrases are read from CSV files and saved into pandas dataframes containing the phrase, success rate, successful occurrences, and total oc-

currences for each successful phrase. The input testimony is tokenized into 5-grams, 4-grams, and then 3-grams in decreasing order so as to identify longer phrases first that may have their own significance. For each round of tokenization, the testimony is analyzed in two ways. It is first searched for exact matches of successful phrases, which are outputted with positive feedback, highlighting that they have been successful in the past. Then, the input n-grams are split into (n-1)-gram chunks. For example, "thank you for your time" may be an n-gram that is found to have a higher success rate than "thank you for your consideration". If the n-gram "thank you for your" is found within the input testimony but is not followed by "time", it may be useful to know that "time" tends to have a higher success rate than "consideration" in this context. Matched n-grams or (n-1)-grams are removed from the input string once they are found, so as not to be counted twice.

The resulting analysis contains positive feedback for phrases that have been previously identified as successful, as well as notes for phrases that are very similar to successful phrases.

4.4 Legislator-Based Comparative Features

It was hypothesized that legislators could empathize more with testimonies that closely mirror their own language patterns, meaning that the language structure can be micro-targeted to individual legislators. The objective is to observe if a testimony presented to a certain legislator is more likely to be successful if the language structure is similar to the legislator's speech patterns. Therefore, the language features are extracted from the collection of the legislator's utterances, averaged, and then compared one by one to the language features of each testimony that was presented to that legislator. This process is explained in further detail below.

4.4.1 Data

To ensure we had enough legislator language data to work with, we selected the 10 most active legislators [17] from the Digital Democracy database [19] to analyze their speech patterns. Every spoken contribution (utterance) from the selected legislators was extracted along with each vote decision ("aye", "nae", or "abstain"). Using this data, we can extract our set of language features from the legislator utterances.

The general public data used in this phase of the research is limited to testimonies that were presented at discussions where at least one of the selected legislators sat on the committee and cast a vote.

4.4.2 Redefining Success

It is very important to note that the definition of "success" has changed for this phase of the research. While success was previously defined as a majority committee vote toward the speaker's desired outcome, it is now defined as a single legislator vote in their favor. For example, a speaker arguing for the passing of a bill where the legislator casts an 'Aye' vote would be considered successful. This means that the legislator was aligned with the speaker on an individual basis.

4.4.3 Language Features

The updated feature set is extracted from the general public testimonies as before, and can be referenced in Table 4.1 (new features) and Figure 4.1 (Gundala features).

The only feature type that changes meaning is the "successful trigrams", "successful quadgrams", "successful pentagrams" set of features. Previously, these were extracted

from the entirety of the general public testimonies dataset and filtered for only phrases with a success rate over 50%. Now, given that we are strictly analyzing language in relation to specific legislators, the "successful phrases" are re-extracted to represent the most common phrases used by a given legislator.

4.4.3.1 Comparative Features

Along with the general public language features, a set of comparative features is extracted as well. This comparative feature set gives us the distance between a general public member's speech patterns and the speech patterns of the legislator they are trying to persuade.

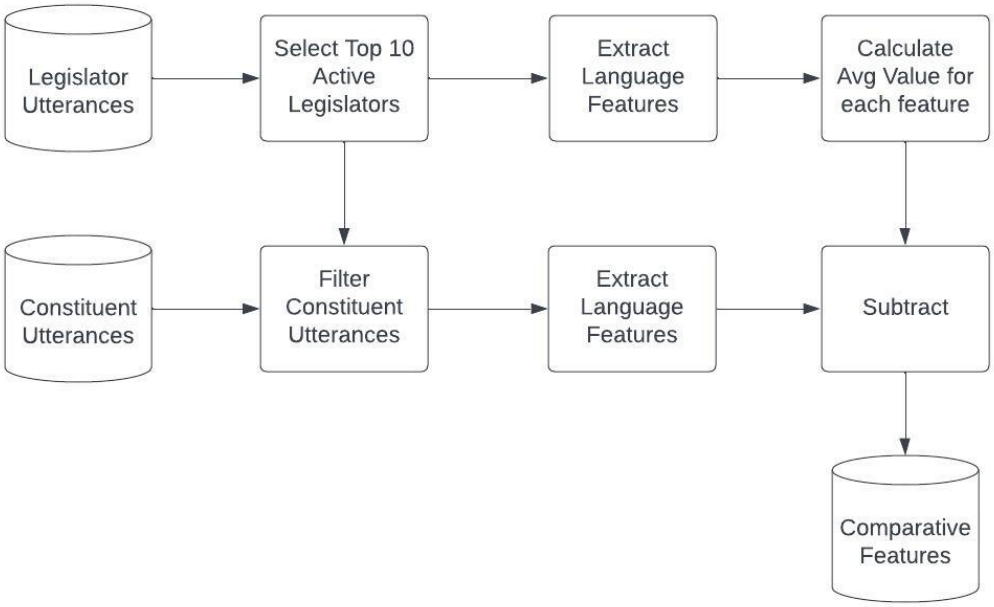

Figure 4.2: Comparative Legislator Language Feature Extractions

The process of generating the comparative feature set is illustrated in Figure 4.2. First, language features are extracted from every utterance from a single chosen legislator. Then, the average value for each feature is computed and stored. For example:

across every one of this particular legislator's testimonies, what was the average number of words per sentence? Then, the same language features are extracted from the all general public testimonies that were presented to that one legislator. This is called the general public language feature set. For each of the testimonies in the general public language feature set, the comparative feature set is computed as the distance between the testimony's extracted language features and the legislator's average value for that feature. This process is repeated for all 10 chosen legislators.

Chapter 5

METHODOLOGY

The research questions at the core of this research are as follows: Can we perform a statistical analysis of real testimonies presented at legislative committee meetings to identify and extract relevant language features with significant correlations to persuasive success? Can this statistical analysis be useful to have when crafting a new testimony?

5.1 Balancing Data For Outcomes

The original dataset is imbalanced, with almost double the number of successful speeches than unsuccessful ones as well as a much higher number of testimonies with alignments"For" than with alignments"Against". This size disparity is illustrated in Table 5.1.

Undersampling techniques were utilized to construct a balanced dataset with an equal number of testimonies for each of the four categories: "Successful For", "Successful Against", "Unsuccessful For", and "Unsuccessful Against". As "Successful Against" was the category with the least amount of data in this case, a random subset of each of the other categories was removed to form the balanced dataset. All experiments run in this research are conducted on both the complete original dataset and the balanced subset.

Table 5.1: Imbalanced Testimonies Count According to Outcome and Alignment

Outcome	Alignment	Testimony Count	Rate
Successful	For	19791	95.6%
	Against	922	4.4%
Unsuccessful	For	1069	10.4%
	Against	9199	89.6%

5.2 Machine Learning Predictive Evaluation

The first phase of testing observes the relevance of the extracted language features to the testimony's success (the vote going the speaker's way). This is done by leveraging machine learning models to predict the outcome of a testimony given the its extracted features. The language features are extracted from the full testimonies database and separated with an 20-80 test-train split. 80% of the testimonies are inputted into the models for training and the trained model then predicts the outcome label (1 for successful and 0 for unsuccessful) of the remaining 20% of testimonies in the test set. Each test is run 10 times and the results are averaged. Finally, we observe the Accuracy, Precision, Recall, and F1 scores for model's prediction capabilities. If the ML models can successfully predict the testimonies' success given the language features, we can conclude that the features contribute in some way to the testimony's success.

In addition to the ML predictive evaluation, the relevance of each feature was evaluated by calculating the correlation between the individual feature and the corresponding outcome (success or failure). The higher the absolute value of the correlation, the more relevant the feature is to the success of a testimony.

5.3 Expert Qualitative Evaluation

The second phase of evaluation is performed on the outputs of the CPAT. As natural language processing has not yet evolved to perform a reliable statistical evaluation on the validity of linguistic suggestions to legislative testimonies, this phase occurred as a validation process experts in the field. 2 experts - Christine Robertson and Dr. Cameron Jones - evaluated the usefulness of the CPAT's outputted statistical analysis. Christine Robertson is a former chief of staff in California state legislature and the current executive director of San Luis Coastal Education Foundation. Dr. Cameron Jones is a lecturer in Cal Poly's department of History and the political action / legislative chair for the California Faculty Association. Both individuals have had a lot of experience with the legislative processes involved in committee meetings, and therefore provided very valuable feedback on the usability of the CPAT.

Chapter 6

RESULTS

The results of this research are evaluated using Machine Learning techniques, statistical evaluations, and feedback from experts in their field. This section highlights the set up and outputs of the experimental results, please see the conclusion for an analysis of these results.

6.1 Natural Language Features

6.1.1 Feature Correlations With Success

Knowing the correlations between a testimony's success and each individual language feature from that testimony can provide valuable insight into the importance of that feature. The two variables we are looking at here are success (which is binary - 0 representing failure and 1 representing success) and a language feature (which is continuous). Because our "success" feature is dichotomous, trying to identify a linear relationship between the two variables is not ideal. Instead, we compute the Point Biserial Correlation coefficient, which compares dichotomous data to continuous data. These computations leveraged python's SciPy [34] library to calculate both the correlation and p-value between each vector of language features and the corresponding success values.

The correlation and p-values can for the top 15 features of the imbalanced and balanced datasets can be referenced in Table 6.1. The results for all 49 features can be found in Appendix A.2.

Table 6.1: Top 15 Feature Correlations with Success - Imbalanced and Balanced Data

Imbalanced		
Feature	**Correlation**	**P Value**
Successful Trigrams	0.686393	0.000000e+00
Successful Quadgrams	0.667406	0.000000e+00
Successful Pentagrams	0.598693	0.000000e+00
Word Menu Pro Con	0.593229	0.000000e+00
Word Menu Opposition	-0.562532	0.000000e+00
Word Menu Support	0.441505	0.000000e+00
Verb Count	-0.241925	0.000000e+00
Sentiment Score	0.227141	0.000000e+00
Preposition Count	-0.216733	0.000000e+00
Noun Count	0.168844	6.367316e-217
Sentence Count	-0.156517	2.553217e-186
Contractions Count	-0.154652	6.659898e-182
Word Count	-0.143304	3.286129e-156
Word Menu Reasoning	-0.139384	8.264485e-148
Expansions Count	-0.138418	8.926644e-146

Balanced Outcome + Alignment		
Feature	**Correlation**	**P Value**
Word Count	-0.101864	0.000030
Medium Length Sentences	0.095363	0.000094
Sentence Count	-0.093804	0.000122
Ambiguity	-0.092821	0.000144
Word Menu Pro Con	0.089072	0.000266
Flesch	-0.087792	0.000326
Preposition Count	-0.087389	0.000347
Contractions Count	-0.077138	0.001597
Word Menu Support	0.075672	0.001959
Successful Pentagrams	0.074809	0.002206
Successful Quadgrams	0.072298	0.003097
Verb Count	-0.071678	0.003363
Very Long Sentences	-0.064380	0.008457
Word Menu Opposition	-0.063480	0.009421
Word Menu Approval	-0.062987	0.009990

The imbalanced dataset contains 5 features with correlation coefficients of over 50% : successful trigrams, successful quadgrams, successful pentagrams, and phrases from Word Menu's "Judgements Pro and Con", "Opposition, Disagreement, and Attack", and "Support, Encouragement, and Agreement" language categories. The remaining 10 features in Table 6.1 are still above 10% correlation, which is a disputed zone for correlation strength. The p-values are extremely low - indicating strong statistical significance - and correlations extracted from the imbalanced dataset are overall much higher than those in the Balanced dataset.

6.1.2 Success Prediction Results

5 different machine learning models are used to to evaluate the usefulness of the new feature set in predicting success: Gaussian Naive Bayes, Multinomial Naive Bayes, Support Vector Machines, and Random Forest models constructed from python's sklearn [27] library. Each model takes a set of testimonies broken down into language features and predicts the testimony's success (0 for success, 1 for failure) based on those features. The performance of these models is measured using statistical metrics

Table 6.2: Success Prediction Results Feature Set Comparisons for Random Forest Model

Unbalanced				
Feature Set	**Accuracy**	**Precision**	**Recall**	**F1**
New	0.851178	0.876000	0.913231	0.894217
Gundala	0.929027	0.942671	0.955351	0.948966
Combined	0.934045	0.947017	0.958095	0.952521
Δ(Combined, Gundala)	0.005018	0.004346	0.002744	0.003555

Balanced Outcome + Alignment				
Feature Set	**Accuracy**	**Precision**	**Recall**	**F1**
New	0.581275	0.584570	0.542881	0.562289
Gundala	0.574369	0.579640	0.556639	0.567433
Combined	0.600398	0.604436	0.595339	0.599444
Δ(Combined, Gundala)	0.026029	0.024796	0.0387	0.032011

of Accuracy, Precision, Recall, and F1 Score. The experiments are performed on the raw imbalanced data, the data balanced on the testimony's outcome (or "success"), and the fully balanced data on combined outcome and alignment.

The experiment is further conducted on 3 different feature sets:

1. Gundala's 16 features alone.

2. The 33 new features alone.

3. The combination of Gundala's features and the new features - 49 total.

We are using Gundala's [18] original 16 features as a baseline from which to compare the performance of the new feature set. The full list of Gundala's features and descriptions can be referenced in Figure 4.1. The full list of new features and their descriptions can be referenced in Table 4.1. In total, the combination of features represents 49 natural language features.

Table 6.3: Success Prediction Results with Gundala Feature Set

Imbalanced				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.809568	0.959833	0.755968	0.845779
Multinomal NB	0.695494	0.694184	0.999555	0.819330
SVM	0.924389	0.947410	0.942993	0.945191
RF	0.929027	0.942671	0.955351	0.948966
Balanced Outcome				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.820142	0.915458	0.706257	0.797349
Multinomal NB	0.845053	0.915631	0.760907	0.831112
SVM	0.903164	0.901148	0.906027	0.903570
RF	0.916741	0.895622	0.942968	0.918677
Balanced Outcome + Alignment				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.533997	0.523612	0.724621	0.607530
Multinomal NB	0.543559	0.542418	0.574185	0.554393
SVM	0.558300	0.556587	0.565656	0.557087
RF	0.574369	0.579640	0.556639	0.567433

Table 6.3 shows the model performance metrics for each of the 5 ML models trained on the *old feature set* (the baseline performance measures). Table 6.4 demonstrates the models performance metrics for the 5 ML models on the *new features alone*. This gives us an idea of the new features' contribution to models' abilities to accurately predict success. Finally, Table 6.5 shows the model performance on the *combined old and new feature sets*.

Tables 6.3, 6.4, and 6.5 show that the Random Forest has the highest performance metrics in most cases, so the feature set comparison is taken from the Random Forest results. Table 6.2 illustrates the performance of each feature set in the Random Forest Model, as well as the difference in performance of the feature set before and after the addition of the new features. As the Δ is positive across the board, we can see that adding the new feature set improves the model's ability to correctly predict the success of a testimony.

Table 6.4: Success Prediction Results with New Feature Set

Imbalanced				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.768881	0.829709	0.837190	0.833396
Multinomial NB	0.775669	0.761651	0.982690	0.858148
SVM	0.819327	0.834408	0.919486	0.874875
RF	0.851178	0.876000	0.913231	0.894217
Balanced Outcome				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.728831	0.701046	0.803434	0.748705
Multinomial NB	0.750177	0.790674	0.684342	0.733629
SVM	0.817025	0.873819	0.738805	0.800536
RF	0.838158	0.857027	0.812982	0.834380
Balanced Outcome + Alignment				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.557238	0.545808	0.675429	0.602639
Multinomial NB	0.558566	0.569868	0.466125	0.512167
SVM	0.571315	0.576609	0.515681	0.543297
RF	0.581275	0.584570	0.542881	0.562289

Table 6.5: Success Prediction Results with Updated Feature Set

Imbalanced				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.804974	0.901596	0.805868	0.851028
Multinomial NB	0.853241	0.858243	0.943591	0.898874
SVM	0.931529	0.947926	0.953267	0.950588
RF	0.934045	0.947017	0.958095	0.952521
Balanced Outcome				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.801677	0.806737	0.789956	0.798203
Multinomial NB	0.790059	0.822738	0.735788	0.776798
SVM	0.918158	0.902392	0.937696	0.919691
RF	0.925195	0.905067	0.949834	0.926899
Balanced Outcome + Alignment				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.559495	0.547079	0.775345	0.640882
Multinomial NB	0.562284	0.583550	0.488968	0.531444
SVM	0.573174	0.570518	0.599060	0.583397
RF	0.600398	0.604436	0.595339	0.599444

6.2 CPAT

To evaluate the Constituent Phraseology Analysis Tool, we originally planned to run a full user study with a collection of journalists, lobbyists, and other individuals with experience in and around legislative committee meetings. After meeting with Robertson and Jones, it was determined that a large scale user study would not be feasible for a few reasons. The background of the research and interpretation of the output was too complex for a large group of people to understand in very limited time, and personal feelings towards machine learning solutions in non-technology centered spaces was too varied to get an accurate representation of the tool's usefulness.

When meeting with Robertson, it became clear that a statistical analysis of vocabulary and language structure may not be appealing to many users who have a clear idea of their message and a strong attachment to their personal voice. A testimony can be deeply personal, with a direct message rooted in strong convictions. Knowing that a certain phrase tends to have a higher percentage chance of success may not be enough for the speaker to want to make changes to their testimony.

In discussing the potential uses of the CPAT with Jones, we uncovered the potential for micro-targeted language. In taking from his extensive experience speaking in legislative committee meetings, Jones suggested that mirroring the committee members' speech patterns could potentially influence their reaction to the arguments being made. From this discussion, it was determined that focusing on the persuasive strength of the language features would ultimately best support the CPAT's usefulness.

Table 6.6: Success Prediction Results with Legislator-Based Comparative Language Features

Original Features				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.801270	0.869632	0.835060	0.851995
Multinomial NB	0.825069	0.846660	0.909249	0.876839
SVM	0.872539	0.867260	0.959465	0.911036
RF	0.883549	0.895329	0.939769	0.917011
Comparative Features + Original Features				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.779757	0.853420	0.820422	0.836596
Multinomial NB	0.806989	0.858876	0.860434	0.859654
SVM	0.876241	0.872223	0.959003	0.913557
RF	0.924909	0.934133	0.958406	0.946114
Performance Δ				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	-0.021513	-0.016212	-0.014638	-0.015400
Multinomial NB	-0.018080	0.012217	-0.048815	-0.017184
SVM	0.003703	0.004963	-0.000463	0.002521
RF	0.041361	0.038804	0.018637	0.029103

6.3 Legislator-Based Comparative Features

There are two distinct feature types in this evaluation:

1. Language features from general public testimonies.

2. Comparative features between the average value for the legislator's language features and the general public testimony's language features.

The results in Table 6.6 are represented in 3 sections.

6.3.0.1 Weighing Comparative Results

In the evaluation of the comparative features and original features, the two feature sets are weighed equally. To have a better idea of the comparative features' importance within the concatenated feature set, the features were assigned custom weight distributions based on their Feature Type. Table 6.7 illustrates the results. First, the prediction models were run with 75% weight on comparative features and 25% weight on original language features (sub-table A). Then, 25% weight on the original features and 75% weight on the comparative features (sub-table B). Δ(A, B) shows that the Gaussian Naive Bayes, Multinomial Naive Bayes, and Random Forest models using concatenated feature set actually perform marginally better with heavier weights assigned to the comparative features than the original features. This is additionally illustrated in Figure 6.1.

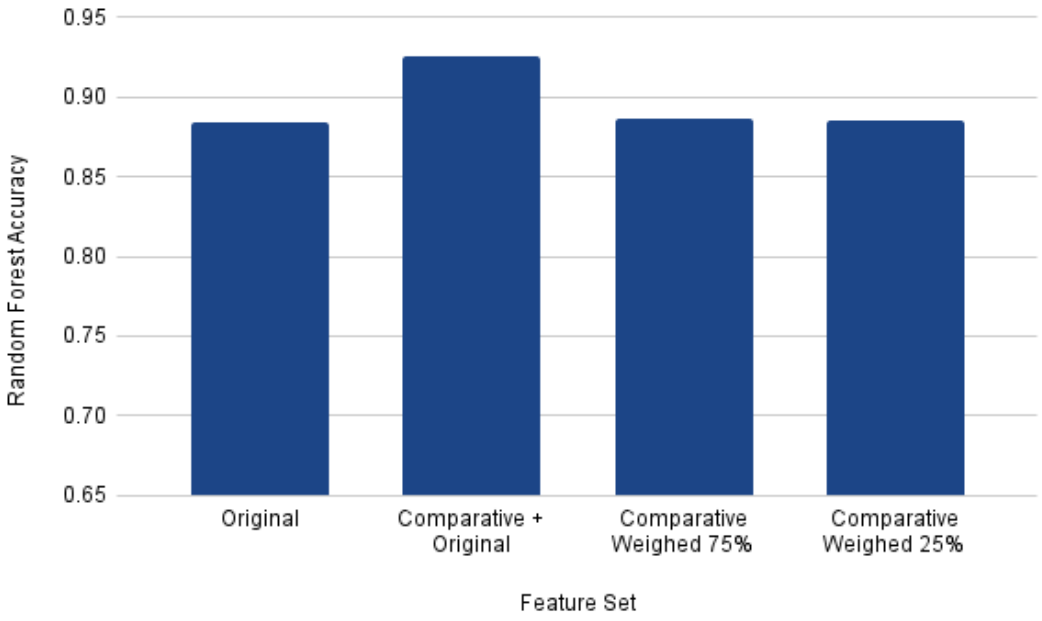

Figure 6.1: Random Forest Accuracy For Comparative Feature Sets

Table 6.7: Success Prediction Results with Weighted Legislator-Based Comparative Language Features

A

Comparative Features 75% / Original Features 25%				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.739399	0.803392	0.820233	0.811725
Multinomial NB	0.754485	0.813015	0.833200	0.822984
SVM	0.840571	0.841760	0.946500	0.891062
RF	0.886086	0.885646	0.957921	0.920367

B

Comparative Features 25% / Original Features 75%				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.734456	0.800157	0.817478	0.808725
Multinomial NB	0.751630	0.811586	0.831312	0.821331
SVM	0.843597	0.842562	0.949283	0.892744
RF	0.885276	0.886250	0.954497	0.919108

Δ (A,B)				
Model	**Accuracy**	**Precision**	**Recall**	**F1**
Gaussian NB	0.004944	0.003235	0.002755	0.003006
Multinomial NB	0.002855	0.001428	0.001888	0.001650
SVM	-0.003026	-0.000802	-0.002783	-0.001683
RF	0.000810	-0.000603	0.003424	0.001266

Chapter 7

CONCLUSION

As we begin to draw conclusions from the results we have observed in this work, we must first acknowledge the assumptions made in the premise of the experiments. It is clear that decisions made in legislative processes are reliant on a complex web of factors outside of the arguments made by constituents. As a result, we cannot truly quantify the extent to which the general public actually can influence legislative votes. We therefore acknowledge that our language corpus may contain testimonies labeled as successful that ultimately had little to no influence the trajectory of a vote. However, we do maintain the objective of analyzing linguistic techniques at whatever level they may help constituents to craft effective persuasive arguments.

7.1 Feature Correlations With Success

The top 15 feature correlations for both balanced and imbalanced data can be referenced in Table 6.1.

7.1.1 Statistical Significance

As was noted in Results section 6.1.1, the calculated p-values are extremely low, particularly for the imbalanced dataset. A low p-value indicates a high statistical significance, meaning that it is very unlikely that the correlation was due to chance.

For example, Word count has a correlation coefficient of -0.143304 and a p-value of 2.386129e-156. A 14% correlation is considered low by many standards, but a p-value of close to 0 illustrates an extremely high significance level.

While it seems counter-intuitive for a low correlation to have a high statistical significance, it may be explained by the immensity of the dataset. The large amounts of data allow for the detection of very small correlations with high degrees of certainty. This theory may also explain the higher p-values in the balanced dataset, as the balanced data is an undersampled subset of the original data and is therefore much smaller (although still quite sizeable).

Word count in the balanced data has a correlation coefficient of -0.101864 and a p-value of 0.000030. While this p-value still passes the 0.05 statistical significance threshold, it is many orders of magnitude larger than the p-value of the same feature in the imbalanced data.

It is important to keep in mind that while significance levels are very high, the p-value measures the significance of the *correlation* and not the predictive importance. So while we can say with a high degree of certainty that the features are more or less correlated with success, this does not directly translate to a strong ability to predict success with just the one feature.

7.1.2 Balanced Data Correlations

The success to language feature correlation coefficients are lower across the board when extracted from the balanced dataset. This could be due to the fact that the imbalanced data has significantly more successful testimonies to reference. Having a more expansive dataset allows for more patterns to be identified with higher confidence levels.

7.2 Success Prediction with New Features

7.2.1 Effects of Balancing Data

Looking at Table 6.2, it is clear that the Random Forest models perform significantly worse on the Balanced dataset than the imbalanced dataset. This pattern is consistent through Tables 6.3, 6.4, and 6.5. This could be explained by the significant shift in data size through undersampling. While the raw, imbalanced dataset contains close to 31,000 testimonies, the balanced outcome + alignment dataset only contains about 3,700 testimonies. This means that the classification models have significantly less data to learn from and train on.

7.2.2 Performance of New Features

The purpose of introducing new language features to the classification models was to have a more broad understanding of language choices that can impact a testimony's success.

Table 6.2 illustrates the performance metrics for the Random Forest classifier on Gundadala's orginal feature set, the new feature set developed in this research, and the combined old and new features. The table also shows the performance delta for Random Forest classifier using just Gundala's features and using the combination of both feature sets. When training on the new language features alone, the models do not perform as well as Gundala's features with a 7.8% lower accuracy and a 5.8% lower F1 score on Random Forest classification on the unbalanced data. The baseline result for binary classifications is 50%, which would be the odds of correctly classifying a testimony at random. Notably, the classification results for the new feature set alone are well over 50%, with an accuracy of 85.1% and an F1 score of 89.4%. This means

that the new features on their own contribute an extra 35-40% to the classification abilities of the model.

In comparing the classification results of models using only Gundala's original features to the classification results of the updated feature set (Gundala's features + the new features), we see a positive trend in success prediction (referenced in Table 6.2). Gundala's feature set already performed very well in the testimony success prediction, with an F1 score of 94.8% and an Accuracy of 92.9%. The slight increase of these performance metrics combined with the significant correlations with success that were identified from some of the new features (Table 6.1) indicates that the new feature set is positively contributing to the classification of testimonial success.

7.3 CPAT

Many individuals have different reactions to the involvement of machine learning and artificial intelligence in non-technical spaces. Those who value the complexity and nuances of linguistic studies may have an aversion to using technology for linguistic analysis. On the other hand, individuals who prioritize persuasiveness may appreciate having access to a statistical evaluation of the language structures of their testimony. This wide range of perspectives combined with limited time availability from individuals in the political sphere made it impractical to evaluate the tool with a large scale user study.

In speaking with experts Robertson and Jones, it was concluded that the Constituent Phraseology Analysis tool has the potential to be a useful tool for providing statistical breakdowns of successful language features. However, the tool may be met with a mixed reaction from users in the political sphere. Developing a strong confidence in

the relevancy of the outputted analysis is crucial to the tool's usability, and therefore requires further development.

7.4 Legislator-Based Comparative Features

The analysis of micro-targeted language towards individual legislators yielded some interesting results. Table 6.6 shows that adding comparative features to the feature set increases the predictive accuracy of the Random Forest model by 4%, with precision, recall, and F1 scores increasing as well. Recall for other models decreased, however, suggesting that more successful testimonies were misclassified as unsuccessful. Because Random Forest classifiers use an ensemble of decision trees to mimick the wisdom of crowds, it could be that the number of erroneously classified testimonies is minimized by the self-correcting properties of the ensemble.

When observing the effects of weighting comparative features above the regular language features in Table 6.7, it is clear that the performance metrics are impacted by a very tiny margin. The largest change in performance when weighing the comparative features at 75% was a 0.4% increase in accuracy for the Gaussian Naive Bayes classifier. Given the small change in performance, it could be reasoned that this change is merely caused by the variance in model performance due to train/test splitting of the data. The size of the dataset used in this experiment could also be a reason for the limited changes in performance. Because this experiment was run on the testimonies presented to 10 select legislators, there may not be enough data to conclusively determine the effects of mirroring legislator language.

Ultimately, there is not currently enough data to conclude that approaching one's language features to mirror a legislator's speech pattern can effect the likelihood of a testimony's success.

Chapter 8

FUTURE WORK

Given more time, this research could be continued in a few ways. First, the exploration and validation of more language features could continue to broaden our understanding of the role of phraseology in legislative persuasion. Natural language is immensely complex, with an expansive set of rules, exceptions, nuances, and contextual dependencies that are ever evolving. The same argument can be made in infinitely many ways, and the language can be interpreted differently from person to person. The way a statement is received depends heavily on the perspective of the listener, given their past experiences, cultural upbringing, relationship with language, and more. There is no end to the number of language features that could be used to analyze the effectiveness of an argument in many different contexts.

Another route by which this research could be furthered is the continuation of the development of the Constituent Phraseology Analysis Tool. It was remarked by expert Cameron Jones that a tool which analyzes a proposed testimony for successful language structures and speech patterns could be very useful in the drafting/editing of a testimony. Expanding the statistical analysis and fine-tuning the outputs to best fit the needs of the user could lead to a very successful tool that would support constituents in effectively communicating their perspectives. One current issue facing the CPAT is the lack of sense disambiguation in the analysis. Much of the analysis is reliant on synonyms from WordNet's Synsets, however, many words carry several different meanings. For example, "read", "scan", "learn", and "study" are all synonyms, but they are not always interchangeable. In order to generate a reliable and useful

language editing suggestion based on statistical success, the suggested replacement must have the same sense (or meaning) as the original phrase.

Finally, further research into the analysis of mirroring legislator phraseology could yield clearer results as to the extent to which mirroring speech patterns affects the persuasive strength of a testimony. Repeating the experiment on a larger set of data could provide more conclusive results on the effects (or lack thereof) of similarity in speech patterns to the persuasive target.

BIBLIOGRAPHY

[1] In Committee | house.gov. https://www.house.gov/the-house-explained/the-legislative-process/in-committee.

[2] Linguistic Features · spaCy Usage Documentation. https://spacy.io/usage/linguistic-featuresnamed-entities.

[3] What is Natural Language Processing? | IBM. https://www.ibm.com/topics/natural-language-processing.

[4] spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, Jan. 2017.

[5] P. A. Alexander and T. L. Jetton. The role of importance and interest in the processing of text. *Educational Psychology Review*, 8:89–121, 1996. Place: Germany Publisher: Springer.

[6] arturomp. Answer to "Expanding English language contractions in Python". https://stackoverflow.com/a/19794953, Nov. 2013.

[7] F. R. Baumgartner, J. M. Berry, M. Hojnacki, D. C. Kimball, and B. L. Leech. Money, Priorities, and Stalemate: How Lobbying Affects Public Policy. *Election Law Journal: Rules, Politics, and Policy*, 13(1):194–209, Mar. 2014.

[8] S. Bird, E. Klein, and E. Loper. *Natural Language Processing With Python*. O'Reilly Media, 2009.

[9] S. Bird and E. Loper. 5. Categorizing and Tagging Words. In *NLTK: The Natural Language Toolkit*. 2009.

[10] A. Budhwar, T. Kuboi, A. Dekhtyar, and F. Khosmood. predicting the vote using legislative speech. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, pages 1–10, New York, NY, USA, May 2018. Association for Computing Machinery.

[11] P. Calderon. VADER Sentiment Analysis Explained. https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9, Mar. 2018.

[12] M. J. Chambliss and R. Garner. Do Adults Change their Minds after Reading Persuasive Text? *Written Communication*, 13(3):291–313, July 1996. Publisher: SAGE Publications Inc.

[13] J. DeJesus. Point Biserial Correlation with Python. https://towardsdatascience.com/point-biserial-correlation-with-python-f7cd591bd3b1, Nov. 2021.

[14] A. Farkiya, P. Saini, S. Sinha, and S. Desai. Natural Language Processing using NLTK and WordNet. 6, 2015.

[15] R. Gandhi. Naive Bayes Classifier. https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c, May 2018.

[16] S. Glazier. *Word Menu*. Random House Inc, New York, NY, USA, first edition, 1992.

[17] J. Grace and F. Khosmood. Feature Engineering for US State Legislative Hearings: Stance, Affiliation, Engagement and Absentees, Sept. 2021. arXiv:2109.08855 [cs].

[18] S. Gundala. Legislative Language of Success. Master's thesis, California Polytechnic State University San Luis Obispo, 2022.

[19] IATPP. Digital Democracy. https://iatpp.calpoly.edu/digital-democracy, 2020.

[20] IBM. What is Random Forest? https://spacy.io/usage/linguistic-featuresnamed-entities, Jan. 2021.

[21] D. Kauffman, F. Khosmood, T. Kuboi, and A. Dekhtyar. Learning alignments from legislative discourse. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o '18, pages 1–2, New York, NY, USA, May 2018. Association for Computing Machinery.

[22] I. S. Kim and D. Kunisky. Mapping Political Communities: A Statistical Analysis of Lobbying Networks in Legislative Politics. *Political Analysis*, 29(3):317–336, July 2021.

[23] W. McKinney. Data Structures for Statistical Computing in Python. pages 56–61, Austin, Texas, 2010.

[24] P. K. Murphy. What makes a text persuasive? Comparing students' and experts' conceptions of persuasiveness. *International Journal of Educational Research*, 35(7):675–698, Jan. 2001.

[25] NCSL. Infographic for Legislative Process. https://www.ncsl.org/portals/1/Media/modal$_b$*ootstrap.html*.

[26] W. S. Noble. What is a support vector machine? *Nature Biotechnology*, 24(12):1565–1567, Dec. 2006.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg,

J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Duchesnay. Scikit-learn: Machine Learning in Python, June 2018. arXiv:1201.0490 [cs].

[28] A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha. Chapter 6 - Machine learning for soil moisture assessment. In R. C. Poonia, V. Singh, and S. R. Nayak, editors, *Deep Learning for Sustainable Agriculture*, Cognitive Data Science in Sustainable Computing, pages 143–168. Academic Press, Jan. 2022.

[29] scikit learn. 1.4. Support Vector Machines. https://scikit-learn/stable/modules/svm.html.

[30] F. A. Sheikha and D. Inkpen. Learning to Classify Documents According to Formal and Informal Style. *Linguistic Issues in Language Technology*, 8, Mar. 2012.

[31] K. P. Shung. Accuracy, Precision, Recall or F1? https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9, Apr. 2020.

[32] P. E. Staff. Implementing 3 Naive Bayes classifiers in scikit-learn. https://hub.packtpub.com/implementing-3-naive-bayes-classifiers-in-scikit-learn, May 2018.

[33] S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, 2 edition, 2003.

[34] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors.

SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3):261–272, Mar. 2020. arXiv:1907.10121 [physics].

APPENDICES

Appendix A

APPENDIX

A.1 Full Word Menu Phrase Lists

The following categorized phrases are extracted from Stephen Glazier's Word Menu [16].

A.1.0.1 Judgements Pro and Con Phrases

"acknowledge", "affirm", "agree", "animadvert", "argue", "assent", "attack", "avouch", "backbite", "bad-mouth", "barrage", "belittle", "berate", "carp", "castigate", "comminate", "complain", "concur", "confront", "contradict", "criticize", "decry", "demur", "denigrate", "denounce", "deplore", "discredit", "disparage", "dispute", "dissuade", "espouse", "execrate", "flatter", "fulminate", "fuss", "hector", "judge", "laud", "lecture", "make fun of", "malign", "mock", "moralize", "nag", "niggle", "object", "opine", "poke fun at", "praise", "profane", "promote", "put down", "question", "quibble", "rate", "rebuke", "recommend", "remonstrate", "reprehend", "repremand", "reproach", "reprove", "revile", "ridicule", "salute", "scandalize", "scold", "score", "slander", "slash", "slur", "smirch", "sound off", "speculate", "support", "talk up", "tout", "trash", "upbraid", "vilify", "vindicate", "vote"

A.1.0.2 Reasoning and Informing Phrases

“accent”, “advise”, “allege”, “allude”, “apprise”, “attest”, “attribute”, “bandy”, “beat around the bush”, “belabor”, “break the news”, “cajole”, “cant”, “chaffer”, “chew over”, “clarify”, “clear the air”, “coax”, “collogue”, “communicate”, “con”, “confer”, “confide”, “constate”, “contend”, “contrast”, “convey”, “convince”, “counsel”, “decree”, “define”, “delineate”, “depict”, “describe”, “disclose”, “discourse”, “discourse”, “discuss”, “divulge, elucidate”, “emphasize”, “enounce”, “equivocate”, “explain”, “expound”, “fill in”, “get across”, “get over”, “gloss over”, “gossip”, “hash over”, “hint”, “hold forth”, “illustrate”, “imply”, “import”, “impute”, “induce”, “inform”, “insinuate”, “instruct”, “interpret”, “itemize”, “lay down”, “lecture”, “limn”, “list”, “name”, “narrate”, “notify”, “offer”, “open”, “outtalk”, “parley”, “persuade”, “philosophize”, “pitch”, “plead”, “postulate”, “powwow”, “preface”, “prelect”, “prescribe”, “proffer”, “propagate”, “propose”, “proselytize”, “psychologize”, “read”, “reason”, “rebut”, “recant”, “recap”, “recapitulate”, “recite”, “recount”, “refute”, “relate”, “relay”, “repeat”, “report”, “represent”, “resolve”, “restate”, “retell”, “retract”, “reveal”, “snitch”, “specify”, “spell out”, “spill”, “spill the beans”, “squeal”, “stipulate”, “stress”, “submit”, “suggest”, “summarize”, “sum up”, “sustain”, “sway”, “synopsize”, “take back”, “talk into”, “talk over”, “talk sense”, “tattle”, “tell”, “transmit”, “treat”, “typify”, “understate”, “unsay”, “unswear”, “unveil”, “validate”, “verify”, “voice”, “vouch”, “wheedle”, “withdraw”, “witness”, “word”, “work in”

A.1.0.3 Reason and Rationale Phrases

“adduce”, “analogy”, “analysis”, “antithesis”, “apologia”, “apprehension”, “argument”, “assumption”, “axiom”, “biconditional”, “brainstorm”, “circumscribe”, “clar-

ification", "clincher", "cogent", "cogitation", "cognition", "cognitive", "coherent", "comprehension", "conception", "confutation", "conjecture", "construe", "contend", "contest", "contradiction", "contradistinction", "contraposition", "controvert", "convention", "convince", "correct", "counterexample", "counterproposal", "criterion", "crux", "cumulative", "data", "datum", "debate", "deduction", "define", "definition", "definitive", "delimit", "demystify", "determine", "dialectic", "didactic", "differentiate", "discursive reasoning", "disjunction", "disprove", "dispute", "disquisition", "dissect", "dissertation", "egghead", "epistemic", "esemplastic", "examination", "excogitate", "explicate", "expostulate", "extrapolate", "fact", "facultative", "fallacy", "forensic", "forethought", "generalization", "gist", "grasp", "hypothesis", "idea", "ideation", "implausible", "implication", "induction", "inference", "information", "intellect", "intellection", "intellectualize", "interpretation", "irrelevant", "ken", "last word", "logic", "lucid", "middle ground", "moot", "notion", "nub", "perception", "perspicacious", "perspicuous", "pertinent", "polemic", "position", "postulate", "prehension", "premise", "proposition", "proviso", "ratiocination", "rational", "rationale", "reason", "reasoning", "reductionism", "reflection", "refutation", "relation", "riddle", "rumination", "sophistry", "sound reasoning", "specious", "speculation", "standard", "statement", "sticky wicket", "stipulation", "syncretism", "synthesis", "tangent", "tenable", "theorem", "theory", "thesis", "thought", "thought-out", "thrust", "treatise", "unilateral", "viable", "watertight", "well-founded", "well-grounded", "well-taken", "wit"

A.1.0.4 Order, Hierarchy, and Systems Phrases

"agenda", "antecedent", "antinomy", "aspect", "basis", "breakdown", "canon", "case", "case history", "catalog", "category", "catena", "chain", "check", "circular reasoning", "class", "classification", "clause", "codify", "coherence", "cohesion", "collate",

“collatoral”, “collegation”, “columniation”, “comparison”, “compartmentalize”, “compendium”, “component”, “comprise”, “concatination”, “concinnity”, “condition”, “connection”, “consecutive”, “consequence”, “construct”, “content analysis”, “continuity”, “continuum”, “convention”, “corallary”, “correlative”, “correspondance”, “course”, “deliniate”, “design”, “dichotomy”, “division”, “doctrine”, “domino effect”, “element”, “example”, “explicit”, “extrinsic”, “facet”, “feedback”, “flowchart”, “form”, “formula”, “framework”, “full circle”, “fundamental”, “go”, “gradiation”, “grade”, “gradient”, “grid”, “hierarchy”, “hyponym”, “increment”, “index”, “infrastructure”, “instnace”, “interconnectedness”, “interface”, “interlocking”, “interpenetration”, “intrinsic”, “juxtapose”, “key”, “knot”, “labrynth”, “lattice”, “level”, “limit”, “linear”, “list”, “logistics”, “matrix”, “maze”, “method”, “model”, “morphology”, “network”, “nexus”, “nomenclature”, “norm”, “offshoot”, “order”, “ordinnance”, “organization”, “outline”, “paradigm”, “paradox”, “parameter”, “partition”, “patchwork”, “pattern”, “pertinent”, “plan”, “point”, “position”, “precindent”, “premise”, “prerequisite”, “principle”, “probability”, “qualitative”, “quantitative”, “random”, “rank”, “reciprocal”, “reference”, “relative”, “relevant”, “reticulate”, “rudiment”, “salient”, “section”, “sector”, “sequence”, “sequitur”, “set”, “side effect”, “sift”, “skeleton”, “solution”, “spectrum”, “step”, “step by step”, “stratagem”, “stratagy”, “structure”, “structuring”, “subcategory”, “subdivision”, “subset”, “subsidiary”, “succession”, “summary”, “superstructure”, “switchery”, “syndetic”, “synoptic”, “syntax”, “system”, “systemic”, “tactics”, “template”, “test”, “thread of an argument”, “track”, “train of thought”, “tree”, “trusswork”, “unity”, “vicious circle”, “warp and woof”, “web”, “well-ordered”, “workable”, “wrinkle”

A.1.0.5 Judgements and Critiques Phrases

"adjudge", "analysis", "appraise", "arbitration", "argue", "assess", "attitude", "candor", "censor", "chord", "commentary", "conclusion", "consensus", "contention", "criticism", "criticize", "critique", "decide", "decision", "deem", "designate", "discretion", "discriminate", "dispose", "distinguish", "editorial", "estimation", "evaluate", "forejudge", "frame of reference", "free will", "gauge", "guess", "hairsplitting", "inclination", "intercede", "judge", "judgement", "judicious", "misjudge", "misread", "notice", "opine", "opinion", "option", "outlook", "overlook", "partiality", "partisan", "pass", "pass judgement", "peremptory", "point of view", "position", "preconception", "predisposed", "prejudge", "prepossession", "rank", "rate", "reconsider", "referee", "regard", "reservation", "reserve judgement", "review", "second thoughts", "sentence", "sentiment", "settle", "slant", "snap", "sound judgment", "speculate", "suppose", "surmise", "tendency", "think better of", "turn the tables", "umpire", "unbiased", "undercurrant", "vacillate", "value judgement", "vantage point", "verdict", "vet", "view", "viewpoint", "vote", "will", "winnow"

A.1.0.6 Approval, Respect and Recognition Phrases

"accept", "acknowledge", "admire", "appreciate", "approbation", "approval", "approve", "begrudge", "bow", "choose", "cite", "coddle", "command respect", "commemorate", "compassion", "condone", "consideration", "coopt", "countenance", "creedence", "credibility", "credit", "cup of tea", "curtsy", "deference", "deserving", "dignify", "elect", "esteem", "estimation", "exemplary", "fancy", "favor", "favorable", "favoritism", "genuflect", "grant", "gratuity", "high opinion", "honor", "idealize", "inclination", "indulge", "laurels", "like", "love", "obeisance", "oblige", "opt", "pardon", "partiality", "pass muster", "pay respects", "pet", "pick", "pick over", "play

up", "popular", "predilection", "prefer", "preference", "presume", "prize", "proclivity", "propensity", "prostrate", "ratify", "recognition", "red carpet", "regard", "relish", "reputation", "repute", "respect", "reward", "rubber stamp", "salute", "sanction", "satisfaction", "save face", "say-so", "seal of approval", "select", "self pride", "self regard", "self respect", "submit", "superiority", "take kindly to", "take to", "testimonial", "think better of", "thumbs up", "tribute", "untouchable", "valuation", "value", "vote of confidence", "vouchsafe", "warm", "well-disposed", "wow"

A.1.0.7 Support, Encouragement, and Agreement Phrases

"adopt", "advocate", "affirm", "agree", "appoint", "assent", "assist", "avouch", "beneficent", "beneficial", "benefit", "boon", "boost", "booster", "buck up", "buoy", "champion", "choice", "cold comfort", "comfort", "commiserate", "concur", "condolence", "confirm", "congratulate", "console", "corroborate", "embrace", "encourage", "encouragement", "endorse", "espouse", "exculpate", "exhort", "exonerate", "find for", "hold up", "hold with", "humor", "indulgence", "ingratiate", "mollify", "name", "nod", "nominate", "nurture", "ok", "okay", "okey-dokey", "on behalf of", "palliate", "partisan", "pass", "pep talk", "persuade", "placate", "please", "prelation", "preselect", "pro", "promo", "promote", "prop up", "proponent", "protestation", "pump up", "recommend", "reinforce", "relent", "second", "see eye to eye", "side with", "single out", "soothe", "suit", "support", "swallow", "unanimous", "uphold", "validate", "vindicate", "vouch for", "whitewash", "win over", "witness", "yea", "yea-say", "yes-man"

A.1.0.8 Disapproval, Disrespect, and Denial Phrases

"adverse", "animosity", "animus", "antipathy", "askance", "aversion", "avoid", "bias", "brush off", "censorious", "chasten", "cold shoulder", "confound", "contradict", "correct", "correction", "critical", "criticism", "cross out", "debase", "decry", "deflate", "denial", "deny", "deprecate", "detest", "detract from", "dim view", "disapproval", "disapprove", "disavow", "discard", "disclaim", "discommend", "discontent", "discount", "discountenance", "discredit", "discriminate", "disenchantment", "disesteem", "disfavor", "disgruntled", "dishonor", "disillusionment", "dislike", "dismiss", "disown", "displease", "disqualify", "disregard", "disrepute", "disrespect", "dissatisfaction", "distaste", "downgrade", "embarrass", "enmity", "eschew", "fair game", "frown on", "harp on", "hate", "hiss", "hoot", "hypercritical", "ignore", "inconsequential", "inexcusable", "inferiority", "in the doghouse", "judgemental", "knock", "libel", "lose face", "low opinion", "make light of", "nay", "nay-say", "neglect", "one-upmanship", "ostracize", "pass up", "pejorative", "pet peeve", "pharisaism", "pick apart", "pick at", "pick on", "play down", "pot shot", "prejudice", "question", "reduce", "reject", "reprehend", "reproach", "reproof", "reprove", "repudiate", "riot act", "segregate", "shun", "skepticism", "slight", "slough over", "snobbishness", "snooty", "snub", "spank", "stereotype", "stigma", "stigmatize", "stricture", "stuck up", "stultify", "supercilious", "take two task", "talk down", "target", "tease", "thumbs down", "toy", "turn down", "uncomplimentary", "underrated", "undervalue", "undeserving", "unfavorable", "unflattering", "unpopular", "unpromising", "unsatisfactory", "unsung", "unworthy", "veto", "vitiate", "write off", "wrong", "x out"

A.1.0.9 Opposition, Disagreement, and Attack Phrases

“abjure”, “admonish”, “adversary”, “afflict”, “argue”, “assail”, “assault”, “at odds”, “attack”, “ban”, “banish”, “blackball”, “blast”, “boo”, “brand”, “bring to terms”, “buck”, “carp”, “cavil”, “challenge”, “chuck”, “combat”, “complain”, “con”, “confront”, “contest”, “counter”, “cow”, “cross”, “damage”, “daunt”, “debunk”, “decline”, “demur”, “differ”, “disagree”, “discourage”, “dispute”, “dissent”, “dissidence”, “dissuade”, “division”, “exclude”, “excommunicate”, “exile”, “fight”, “find against”, “flak”, “forbid”, “gainsay”, “harrass”, “heckle”, “hostile”, “hostility”, “hound”, “impeach”, “impugn”, “lash out”, “niggle”, “nit pick”, “object”, “obstruct”, “offend”, “offense”, “onslaught”, “opponent”, “oppose”, “opposed”, “oppugn”, “plaint”, “pulemic”, “press”, “protest”, “quibble”, “rebut”, “recriminate”, “redbait”, “renounce”, “reprobation”, “repudiate”, “resist”, “resistance”, “run counter to”, “squelch”, “stimy”, “taboo”, “take issue with”, “take on”, “take sides”, “tear into”, “tee off”, “torment”, “variance”, “withstand”, “write down”, “wrong”

A.2 Contractions, Expansions, and Contrast Phrases

A.2.0.1 Contractions

“ain’t”, “aren’t”, “can’t”, “’cause”, “could’ve”, “couldn’t”, “couldn’t’ve”, “didn’t”, “doesn’t”, “don’t”, “hadn’t”, “hasn’t”, “haven’t”, “he’d”, “he’d’ve”, “he’ll”, “he’s”, “how’d”, “how’d’y”, “how’ll”, “how’s”, “i’d”, “i’d’ve”, “i’ll”, “i’ll’ve”, “i’m”, “i’ve”, “isn’t”, “it’d”, “it’d’ve”, “it’ll”, “it’s”, “let’s”, “ma’am”, “mayn’t”, “might’ve”, “mightn’t”, “mightn’t’ve”, “must’ve”, “mustn’t”, “mustn’t’ve”, “needn’t”, “needn’t’ve”, “oughtn’t”, “oughtn’t’ve”, “shan’t”, “sha’n’t”, “shan’t’ve”, “she’d”, “she’d’ve”, “she’ll”, “she’s”, “should’ve”, “shouldn’t”, “shouldn’t’ve”, “so’s”, “that’d”, “that’d’ve”, “that’s”, “there’d”,

“there’s”, “they’d”, “they’ll”, “they’re”, “they’ve”, “wasn’t”, “we’d”, “we’d’ve”, “we’ll”, “we’re”, “we’ve”, “weren’t”, “what’ll”, “what’re”, “what’s”, “what’ve”, “when’s”, “when’ve”, “where’d”, “where’s”, “where’ve”, “who’ll”, “who’s”, “who’ve”, “why’s”, “why’ve”, “will’ve”, “won’t”, “won’t’ve”, “would’ve”, “wouldn’t”, “wouldn’t’ve”, “y’all”, “you’d”, “you’ll”, “you’re”, “you’ve”

A.2.0.2 Expansions

“am not”, “are not”, “is not”, “has not”, “have not”, “are not”, “am not”, “cannot”, “because”, “could have”, “could not”, “could not have”, “did not”, “does not”, “do not”, “had not”, “has not”, “have not”, “he had”, “he would”, “he would have”, “he shall”, “he will”, “he has”, “he is”, “how did”, “how do you”, “how will”, “how has”, “how is”, “how does”, “i had”, “i would”, “i would have”, “i shall”, “i will”, “i shall have”, “i will have”, “i am”, “i have”, “is not”, “it had”, “it would”, “it would have”, “it shall”, “it will”, “it has”, “it is”, “let us”, “madam”, “may not”, “might have”, “might not”, “might not have”, “must have”, “must not”, “must not have”, “need not”, “need not have”, “ought not”, “ought not have”, “shall not”, “shall not”, “shall not have”, “she had”, “she would”, “she would have”, “she shall”, “she will”, “she has”, “she is”, “should have”, “should not”, “should not have”, “so as”, “so is”, “that would”, “that had”, “that would have”, “that has”, “that is”, “there had”, “there would”, “there has”, “there is”, “they had”, “they would”, “they shall”, “they will”, “they are”, “they have”, “was not”, “we had”, “we would”, “we would have”, “we will”, “we are”, “we have”, “were not”, “what shall”, “what will”, “what are”, “what has”, “what is”, “what have”, “when has”, “when is”, “when have”, “where did”, “where has”, “where is”, “where have”, “who shall”, “who will”, “who has”, “who is”, “who have”, “why has”, “why is”, “why have”, “will have”, “will not”, “will not

have", "would have", "would not", "would not have", "you all", "you had", "you would", "you shall", "you will", "you are", "you have"

A.2.0.3 Contrast Phrases

"although", "besides", "but", "compared with", "conversely", "differ", "even though", "furthermore", "however", "in contrast to", "instead", "less than", "more than", "nevertheless", "notwithstanding", "on the other hand", "otherwise", "rather than", "regardless", "though", "unless", "unlike", "while", "yet"

Table A.1: All Feature Correlations with Success - Raw and Balanced Data

	Imbalanced		Balanced	
Feature	Correlation	p_value	Correlation	p_value
Successful Trigrams	0.686393	0.000000e+00	0.035239	0.149301
Successful Quadgrams	0.667406	0.000000e+00	0.058105	0.017361
Successful Pentagrams	0.598693	0.000000e+00	0.044491	0.068610
Word Menu Pro Con	0.593229	0.000000e+00	0.068917	0.004763
Word Menu Opposition	-0.562532	0.000000e+00	-0.094412	0.000108
Word Menu Support	0.441505	0.000000e+00	0.053114	0.029678
Verb Count	-0.241925	0.000000e+00	-0.064665	0.008094
Sentiment Score	0.227141	0.000000e+00	0.026772	0.273345
Preposition Count	-0.216733	0.000000e+00	-0.094743	0.000103
Noun Count	0.168844	6.367316e-217	0.096769	0.000072
Sentence Count	-0.156517	2.553217e-186	-0.033321	0.172727
Contractions Count	-0.154652	6.659898e-182	-0.069631	0.004345
Word Count	-0.143304	3.286129e-156	-0.017930	0.463224
Word Menu Reasoning	-0.139384	8.264485e-148	-0.055968	0.021943
Expansions Count	-0.138418	8.926644e-146	-0.015781	0.518541
Word Menu Rationale	-0.130848	2.397153e-130	-0.024498	0.316193
Contrast Count	-0.129254	3.264779e-127	-0.006831	0.779887
Determinants Count	-0.121654	8.240112e-113	-0.015501	0.525990
Passive Percent	-0.118330	8.561183e-107	0.001735	0.943418
Smog Score	-0.117359	4.553994e-105	-0.018663	0.445134
Short Sentences	-0.115479	9.080303e-102	-0.048131	0.048824
Person References	0.113791	7.505322e-99	0.044068	0.071289
Question Count	-0.108141	2.074823e-89	-0.062625	0.010335
Cardinal References	-0.105159	1.267811e-84	-0.050560	0.038484
Medium Length Sentences	0.100896	5.110352e-78	0.038874	0.111637
Word Menu Approval	-0.095514	4.489299e-70	-0.032111	0.188861
Max Connections	-0.090474	4.901714e-63	0.005784	0.812954
Very Short Sentences	-0.089054	4.021781e-61	-0.051649	0.034491
Very Long Sentences	-0.087442	5.505863e-59	-0.003258	0.893959
Word Menu Judgements	-0.087303	8.372125e-59	0.007723	0.752062
Date References	-0.071878	2.199389e-40	-0.034107	0.162816
Word Menu Order	-0.070508	6.392217e-39	0.038686	0.113382
GPE References	-0.067128	1.971393e-35	-0.009770	0.689403
Long Sentences	-0.064720	4.742485e-33	0.023120	0.344193
Flesch score	-0.063497	7.124841e-32	-0.087671	0.000326
Law References	-0.055479	1.032037e-24	-0.020748	0.395956
Word Menu Disapproval	-0.053043	9.999522e-23	-0.006073	0.803787
Money References	-0.042877	2.201818e-15	-0.032025	0.190055
Average Connections per Word	-0.038433	1.186375e-12	0.035189	0.149871
Average Sentence Length	-0.038433	1.186375e-12	0.035189	0.149871
Average Connections per Sent.	-0.038433	1.186375e-12	0.035189	0.149871
References to Other Bills	-0.028976	8.449027e-08	-0.022394	0.359545
Quantity References	-0.024076	8.539894e-06	-0.031897	0.191829
Percentage Refereces	-0.022488	3.219430e-05	-0.020570	0.400015
FAC References	0.006166	2.543897e-01	0.014371	0.556590
Ambiguity Score	0.003950	4.653067e-01	-0.038974	0.110715
References to Current Bill	0.002863	5.967034e-01	0.026890	0.271234
Adjectives Count	-0.002802	6.044463e-01	0.044175	0.070606
NORP References	-0.001500	7.815348e-01	-0.009952	0.683922