LEGISLATIVE LANGUAGE FOR SUCCESS

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

Sanjana Gundala

June 2022

COMMITTEE MEMBERSHIP

TITLE: Legislative Language For Success

AUTHOR: Sanjana Gundala

DATE SUBMITTED: June 2022

COMMITTEE CHAIR: Maria Pantoja, Ph.D.

Professor of Computer Science

COMMITTEE MEMBER: Foaad Khosmood, Ph.D.

Professor of Computer Science

COMMITTEE MEMBER: Jonathan Ventura, Ph.D.

Professor of Computer Science

ABSTRACT

Legislative Language For Success

Sanjana Gundala

Legislative committee meetings are an integral part of the lawmaking process for local and state bills. The testimony presented during these meetings is a large factor in the outcome of the proposed bill. This research uses Natural Language Processing and Machine Learning techniques to analyze testimonies from California Legislative committee meetings from 2015-2016 in order to identify what aspects of a testimony makes it successful. A testimony is considered successful if the alignment of the testimony matches the bill outcome (alignment is "For" and the bill passes or alignment is "Against" and the bill fails). The process of finding what makes a testimony successful was accomplished through data filtration, feature extraction, implementation of classification models, and feature analysis. Several features were extracted and tested to find those that had the greatest impact on the bill outcome. The features chosen provided information on the sentence complexity and type of words used (adjective, verb, nouns) for each testimony. Additionally all the testimonies were analyzed to find common phrases used within successful testimonies. Two types of classification models were implemented: ones that used the manually extracted feature as input and ones that used their own feature extraction process. The results from the classification models and feature analysis show that certain aspects within a testimony such as sentence complexity and using specific phrases significantly impact the bill outcome. The most successful models, Support Vector Machine and Multinomial Naive Bayes, achieved an accuracy of 91.79% and 91.22% respectively.

ACKNOWLEDGMENTS

Thank you to my advisor, Dr. Maria Pantoja, for her constant support and guidance throughout my time at Cal Poly, especially during this research, and to my committee members, Dr. Foaad Khosmood and Dr. Jonathan Ventura, for their valuable advice and feedback.

Thank you to the Institute for Advanced Technology and Public Policy as well as the Digital Democracy team for providing the data necessary to conduct this work. Thank you to the Cal Poly Computer Science department for providing me with the resources to accomplish this work.

Lastly, thank you to my friends and family for their constant love and support. Special thank you to my mom for her guidance, encouragement, and always being my biggest supporter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Purpose of the Problem

Legislative meetings are an integral part of the lawmaking process for local and state bills. The testimonies presented during these meetings are a large factor in the outcome of the proposed bill. The purpose of this work is to build upon the Digital Democracy government transparency platform in order to understand what language makes a testimony successful. To remove any external biases such as the status of the speaker or possible affiliations, only testimonies from lobbyist or general public members are analyzed in this research. The language used in these testimonies has a greater impact on the bill outcome compared to testimonies by legislative members or elected officials. This information could be useful to lawyers, citizens, advocacy groups, and other public interest organizations by helping them recognize what elements would improve their testimonies. Vice versa, this information could also aid legislators in removing any sentiment or bias from a testimony, allowing them to focus on the bill itself.

1.2 Background

1.2.1 Language's Impact on Thought

In the mid 1900s, Linguists Edward Sapir and Benjamin Lee Whorf created the idea titled linguistic relativity, which is a principle suggesting that the structure of a language affects its speakers' perception or cognition. Essentially this means the type of language used can affect the way people think and receive information. This is a critical idea in fields that rely heavily on language, such as the legal and legislative sector [25].

There are two famous experiments in the area of linguistic relativity done by Loftus and Palmer in 1974 titled, "Reconstruction of automobile destruction: An example of the interaction between language and memory." In the first experiment, forty-five American students were shown a footage of a traffic accident and the participants were asked to estimate the speed at which the cars were going.

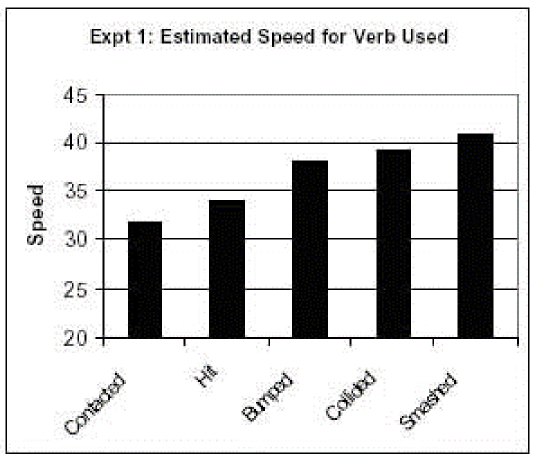

Figure 1.1: Results from Loftus and Palmer's first experiment [30]

As depicted in the figure above, the results showed that changing the words describing a car crash in a video affected the speed estimated by participants although they viewed footage from the same crash. In groups where words like "collided", "smashed", or "hit" were used, the participants reported higher speeds compared to groups who were asked "About how fast were the cars going when they contacted each other?"

In the second experiment, three groups of participants watched footage of car crashes. Two groups were asked a question prior to the footage, one using the word hit and one using the word smash, and the last group was not asked any questions. A week later, all participants were asked if there was a broken glass as a result of the crash in the footage. Those primed with questions had a higher percentage of "yes" answers even though the footage had no glass at all. Although there are external factors which may affect the participants responses, both these experiments showcase that altering the wording does provoke different answers of the same scenario [14].

1.2.2 Forensic Linguistics

Naturally linguistics plays a huge role in the legal field since much of the work in law is dependent on language, either spoken or written. Lawsuits, indictments, pleadings, briefs, legal opinions, testimonies, etc. are all documented and preserved in writing. This relationship between law and linguistics led to a new area of study titled Forensic linguistics [1]. Forensic linguistics involves the application of scientific knowledge to language in the context of criminal and civil law. Linguists have testified and consulted in all types of civil cases including trademark disputes, product liability, discrimination cases, business fraud, and contract disputes.

Forensic linguistics study the language of written law, its complexity and origin, as well as the use of language in forensic procedures. Some forensic linguists focus on features such as punctuation, spelling, vocabulary choices, and grammatical aberrations in the writing, whereas others focus more on syntactic features such as number of words per sentence, sentence length, number of long sentences, total number of sentences, etc. These syntactic features are typically things the speaker or writer is less likely to be consciously aware of [42].

1.2.3 Digital Democracy

Lawmakers in California introduce an average of 5,000 bills during each legislative session, which is a two year time period. However, they do not produce any transcripts or minutes that capture the testimony, debate, and negotiations that occur. The Brown Act was created to increase transparency for local governments, by requiring governmental bodies to provide public notice of their meetings, post agendas of the discussion topics, and provide public access to the meetings [9]. Since the Brown Act applies only to local government, California's state legislators exempt themselves these requirements[2]. The majority of state legislature hearings are audio or video recorded and made available to public through services like The California Channel [49] for a maximum of 3 months. However, since the legislature does not provide transcripts of these discussions, these recordings cannot be searched efficiently. This requires constituents to scan potentially hours of video to find topics of interest. As a result, the news, media, and public have no easy way to find out what happens during these legislative sessions. This lack of transparency prevents civic engagements and accountability.

In 2012, former State Senator Sam Blakeslee founded the Institute for Advanced Technology and Public Policy (IATPP), a nonprofit, bipartisan organization at California Polytechnic State University (Cal Poly) in San Luis Obispo [18]. Three years later, through private donations and student development, the IATPP launched Digital Democracy, an online platform for increasing government transparency and accountability [17]. Digital Democracy transcribes all state legislative committee hearing video recordings and creates a database of search able transcripts available to users. This tool allows users to efficiently search videos by keyword, topic, date, or speaker. These transcripts combined creates a robust database containing speakers, testimonies, speaker's positions, organizations involved, donations, bill outcomes, etc.

In addition to providing access to this information, Digital Democracy also focuses on how this data can be meaningfully interpreted and acted upon, through various projects such as "Predicting the Vote Using Legislative Language" and "Learning Alignments from Legislative Discourse" which are discussed in more detail in Chapter 2. California was the initial focus of Digital Democracy, however this platform was later expanded to Texas, New York, and Florida. For some time, the Digital Democracy platform was the only source of state legislative records for one third of US citizens. Unfortunately this project was discontinued in 2018 due to lack of funding.

This research utilizes the Digital Democracy database to provide valuable insight on the type of language used in legislative meetings and analyzes what aspects of testimony make it successful. With further development, this can become a powerful tool for lobbyists, legislative members, and general public involved in state legislature. Additionally this research applies to both traditional and newer machine learning techniques to a diverse database and evaluates which models are best for text classification.

Chapter 2

RELATED WORKS

2.1 Predicting the Vote Using Legislative Language

Predicting the Vote Using Legislative Language by Cal Poly student, Aditya Budhwar, is one of the major works related to this research and also utilized the Digital Democracy data set [6]. This research analyzed whether verbal utterances made by legislators during the legislative process can indicate their intent on a future vote, and therefore can be used to automatically predict said vote to a significant degree. The statements made by lawmakers are only one factor in determining the vote, however this research questions if those statements alone can be predictive to a significant degree. The authors examined thousands of hours of legislative deliberations from the California state legislature's 2015-2016 session to form models of voting behavior for each legislator and used the models to train classifiers and predict the votes that occur subsequently.

Features such as volume of speech (number of words within the speech), number of speech interruptions, speech sentiment, positive utterance ratio, negative utterance ratio, and question count were extracted from the verbal utterances in the Digital Democracy dataset, prior to ingestion into classifiers. The feature set was tested with several supervised learning algorithms: Support Vector Machines, Random Forests, and Keras with Tensorflow. Various combination of features were tested to find those that produced the highest accuracy for each classifier.

The authors were able to achieve legislator vote prediction as high as 83% and for bill prediction they achieved 76% accuracy both with the Keras and Tensorflow model.

This research follows a similar process to Budhwar's. Feature extraction was done on the Digital Democracy data set, various classification models were trained with those features and the vote outcomes. However, the statements provided in Budhwar's work come from only legislators not lobbyists or general public members. This is because Budhwar's work focuses on predicting legislators votes so it would not be necessary to analyze lobbyists or general public's testimonies. This research focuses primarily on the type of language used by non legislative members in statements rather than other contextual features such as volume of speech [6].

2.2 Learning Alignments from Legislative Discourse

Cal Poly Professor, Daniel Kauffman continued the analysis of legislative statements through his work, titled "Learning Alignments from Legislative Discourse [22]." The goals of this research are to detect any biases legislators have when voting, such as personal opinions or affiliation with an organization, and predict the frequency of agreement between legislators and stakeholders over a collection of proposed laws. This can determine the extent to which a legislator's spoken language indicates their degree of alignment toward an entity (a specific organization or another individual) known for possessing certain views. For this study, bill discussion transcripts from Digital Democracy, organization positions, organization donations, and legislator votes were used along with natural language processing methods to predict alignment scores between each member of the California state legislature and a select set of state-recognized organizations.

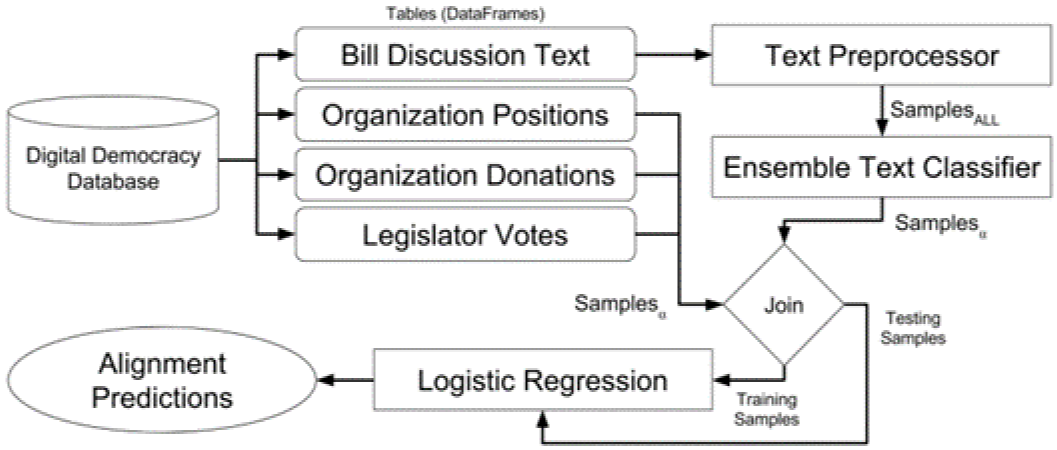

Figure 2.1: System Architecture using Digital Democracy database

The methods implemented achieved up to 78 percent predictive accuracy using a combination of discourse and legislator-related features [22]. This work provides a solid foundation to build upon for this research, since techniques, and machine learning models such as token tagging, n-grams generation, and Naive Bayes are applicable to both [32].

2.3 Understanding the Language of Political Agreement and Disagreement in Legislative Texts

The desire to increase transparency within legislation, has inspired other researchers beyond the Digital Democracy team. Davoodi, Waltenburg, and Goldwasser research the dynamics that lead to adopting national policies using a large-scale dataset they collected [28]. The dataset contains state bills, legislator information, geographical information about legislators' districts, and donations and donors' information. The initial task in this research is to predict the legislative body's vote breakdown for a given bill, according to different criteria such as gender, rural-urban, and ideological splits. The authors utilized a joint graph and text embedding model to represent the nodes and their textual attributes in the legislative graph which is used for the roll-call prediction and aggregation. This system's architecture uses BERT's pretrained embedding to form an initial representation for the textual information of the nodes in the legislative graph and text attributed Relational Graph Convolutional Layers to generate an embedding for the nodes based on their relations. These were then combined to build a representation of edges in the group for relation prediction and aggregated vote relations.

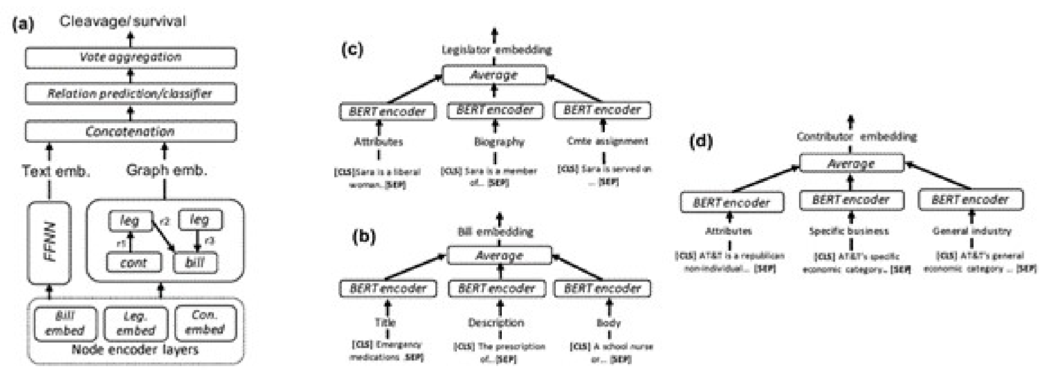

Figure 2.2: System Architecture using BERT and RGCN

The experiments conducted show that using a joint text and graph prediction model and providing the legislative context in which the bill is presented (such as gender, ideology, geography, and party affiliation) helps improve the prediction over strong text-based models, outperforming each of the models (BERT and RGCN) in isolation [28].

2.4 Get out the Vote

"Get out the Vote: Determining support or opposition from Congressional floor-debate transcripts" by authors Thomas, Pang, and Lee investigates whether transcripts of U.S Congressional floor debates can be used to determine the alignment (support or opposition) of speeches for a proposed piece of legislation [44]. In this context, a speech is a continuous single-speaker segment of text, similar to this research. The U.S floor debate transcripts were extracted from GovTrack, an independent website that collects public available data on legislative activities of U.S congress, for the year 2005.

The 3268 pages of transcripts contained voting records for all roll-call votes during that year.The authors focused on debates for "controversial" bills, meaning the losing side of the bill gave at least 20% of the speeches. Each debate contains a series of speech segments. These segments were labeled with the vote cast (aye or nay) by the speaker of the segment for that proposed bill. One-sentence utterances were disregarded, since they typically were pertaining to issues not related to the bill, such as yielding time.

The data was randomly split into training, testing, and development sets representing about 70%, 20%, and 10% of the data, with 38 debates in the training set, 10 in the test set, and 5 in the development set. For the evaluation process, the authors created 3 models: a baseline model, a SVM light classifier with only speech segments, and a SVM classifier with speech segments and agreements. In House discourse, it is common for one speaker to make reference to another, either agreeing or disagreeing with the topic in question. To incorporate this aspect in the SVM classifiers, the authors classified each reference connecting two speakers with a positive or negative label depending on whether the two voted the same way on the bill being discussed. These labels were then used to train the SVM classifier.

The baseline model achieved an accuracy of 59%, the first SVM light classifier achieved an accuracy of 70%, and the second SVM light classifier with agreements achieved an accuracy of 89%. These results show that including even limited information about segment relationships can significant improve the models ability in predicting alignment: support or opposition [44].

2.5 BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Bidirectional Encoder Representations from Transformers, also known as BERT, is a transformer-based machine learning technique for natural language processing pre-training developed by Google [19]. Since 2019, Google has been utilizing BERT to better understand user searches. Transformer is a deep learning model and attention mechanism that learns contextual relations between words in a text [48]. BERT's goal is to a generate a language model with Transformer's encoder mechanism to read text input. Unlike directional models which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once which is why it is considered bidirectional. This allows the model to learn the context of a word based on all its surroundings. As a result, BERT can be fine-tuned with one additional output layer to create state-of-the-art models for various tasks, such as question answering, sentiment analysis, sentence prediction, and more.

The input representation for BERT represents both a single sentence and a pair of sentences in one token sequence. A "sentence" is an arbitrary length of continuous text and does not have to be a grammatically correct sentence whereas a "sequence" can be either a single sentence, or two sentences put together. If sentence pairs are put together into a single sequence, the sentences are differentiated in two ways: by separating them with a special token [SEP] or by adding a learned embedding to every token to indicate whether it belongs to sentence A or sentence B. Fig 2.4.1 shows an example of BERT learning to model relationships between sentences. Given two sentences A and B, BERT determines if B is the next sentence that comes after A in the corpus, or just a random sentence

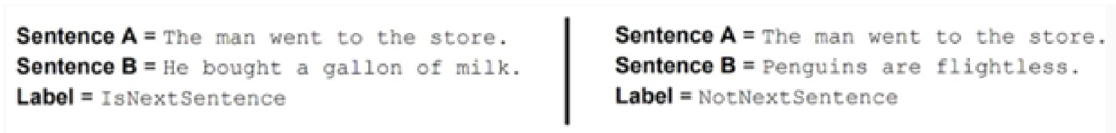

Figure 2.3: BERT learning to model relationships

For a given token, the input representation is generated by combining the corresponding token, segment, and position embeddings as shown in Fig 2.4.2 below. The input embeddings are the sum of the token embeddings, the segmentation embed-dings, and the position embeddings

Figure 2.4: BERT input representation

BERT's model architecture is a multi-layer bidirectional Transformer encoder with initially only two model sizes: BERTBASE (L=12, H=768, A=12, Total Parameters=110M) and BERTLARGE (L=24, H=1024, A=16, Total Parameters=340M). Here L is the number of layers also known as Transformer blocks, H is the hidden size, and A represents the number of self-attention heads. Now, multiple BERT models are available with different variations of these parameters which makes BERT applicable to various types of data. BERT models are pre-trained from unlabeled data extracted from the BooksCorpus with 800M words and English Wikipedia with 2,500M words. BERT consists of two steps: pre-training and fine-tuning. BERT utilizes two unsupervised pre-training tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP) to improve the fine-tuning-based approach.

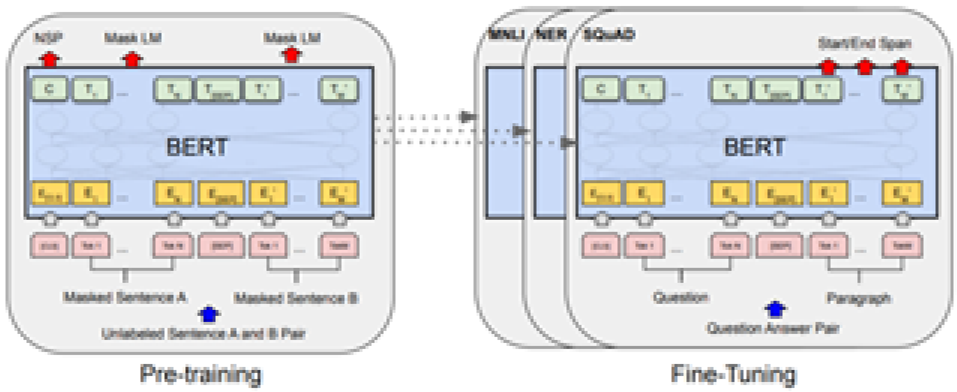

Figure 2.5: Overall pre-training and fine-tuning procedures for BERT

These tasks in combination with pre-training on Transformer allows BERT to better understand the context of a word based on its surroundings and sentence relationships. Unlike directional language model pre-training, the MLM objective allows the input representation to combine the left and right context, which is used to pre-train a deep bidirectional Transformer. During pre-training, the model is trained on unlabeled data over different pre-training tasks. During fine tuning, the BERT model is initialized with the pre-trained parameters and then all of the parameters are fine-tuned using labeled data from the downstream tasks [19].

Chapter 3

SOFTWARE TOOLS

3.1 NLTK

Natural Language Toolkit (NLTK) is a platform for building Python programs that work with language data. NLTK provides a simple interface and has over 50 corpora and lexical resources for text processing, tokenization, parsing, tagging, and semantic reasoning. In this research NLTK was used for text processing tasks, such as N-gram generation and Part of Speech (POS) tagging [24].

3.2 spaCy

spaCy is an open-source software library for natural language processing written in Python. spaCy features convolutional neural network models for part of speech tagging, dependency parsing, text categorization, named entity recognition (NER) and natural language processing. In this research spaCy's English corpus was used to first identify the proper nouns in text [15].

3.3 Sklearn

Sklearn is a Python module containing numerous Python packages for classic machine learning algorithms such as numpy, matplotlib, and preprocessing. In this research sklearn's preprocessing package was used to encode features prior to ingestion into the model. Additionally, sklearn's Naive bayes and SVC packages were used to implement the Naïve Bayes model and Support Vector Model. Lastly sklearn's metrics package was used to provide accuracy, precision, and recall values for each model [34].

3.4 Tensorflow

Tensorflow is an open-source software library for machine learning and artificial intelligence developed by the Google Brain team. It can be used for a wide range of tasks but is particularly used for training and inference of deep neural networks. In this research Tensorflow was used to implement BERT as well as a fully connected neural network [27].

3.5 Keras

Keras is a deep learning API written in Python on top of TensorFlow developed with a focus on enabling fast experimentation. In this research Keras was used to implement a fully connected neural network using a Sequential model, which is a linear stack of layers [8].

3.6 Pandas

Pandas is a fast and flexible open-source data analysis and manipulation tool, built on top of Python. In this research Pandas was used to read in data from CSV files and store the data in a DataFrame objects. Storing data as DataFrames allowed it to be easily filtered, transformed, reshaped, and indexed. This was particularly useful for filtering duplicates and unanimous bills from the dataset [29].

3.7 Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. In this research Matplotlib was utilized to do feature analysis, specifically to generate a correlation matrix between features [16].

3.8 Google Colab

Google colab is a free Juptyer notebook environment that runs entirely in the cloud. Colab allows anyone to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. CoLab allows for easily integration with PyTorch, TensorFlow, Keras, and OpenCV. For this research Google CoLab was used to implement each BERT model and generate a correlation matrix between features [21].

Chapter 4

EXPERIMENTAL DESIGN

4.1 System Overview

This system consists of several components and interconnected processes. Fig. 4.1.1 below clearly outlines the system design.

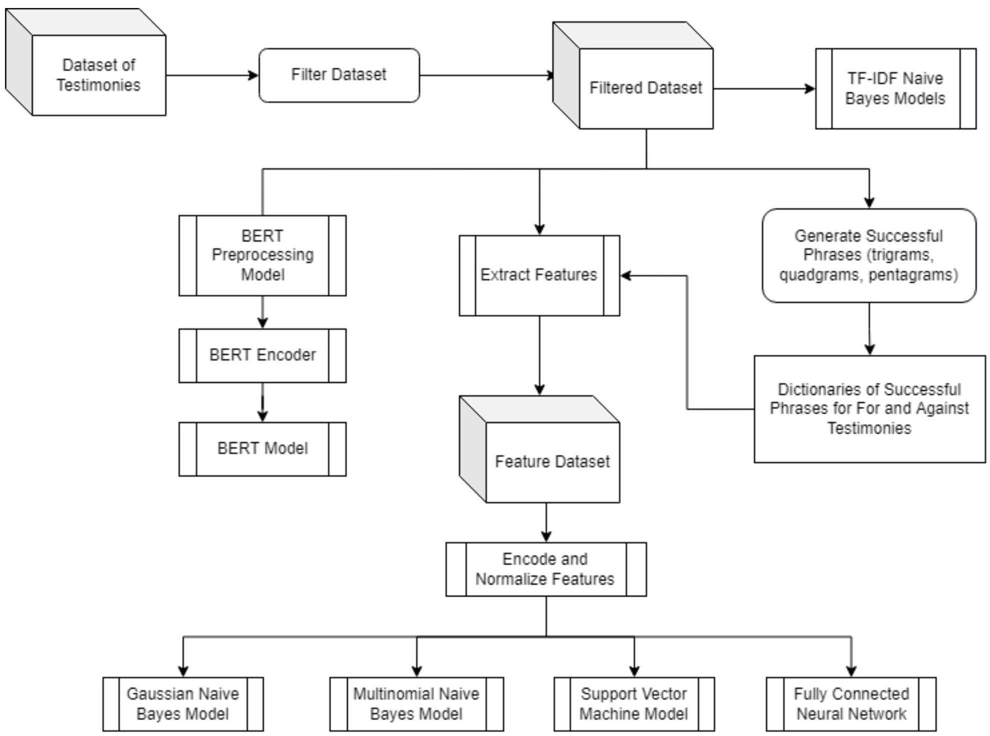

Figure 4.1: System Overview

The initial dataset is filtered and a filtered dataset is outputted. This dataset is used as input for both the BERT models and the TF-IDF Naive Bayes models. For the remaining models, the filtered dataset is then used to extract features. A side step in this process is generating the successful phrases. This step iterates through all testimonies three times to generate six dictionaries: three-word phrases (trigrams) that appear in testimonies with “For” alignments, three-word phrases that appear in testimonies with “Against” alignments, four-word phrases (quadgrams) that appear in testimonies with “For” alignments, four-word phrases that appear in testimonies with “Against” alignments, five-word phrases (pentagrams) that appear in testimonies with “For” alignments, and five-word phrases that appear in testimonies with “Against” alignments.

Although the system overview can be interpreted as this step occurring simultaneously to the feature extraction, it actually occurs beforehand. All six dictionaries and the filtered dataset are required as input for feature extraction. After feature extraction is completed a features dataset is outputted. This data is then encoded and normalized prior to ingestion into the four classification models: Gaussian Naive Bayes, Multinomial Naive Bayes, Support Vector Machine, and Fully Connected Neural Network.

4.2 Finalizing the Dataset

When initially approaching this problem, the primary focus was to parse the JSON transcripts and generate a basic dataset, which was accomplished. However, after a few months of data parsing and development, a more detailed dataset was created via the Digital Democracy research team [17]. For this new dataset, an automated transcript was created for each legislative meeting recording. From there the research team hand labelled the speaker, timestamps, and alignments for each utterance. The automated transcripts were also reviewed to ensure the correct content was stored in the transcript.

Since features in the new dataset were hand tagged, specifically Alignment and Speech Type, it proved to be more accurate than the initial dataset. With the older dataset, time stamps were used to identify longer speeches and these speeches were assumed to be testimonies, the newer dataset removes this assumption. Additionally the newer dataset contained IDs for each bill, discussion, vote, and person, which made filtering the dataset more efficient. For the purpose of accuracy and scalability this newer dataset was utilized over the manually generated one.

4.3 Initial Dataset Generation

In order to gather the necessary information from the JSON transcripts, large testimonies were extracted based on time stamps and any non-unanimous bills were filtered out. Each legislative meeting was translated and stored in a JSON file. The JSON files contained "hearing transcript" and "voting result" sections, which is the testimony information and voting portion of the bill discussion. The figures below show snippets of one of the JSON files.


```
hearing. </div><div class = 'note'>[7] The mentions of other bills are recognized automatically from the t
hearing. </div></div><div class='endnotes_text'>All Rights Reserved (c). AI4Reporters, 2020.</div></div>",
    "endnotes_text": "All Rights Reserved (c). AI4Reporters, 2020.",
    "video_duration": 488,
    "leginfo_bill_link": "http://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB
    "hearing_transcript": [
        {
            "start_time": 1392,
            "end_time": 1395,
            "vid_file_id": "ce5e8461be27c9edfef9df6f196f7024",
            "speaker_pid": 107,
            "speaker_first": "Ricardo",
            "speaker_last": "Lara",
            "utterance": "I see. The floor is yours.",
            "section_label": "bill_intro"
        },
        {
            "start_time": 1395,
            "end_time": 1424,
            "vid_file_id": "ce5e8461be27c9edfef9df6f196f7024",
            "speaker_pid": 87,
            "speaker_first": "Reginald",
            "speaker_last": "Jones-Sawyer",
            "utterance": "The best for last. Thank you, Mr. Chair. Senate amendments do two things. First,
Governor with the authority to negotiate and enter into agreements on behalf of the state with the Interna
Committee and the International Paralympic Committee. Second, it provides financial assistance that is req
Olympic Games.",
            "section_label": "bill_intro"
        },
```


Figure 4.2: Hearing Transcript Section of JSON file

The "voting result" sections were parsed to identify the outcome of the bill. N-grams were generated on the utterance containing the voting outcome and these n-grams were then filtered to extract any containing the key words "favor", "against" or "abstained." For each bill the voting section follows the same uniform structure, the number of individuals in favor is always in the beginning of the statement, followed by number against, and lastly number of abstentions. The n-grams containing the key words were then ran through Word2Vec's word_to_num module which converts written numbers into their numerical form. This process extracted the number of ayes, naes, and abstains for each bill presented. However, if this technique were to be implemented for other transcripts which may not follow the same structure, values would have to be extracted based on context words.

Each entry in the "hearing transcript" section of the JSON file contains the speaker's first name, speaker's last name, utterance, start time, and end time. The filter developed iterates through each utterance in each transcript and populates a new JSON dataset. As the utterances are iterated through, the previous speaker, current speaker, current utterance, "current speech", and "current speech time" are stored. The "current speech" is formed as each new utterance is added to it, if the current speaker is the same as the previous speaker. Once the speaker changes "current speech time" is checked to see if it exceeds a given time threshold. If it surpasses the threshold, the current speech is included as a new entry in the JSON dataset. This JSON dataset is later read from in order to do feature extraction.

```
[
    {
        "Speaker": "RichardBloom",
        "Time": 136,
        "Speech": "Thank you Mr. Chair and Members, I would like to begin by accepting the Committee amendments that clarify t
        "Bill Outcome": "Passed",
        "FileName": "CA_201720180AB2528_2_254716_111075.json",
        "Sentiment": "SUPPORT",
        "POS": "NNP PRP NNP NNP CC NNP PRP MD VB TO VB IN VBG DT NNP NNS WDT VBP DT NN IN NN NN NN DT NN VBZ TO VB NNP NNS RB
    },
    {
        "Speaker": "JoseGonzalez",
        "Time": 191,
        "Speech": "Hello again, Mr. Chair and Members of the Committee. Again, my name's Jose Gonzales. I'm the founder of Lat
        "Bill Outcome": "Passed",
        "FileName": "CA_201720180AB2534_3330_254610_110269.json",
        "Sentiment": "SUPPORT",
        "POS": "NNP VBZ NNP NNP CC NNP IN DT NNP NNP PRP$ NN NNP NNP NNP DT NN IN NNP NNP IN PRP VBD NN NN DT JJ NNP CC JJ JJ
    },
    {
        "Speaker": "JoseMedina",
        "Time": 254,
        "Speech": "I present Assembly Bill 2772. A bill that will add a course of ethnic studies to be a high school graduatio
        "Bill Outcome": "Passed",
        "FileName": "CA_201720180AB2772_81_254161_108754.json",
        "Sentiment": "SUPPORT",
        "POS": "PRP VBP RB NNP CD NNP NN WDT MD VB DT NN IN JJ NNS TO VB DT JJ NN NN NN VBG DT NN NN CD CC IN VBG IN DT NN PRP
    },
    {
        "Speaker": "MargaritaBerta-Avila",
        "Time": 133,
        "Speech": "Okay, thank you. Good afternoon to the Members of the Assembly Education Committee, my name is Dr. Margarit
        "Bill Outcome": "Passed",
        "FileName": "CA_201720180AB2772_81_254161_108754.json",
        "Sentiment": "SUPPORT",
        "POS": "NNP NN RB JJ NN TO DT NNS IN DT NNP NNP NNP PRP$ NN VBZ NNP NNP NNP PRP VBP RB IN NN IN JJ NNS RB NNP NNP RB
    },
```


Figure 4.3: Excerpt of Initial Dataset Generated

4.4 Current Dataset

The current dataset contains 18 different fields listed below. For the purpose of this research the following fields were utilized: Person Type, Organization, Alignment, Ayes, Naes, and Text. The "text" field consists of each utterance spoken during the legislative hearing. An utterance is a single comment or sentence. Using the Person Id, Bill Id, and Discussion Id, these utterances were combined to extract the full testimony presented by each speaker. These testimonies were then used to generate more features including readability scores and successful phrases.

Field Name	Description
First	First Name of Speaker
Last	Last Name of Speaker
Person Id	Unique Id for each person
Person Type	Type of Person Speaking (Lobbyist, Legislative Member, General Public)
Organization	Any organization the speaker is affiliated with
Bill Id	Unique Id for each bill
Discussion Id	Unique Id for each bill discussed at a specific date and committee
Speech Type	Purpose of the speech
Alignment	For, Against, Neutral
Date	Date of Testimony
Vote Id	Unique Id for the vote (Helps distinguish votes if multiple have occurred in one discussion)
Ayes	Number of votes in favor of the bill question
Naes	Number of votes against the bill in question
Abstain	Number of votes abstaining for the bill in question
Motion Id	Unique Id of the motion
Motion Text	Actual motion to be voted on
doPass	If the bill has passed or not
Text	Utterance within testimony

Table 4.1: List of fields in current dataset

4.5 Data Extraction

This dataset was store in a .tsv file which was processed and stored in a dataframe using Pandas [29] read_csv() module. In order to avoid overfitting and extract accurate data, it was important to then filter this dataframe accordingly. This included removing any duplicate entries of utterances and any unanimous bills prior to feature extraction. Duplicate testimonies were removed using Pandas dropduplicate() module and using the Person Id, Bill Id, and Disucssion Id from the new dataset. Unanimous bills were filtered based on the "Ayes" and "Naes" columns of the dataframe. Any entry containing 0 "Ayes" or 0 "Naes" was removed from the dataframe. This eliminates any bias bills which may have passed or failed solely on the context of the bill and regardless of the language used in the testimony.

4.6 Feature Extraction

4.6.1 Alignment

Initially for each testimony, a simple sentiment analysis was done to determine the tone: in support, against, or neutral. A small sample set of words which indicate support and a set of words which indicate opposition were created and then compared against the speech to identify if the speech contained more support or opposition words. If there was no majority of support or opposition words, the speech was declared as neutral. However, this sentiment analysis was later removed when the new dataset was introduced since it was not as accurate as the hand-tagged alignment labels in the newer dataset.

The Alignment field indicates if the testimony being presented was in favor or against the bill being presented. The possible values for Alignment include “For”, “Against”, “Indeterminate”, “For_if_amend”, and “Against_unless_amend”. the testimony does not contain strong language for or against, the alignment is labelled Indeterminate. As mentioned previously, the labels for alignment were hand tagged by members of the Digital Democracy research group.

Since hand tagged labels can be error prone, a sample of 64 testimonies, 10% of the smaller dataset after filtering, was generated to test the accuracy of the alignments. The given hand tagged alignments were hidden and the alignment for each testimony was redetermined, as shown in Table 4.6.1 below.

Disparity in Alignments?	Confirmed Alignment	HandTagged Alignment	Text
1	For	For	Ed Berens on behalf of the Los Angeles County Board of Supervisors in support.
1	For	For	Good afternoon, Usha Mutschler on behalf of the California State Sheriffs Association in support.
1	For	For	Bruce Palmer, Jewish Community Relations Council of Sacramento, in strong support
1	For_if_amend	For_if_amend	Thank you. Bill. Mr. Chairman and Members, Bill Magavern at the Coalition for Clean Air. We appreciate Senator Beall's commitment to public transit and to active transportation., But we support the amendments that have been recommended by the Bike Coalition and the Lung Association. And we also suggest that the bill be amended so that the Trade Corridor funding be contingent on adherence to the guiding principles of the Sustainable Freight Action Plan, which was published by the administration last year. Thank you
1	For	For	My name is Brian Hamlin, I'm the co-executive Director of the California Renters Legal Advocacy and Education Fund., We're a non-profit that works to make housing more affordable and accessible for all Californians. And I enthusiastically support SB 35. Thank you.>>Good., Okay.>>We have a number of tweeners here., Brian Augusta on behalf of the California Rural Legal Assistance Foundation. We, too, do not yet have a position on it. We have some concerns with the bill which we've shared with the author. And the author and his staff have been working carefully with us to address those concerns., So we just wanted to note that for the record.
0	Indeterminate	Against	Mr. Chairman, Members of the Committee, James Lumbar on behalf of the California Motorcycle Dealers Association. We've been involved in every reauthorization since May in 87 and we would hope to be involved in this one as well, because we're the ones that sell the motorcycles to the users and we always would request that they drive responsibly. Thanks.

Table 4.2: Alignment verification for random sample set

Table 4.6.1 shows each testimony with the hand tagged alignment, confirmed alignment, and if there was any disparity between these two labels. No difference is signified by a 1, whereas a difference is signified by a 0. The average of the disparity column was 92.06% which means the alignments in the given data set are 92% accurate.

4.6.2 Readability Scores

Readability tests designed to indicate how difficult a passage in English is to understand. For testimonies these reliability tests can provide a quantitative value for the complexity of each statement. Four basic readability tests were extracted for each testimony: Flesch reading-ease, Gunning Fog Index, Smog Index, and Dale-Chall readability score. For each readability test the output calculated signifies a different attribute of the text [41].

4.6.2.1 Flesch Reading-Ease

In the Flesch reading-ease test higher scores indicate the testimony is easier to read whereas lower numbers indicate the testimony is more difficult to read[35]. A score of 100.00-90.0 would indicate a 5th grade reading level, whereas 10.0-0.0 indicates an above college graduate reading level which is extremely difficult to read. The Flesch reading-ease scores are calculated for each testimony using the formula below:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

Figure 4.4: Flesch Readability Score Formula

4.6.2.2 Gunning Fog Index

For Gunning Fog Index, the ideal score for readability is 7 or 8 [36]. Anything above 12 is too hard for most people to read, for example The Bible, Shakespeare, and Mark Twain have Gunning Fox Indexes around 6 and leading magazines, like Time, Newsweek, and the Wall Street Journal average around 11. The Gunning Fog Index can be calculated using the formula below:

Gunning Fog Index = 0.4 (ASL + PHW)

where ASL = Average Sentence Length and PHW = Percentage of Hard Words. "Hard words" are counted as words with three or more syllables that are not proper nouns, combinations of easy words or hyphenated words, 3-syllable verbs ending in -es and -ed.

4.6.2.3 Dale-Chall Score

The Dale-Chall Formula calculates the US grade level of a text sample based on sentence length and the number of 'hard' words [37]. The Dale-Chall Formula is unlike other formulas that use word-length to assess word difficulty. 'Hard' words are defined as words that do not appear on a specially designed list of common words familiar to most 4th-grade students.

Raw Score = 0.1579 * (PDW) + 0.0496 *ASL

where PDW = percentage of difficult words and ASL = Average Sentence Length in words. If PDW is greater than 5% then: Adjusted Score = Raw Score + 3.6365, otherwise Adjusted Score = Raw Score.

4.6.2.4 SMOG Grade

Smog stands for "Simple Measure of Gobbledygook." SMOG estimates the years of education the average person needs to understand any piece of writing, also known as the SMOG Grade [38]. For testimonies with less than 3 sentences, the default SMOG score is 0. For testimonies with less than 30 sentences the SMOG Grade is calculated using the following formula, where poly_syllab is the number of words containing more than 2 syllables, avg_poly_syllab is the average number of polysyllable words in a sentence, and num_sentences is the total number of sentences in the testimony:

SMOG = 3 + (sqrt (poly_syllab + (avg_poly_syllab * (30 - num_sentences)))

For testimonies with more than 30 sentences the SMOG Grade is calculated using the following formula, using the same variables as above:

SMOG = 1.0430 * (sqrt (30 * poly_syllab / num_sentences)) + 3.1291

The output of SMOG correlates to a grade level, therefore, a testimony with a higher SMOG grade signifies it requires a high-level education to understand.

Each readability score was tested but only flesch and smog score were chosen as features. The Gunning Fog Index Readability Formula is more ideal for education material such as business magazines and journals so it was not chosen. Dale-Chall and Flesch readability scores are both suitable for all kind of texts, however Flesch scores are typically used for military and government agencies which is why it was a better fit for this research. Lastly SMOG Readability Formula was chosen since it is typically used for text aimed at higher level readers [41].

4.6.3 Person Type

One of the features given in the dataset that was utilized for each classification model was Person Type. This feature refers to the type of person giving the testimony: "Lobbyist" or "General Public." According to California's Fair Political Practices Commission [10], a lobbyist is an individual who is compensated to communicate directly with any state, legislative or agency official to influence legislative or administrative action on behalf of his or her employer or client. This feature provides more context to the testimony and helps determine if the type of person speaking has an effect on the bill outcome.

4.6.4 Successful Phrases

For each testimony, n-grams were generated using NLTK's n-gram module. In order to generalize the ngrams, spaCy's 'en_core_web_sm' corpus was used to first identify the proper nouns in the text. These proper nouns were then replaced with their respective entities, such as name, place, or organization. For example the statement "with the Howard Jarvis Tax Preparer's Association in opposition to" is generalized to "with ORG in opposition to."

Once a new text was created with entities in place of proper nouns, the new text was tokenized and used generate large n-grams of about six or seven words. These n-grams are essentially generalized phrases in each testimony. The n-grams were then iterated through and placed in a dictionary to count how many times each phrase appeared. However, this showed there was not much commonality in large phrases between testimonies. This process was then redone with smaller n-grams of three, four, and five words which occurred much more frequently within testimonies.

In order to further generalize the phrases for ingestion into the classification models, the n-grams were filtered to find "successful phrases." First the bill outcome and alignment of the testimony are extracted from the filtered dataset. After replacing proper nouns with their respective entities, each n-gram was stored in two of four dictionaries to see how many times it aligns with the bill outcome. The ngram aligns with the bill, if the bill passed and the testimony alignment was "For" or "Indeterminate" or if the bill failed and the alignment was "Against" for "Indeterminate."

Each dictionary contains the phrase as the key, but the values varies. In the first dictionary the value is the number of times the phrase appears in testimonies with an alignment of “For” or “Indeterminate.” In the second dictionary, the value is the number of times the phrase appears in testimonies with an alignment of “For” or “Indeterminate,” and the bill passes. In the third dictionary the value is the number of times the phrase appears in testimonies with an alignment of “Against” or “Indeterminate,” and in the last dictionary the value is the number of times the phrase appears in testimonies with an alignment of “Against” or “Indeterminate” and the bill fails.

These dictionaries were then used to calculate the success rate of the phrase. First any phrase that does not appear in more than 10 testimonies is filtered out. The positive success rate is the calculated as the (the number of times the phrase alignment is “For” or “Indeterminate” and the bill passes) / (number of “For” or “Indeterminate” testimonies the phrase appears in). Likewise, the negative success rate is the calculated as the (the number of times the phrase alignment is “Against” or “Indeterminate” and the bill fails) / (number of “Against” or “Indeterminate” testimonies the phrase appears in). These percentages are stored in two separate dictionaries which are then filtered to only store phrases with a success rate over 49%. This process results in two lists of “positive successful phrases” and “negative successful phrases” which are successful phrases in testimonies with an alignment of ”For” and successful phrases in testimonies with an alignment of ”Against” respectively. Example five word phrases are shown below, along with the success rate of the phrase. The full list of 3 word phrases, 4 word phrases, and 5 word phrases for each alignment type can be found in the Appendix A.

5 word Phrase (Pentagram)	Success Rate
('have', 'an', 'official', 'position', 'on')	0.8
('madame', 'chair', 'and', 'members', 'of')	0.5
('of', 'the', 'org', 'for', 'the')	0.5
('forward', 'to', 'continuing', 'to', 'work')	0.833333333
('aye', 'vote', 'on', 'this', 'bill')	0.5
('person', 'for', 'his', 'leadership', 'on')	0.5
('that', 'this', 'is', 'going', 'to')	1
('continuing', 'to', 'work', 'with', 'the')	0.666666667
('id', 'like', 'to', 'thank', 'the')	0.666666667
('with', 'org', 'in', 'opposition', 'thank')	1
('live', 'in', 'gpe', 'and', 'i')	0.857142857
('the', 'intent', 'of', 'the', 'bill')	0.5
('on', 'behalf', 'of', 'the', 'urban')	0.5
('we', 'ask', 'for', 'your', 'no')	1
('do', 'want', 'to', 'thank', 'the')	0.5
('for', 'your', 'no', 'vote', 'on')	1
('we', 'do', 'have', 'some', 'concerns')	0.875
('org', 'in', 'opposition', 'thank', 'you')	0.5
('you', 'to', 'oppose', 'this', 'bill')	1

Table 4.3: Excerpt of successful 5 word phrases for testimonies with "For" or "Indeterminate" alignment

5 word phrase (pentagram)	Success Rate
('to', 'move', 'america', 'and', 'cardinal')	1
('the', 'comments', 'of', 'my', 'colleagues')	0.5
('with', 'org', 'org', 'of', 'gpe')	1
('more', 'time', 'to', 'read', 'thank')	1
('to', 'oppose', 'along', 'with', 'org')	1
('on', 'behalf', 'of', 'org', 'california')	0.5
('councils', 'org', 'fair', 'rents', 'for')	1
('with', 'org', 'county', 'and', 'municipal')	0.5
('gpe', 'org', 'of', 'gpe', 'hi')	1
('org', 'org', 'beyond', 'the', 'arc')	1
('with', 'org', 'here', 'date', 'on')	0.5
('partnership', 'for', 'justice', 'org', 'coalition')	1
('public', 'advocates', 'here', 'to', 'oppose')	1
('community', 'technologies', 'community', 'health', 'councils')	1
('senators', 'person', 'on', 'behalf', 'of')	0.666666667
('to', 'take', 'up', 'more', 'time')	1
('thank', 'you', 'org', 'esperanza', 'community')	1
('person', 'with', 'public', 'advocates', 'here')	1
('oppose', 'along', 'with', 'org', 'org')	1

Table 4.4: Excerpt of successful 5 word phrases for testimonies with "Against" or "Indeterminate" alignment

Table 4.6.2 indicates that half the time the phrase "madame chair and members of" appears in a testimony with a "For" or "Indeterminate" alignment the bill passes since the success rate is 50%. Similarly, Table 4.6.3 indicates that every time the general phrase "PERSON with public advocates here" appears in a testimony with an "Against"or "Indeterminate" alignment, the bill fails since the success rate is 100%.

After these lists are generated, the testimonies are iterated though again. For each testimony, if the alignment is "For" then the "positive successful phrases" list is cross referenced but if the alignment is "Against" the "negative successful phrases" list is cross referenced. The number of either positive or negative successful phases within the testimony is then used as a feature for the classification models. These positive and negative successful phrases were created for trigrams, quadgrams, and pentagrams.

4.6.5 Successful "Part of Speech" Phrases

The part of speech for each testimony was initially extracted to evaluate if the sentence structure of testimonies made an impact on the outcome of bill. However, after looking into the number of occurrences of each POS tagged sentence it was found that similar to larger n-grams there was not much commonality between POS tagged sentences. Therefore, this feature was then altered to be the number of successful POS tagged phrases in a testimony. Generating a list of successful POS n-grams was the same process as generating a list of successful phrases discussed previously. Instead of replacing proper nouns in the testimony, NLTK's pos_tag module [24] was used to replace each sentence in the testimony with tagged sentences. These new tagged sentences were then used to generate POS n-grams. Once again, the list of successful POS n-grams was cross referenced to find the number of successful POS n-grams in each testimony and this number was used for ingestion into classification models. However, this feature produced the same results as the "Successful Phrases" feature and not provide any additional information. Therefor,this feature was later removed and changed into proportions of parts of speech, since that provided better insight into the type of words used in testimonies.

4.6.6 Proportion of Parts of Speech

Proportions of Parts of Speech provides insight into what type of language is used in testimonies. For example, a descriptive testimony would contain more adjectives. These features were extracted by tagging the Part of Speech (POS) of each word within each testimony. The POS tagging was done using NTLK's pos_tag module [24] to generate POS tagged sentences and the Counter module [11] to count the occurrences of each tag within the testimony. Below is a full list of possible POS tags, however for the purpose of this research the POS tags were combined into broader categories: DT, JJ, NN, VB and PRP.

CC	Coordinating conjunction
CD	Cardinal digit
DT	Determiner
EX	Existential there (ex: "there is")
FW	Foreign word
IN	Preposition subordinating conjunction
JJ	Adjective (ex: "big")
JJR	Adjective, comparative (ex: "bigger")
JJS	Adjective, superlative (ex: "biggest")
LS	List marker (ex: 1.)
MD	Modal (ex: "could", "will")
NN	Singular noun
NNS	Plural noun
NNP	Singular proper noun
NNPS	Plural proper noun
PDT	Predeterminer (ex: "all the kids")
POS	Possessive ending
PRP	Personal Pronoun (ex: I, he, she)
PRP$	Possessive Pronoun (ex: my, his, hers)
RB	Adverb (ex: "very", "silently")
RBR	Adverb, comparative (ex: "better")
RBS	Adverb, superlative (ex: "best")
RP	Particle (ex: "give up")
TO	To (ex "to go to the store")
UH	Interjection (ex: "uhmmm")
VB	Verb, base form (ex: "take")
VBD	Verb, past tense (ex: "took")
VBG	Verb, gerund/present participle (ex: "taking")
VBN	Verb, past participle (ex: "taken")
VBP	Verb, present, non-3rd person, singular (ex: "take")
VBZ	Verb, present, 3rd person, singular (ex: "takes")
WDT	Wh-dterminer (ex: "which")
WP	Wh-pronoun (ex: "who", "what")
WP$	Possessive wh-pronoun (ex: "whose")
WRB	Wh-adverb (ex: "where", "when")

Table 4.5: Part of Speech Acronyms and Descriptions [46]

For example, the occurrences of NN, NNS, NNP, and NNPS were added to find the total number of nouns. This was done with adjectives and proper nouns as well. Once again, the Counter module was used to find the proportions of each category within the testimony. The proportions for DT, JJ, NN, VB, and PRP were then ingested as features into the classification models.

4.6.7 Dependency Parsing

Dependency parsing is a process used to analyze the grammatical structure of a sentence and identify the relationship between words in the sentence [20]. Spacy's [15] dependency parser was implemented to extract features which provide insight into the sentence structure and complexity: highest number of connections in testimony, average number of connections per word, and average number of connections per sentence. Each testimony was broken down into sentences and then tokenized to find the number of children for each word in the sentence. A variable, max, was created and used to keep track of the highest number of connections for an individual word in the testimony. For the averages, the number of connections for each word was totaled and then divided by the number of words in the testimony and number of sentences.

Consider the statement "Deemed universities charge huge fees." As shown in the figure below, some words are dependent on others. Each dependency is considered to be a "connection," so the word "charge" has two connections.

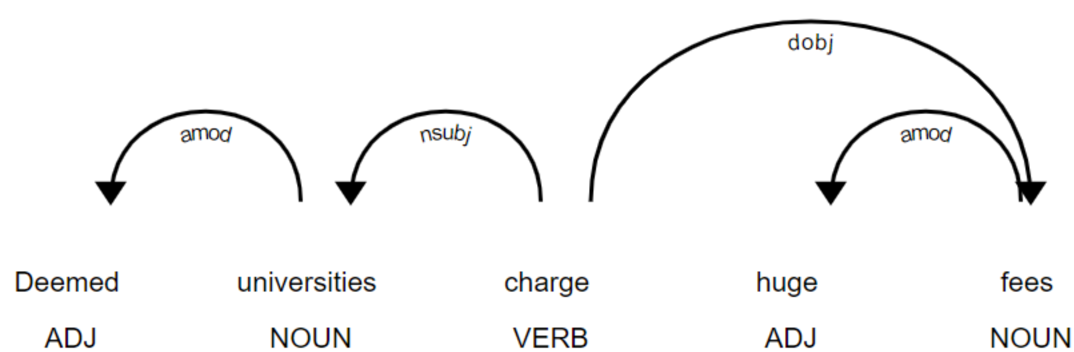

Figure 4.5: Dependency Parsing Example [20]

The max number of connections for this sentence would be ”charge” with 2 connections. The average number of connections per word would be 4/5= 0.8 and the average number of connections per sentence would be 4/1=4.

4.6.8 Additional Metrics

In addition to the features listed above, the models were also trained with metrics such as word count, sentence count, and average sentence length. These metrics provide more information on the complexity and length of the testimony. For word count and sentence count python’s nlp module [4] was used and average sentence length was calculated as (word count)/(sentence count).

4.6.9 Outcome

The outcome field is the label class for the classifier models. In this research it can be either be 0 or 1 to indicate if the testimony was successful. 0 means the bill outcome does not match the alignment of the testimony (the alignment is “For” and the bill fails or the alignment is “Against” and the bill passes). 1 means the bill outcome does match the alignment of the testimony. For testimonies with an ”Indeterminate” alignment the outcome is set to 1. Only 4.415% of the dataset is comprised of these testimonies so setting the outcome to 1 should not greatly impact the accuracy of the models.

4.6.10 List of Finalized Features

Although several features were extracted and implemented, not all provided valuable insight or seemed to have a significant impact on the bill outcome. Using the smaller dataset of 10,000 testimonies, various combinations of features were tested to find the ones that had the greatest impact on the bill outcome. After filtration, the smaller dataset was used to extract features and create inputs for two classifier models: Gaussian Naive Bayes and Support Vector Machine. The accuracy, precision, and recall of these models were documented for each various feature sets. As additional features were extracted the models were run to see if there was improvements in accuracy. Results of testing two feature sets are shown in the tables below. The rest of the results from this testing can be found in the Appendix A.

Features: speech, person type, alignment, flesch score, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.21179625	0.8	0.01346801
	2	0.80965147	0.8228883	0.98051948
	3	0.20643432	0.8333333	0.01666667
	Average	**0.40929401**	**0.8187405**	**0.33688472**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
SVM	1	0.77211796	0.772118	1
	2	0.79088472	0.7908847	1
	3	0.81769437	0.8176944	1
	Average	**0.79356568**	**0.7935657**	**1**

Table 4.6: Feature Testing Results 1

Features: speech, person type, alignment, flesch score, smog score, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.80428954	0.8483965	0.93269231
	2	0.8150134	0.8319328	0.97058824
	3	0.79892761	0.8092643	0.98344371
	Average	**0.80607685**	**0.8298645**	**0.96224142**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
SVM	1	0.80160858	0.8016086	1
	2	0.80697051	0.8069705	1
	3	0.80697051	0.8069705	1
	Average	**0.8051832**	**0.8051832**	**1**

Table 4.7: Feature Testing Results 2

Table 4.6.5 below shows the final list of features chosen to be ingested into the classification models.

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)

Table 4.8: List of Finalized Features with Descriptions

Here is an example testimony as well as the feature values extracted from the text:

Thank you, Mr. Chair, and Committee Members. Aaron Fox with the Los Angeles LGBT Center, also on behalf of California HIV Alliance. SB 1021 further quantifies long established scientific best practices and universally accepted standard of care for the treatment and prevention of HIV AIDS, which is a single tablet regimen, one pill. This will not cause the state Very much money, if at all, and we believe that overall this will save the state dollars as it will prevent, new HIV infections in California, saving our overall healthcare system dollars. The charge that we are trying to protect market share for a specific pharmaceutical company is an affront to those living with

HIV,, and those at risk for it. And we would like to note that for the record for the Committee that we believe that this charge is specifically insulting to our community.

Feature	Value
persontype	General Public
alignment	For
word_count	163
sentence_count	6
avg_sentence_length	27.16666667
flesch	69.28
smog	12.75
successful_trigrams	9
successful_quadgrams	3
successful_pentagrams	1
DT	0.110344828
JJ	0.082758621
NN	0.324137931
PRP	0.062068966
VB	0.117241379
max_connections	10
avg_connections_word	0.963190184
avg_connections_sent	26.16666667
outcome	1

Table 4.9: Feature values for Example Testimony

Chapter 5

MACHINE LEARNING TECHNIQUES

5.1 Background

Text classification is a machine learning technique that assigns a set of predefined categories to open-ended text. For example, many email providers, such as Google and Outlook, categorize emails into Important, Other, Spam, etc. This categorization process is done through text classification. The subject line and contents of the email are analyzed for specific features that indicate which category the email belongs to. A simple example of this is shown in the figure below.

Hello, My car is stuck on the highway, it's	Important
I wanted to have some information on yo	Not Important
Do you offer discount if I buy two insuran	Not Important
All my belongings have been stolen, I do	Important

Figure 5.1: Text Classification Example [32]

Machine learning text classification is utilized for scalability, real-time analysis, and consistent criteria. There are three types of systems for automatic text classification: rule-based systems, machine learning-based systems, and hybrid systems [32]. Some of the most popular text classification algorithms include the Naive Bayes family of algorithms, support vector machines, and neural networks. Four different types of performance metrics were used to assess the performance of each classifier: Accuracy, Precision, Recall, and F1 Score. Accuracy is calculated as the ratio between the number if correct predictions to the total number of predictions [12].

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{True}_{\boldsymbol{positive}} + \boldsymbol{True}_{\boldsymbol{negative}}}{\boldsymbol{True}_{\boldsymbol{positive}} + \boldsymbol{True}_{\boldsymbol{negative}} + \boldsymbol{False}_{\boldsymbol{positive}} + \boldsymbol{False}_{\boldsymbol{negative}}}$$

Figure 5.2: Accuracy Formula

Precision is calculated as the ratio between the number of correctly classified positive samples to the total number of classified positive samples (either correctly or incorrectly). Precision measures the model's accuracy in classifying a sample as positive [12].

$$\boldsymbol{Precision} = \frac{\boldsymbol{True}_{\boldsymbol{positive}}}{\boldsymbol{True}_{\boldsymbol{positive}} + \boldsymbol{False}_{\boldsymbol{positive}}}$$

Figure 5.3: Precision Formula

Recall is calculated as the ratio between the number of correctly classified positive samples to the total number of positive samples. Recall measures the model's ability to detect positive samples. For binary classification, recall can also be called sensitivity. The higher the recall, the more positive samples detected [12].

$$\boldsymbol{Recall} = \frac{\boldsymbol{True}_{\boldsymbol{positive}}}{\boldsymbol{True}_{\boldsymbol{positive}} + \boldsymbol{False}_{\boldsymbol{negative}}}$$

Figure 5.4: Recall Formula

F1 score is defined as the harmonic mean of precision and recall. The F1 Score combines Precision and Recall into a single metric so it is directly correlated to both metrics. If the precision and recall are high, the f1 score will also be high. If the precision and recall are low, the f1 score will also be low.

$$F1\ score = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

Figure 5.5: F1 Score Formula

5.2 Data Preparation

Prior to ingestion into a classification model, each of the features was encoded or normalized. Categorical features: PersonType and Alignment were encoded using sklearn's preprocessing Label Encoder module [34]. Numerical features: word count, sentence count, average sentence length, flesch score, smog score, successful trigrams, successful quadgrams, and successful pentagrams were normalized using sklearn's preprocessing module which uses L2 normalization technique. The pos proportion features: DT, JJ, NN, PRP, and VB were not normalized.

Thank you, Mr. Chair, and Committee Members. Aaron Fox with the Los Angeles LGBT Center, also on behalf of California HIV Alliance. SB 1021 further quantifies long established scientific best practices and universally accepted standard of care for the treatment and prevention of HIV AIDS,, which is a single tablet regimen, one pill. This will not cause the state Very much money, if at all, and we believe that overall this will save the state dollars as it will prevent, new HIV infections in California, saving our overall healthcare system dollars. The charge that we are trying to protect market share for a specific pharmaceutical company is an affront to those living with HIV,, and those at risk for it. And we would like to note that for the record for the Committee that we believe that this charge is specifically insulting to our community.

```
persontype                0.000000
alignment                 2.000000
word_count                0.054036
sentence_count            0.041667
avg_sentence_length       0.352179
flesch                    0.526383
smog                      0.656202
successful_trigrams       0.176471
successful_quadgrams      0.157895
successful_pentagrams     0.100000
DT                        0.110345
JJ                        0.082759
NN                        0.324138
PRP                       0.062069
VB                        0.117241
max_connections           0.444444
avg_connections_word      0.946346
```

Figure 5.6: Normalized and Encoded Features

5.3 Naive Bayes

Naïve Bayes is a popular classification algorithm mostly used to get the base accuracy of the dataset. Naive Bayes assumes that the features, also known as predictors, are independent which means changing the value of one feature does not influence the values of the other variables all predictors have equal impact on the outcome. It is easy and efficient for predicting classes and performs well in multi-class prediction[7]. Naive Bayes performs better in cases with categorical input variables compared to numerical variables and compared to other models like logistic regression when the assumption of independence holds. Therefore, Naïve Bayes classifiers are often used in text classification and is best used when the dataset is huge and when the training set is small. The major disadvantage of Naive Bayes is the assumption of independent predictors since in real life, it is almost impossible a set of predictors are completely independent [40].

For this research two variations of Naive Bayes were implemented using Sklearn's Naive Bayes module [34]: Multinomial Naive Bayes and Gaussian Naive Bayes. Gaussian Naive Bayes is useful when working with continuous values whose probabilities can be modeled using Gaussian distribution also known as normal distribution as shown in the curve below [33].

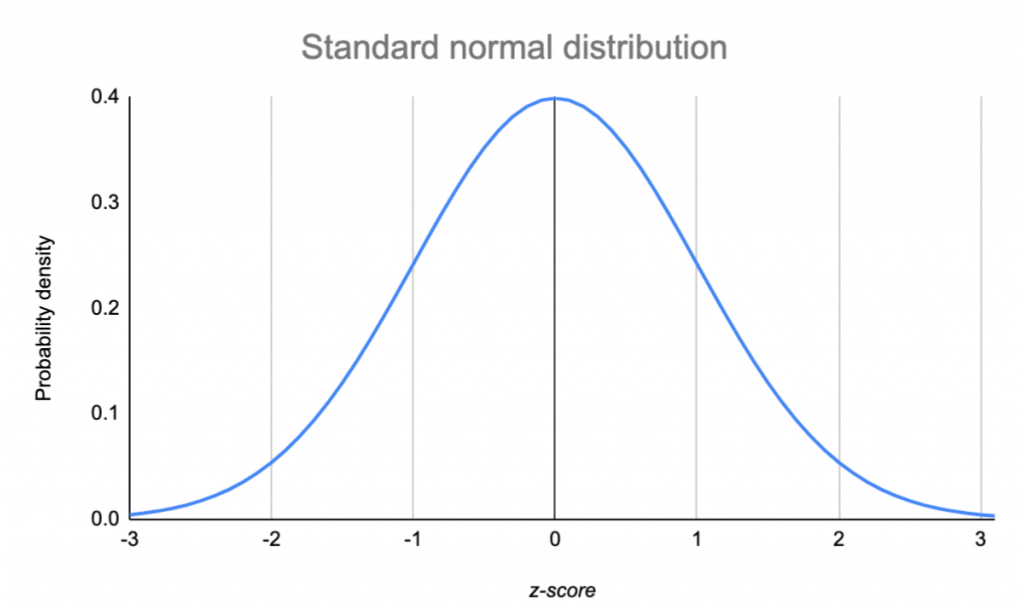

Figure 5.7: Normal Distribution Curve [3]

Multinomial Naive Bayes is useful to model feature vectors where each value represents the number of occurrences of a term or its relative frequency [33].

5.4 Support Vector Machine

Support Vector Machines (SVM) are a set of supervised learning methods used for classification, regression, and outlier detection [13]. SVM draws a line or "hyperplane" that divides a space into two sub-spaces. One subspace contains vectors (tags) that belong to a group, and another subspace contains vectors that do not belong to that group [32].

The distance between support vectors is known as the margin, the optimal hyperplane is the one with the largest margin, as shown in Figure 5.4.1 below.

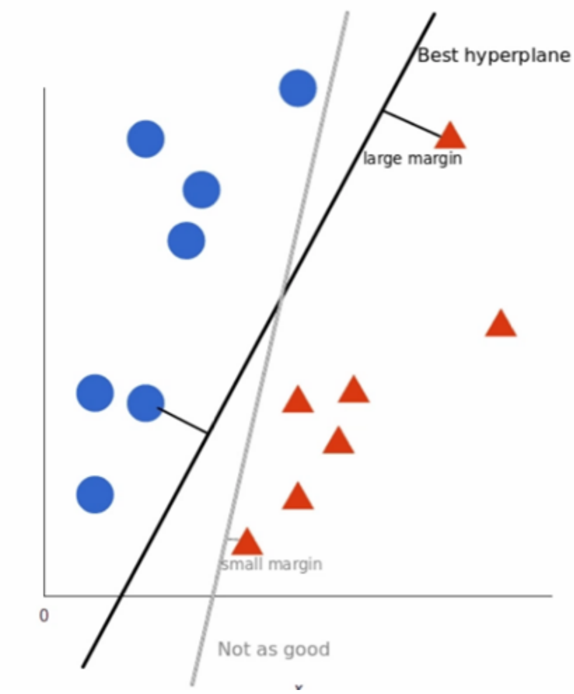

Figure 5.8: Support Vector Machine Diagram [32]

The objective of SVM is to select a hyperplane with the maximum possible margin between support vectors in the given dataset. SVM does this in two steps: generating hyperplanes which segregate the classes in the best way and then selecting the right hyperplane with the maximum segregation between the nearest datapoints on either side. The advantages of SVM include staying effective in situations where the number of dimensions is greater than the number of samples and using a subset of training points in the decision function, so it is memory efficient. Support vector machines don't need much training data to start providing accurate results, however it does require more computational resources than Naïve Bayes, but the results are even faster and more accurate [13]. For this research Sklearn's SVC module was utilized to implement the SVM module [34].

5.5 Neural Networks

A neural network is a series of algorithms that is used to recognize underlying relationships in a dataset in a way that mimics how a human brain operates. Neural networks can adapt to changing input, so the network generates the best possible result without needing to redesign the output criteria. A neural network works similar to the human brain's neural network. As information enters the brain, each layer, or level, of neurons does its particular job of processing the incoming information, deriving insights, and passing them on to the next and more senior layer. In a neural network a "neuron" is a mathematical function that collects and classifies information according to a specific architecture [45].

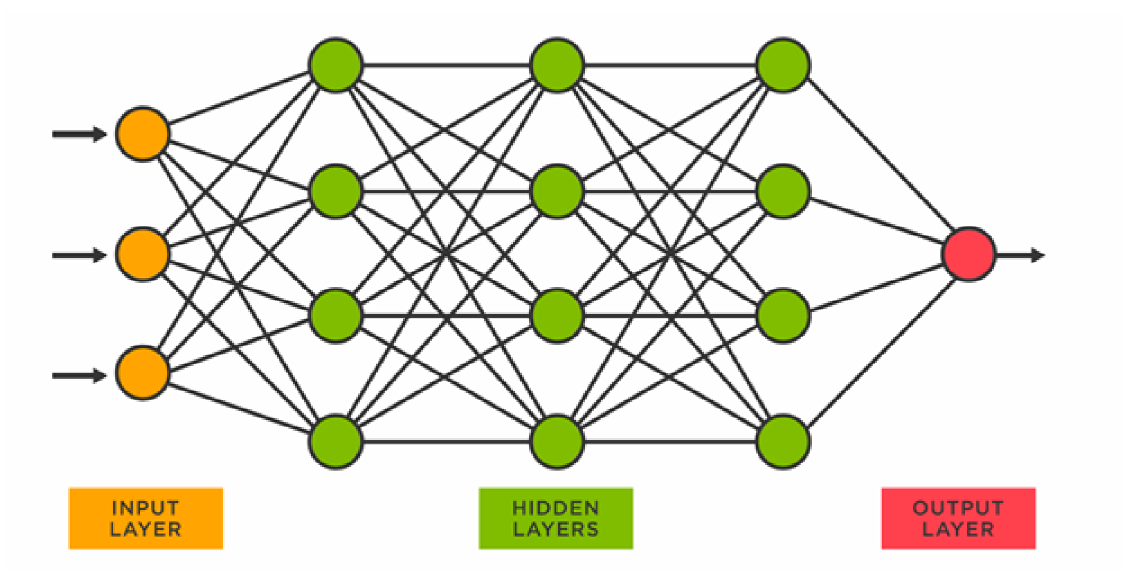

Figure 5.9: Simple Neural Network Diagram [45]

As shown in the figure above, neural networks try to imitate this multi-layered approach for processing various information and basing decisions on them.

There are several types of neural networks, for the purpose of this research Convolutional Neural Networks (CNNs) and Fully Connected Neural Networks (FCNNs) were considered. A fully connected neural network contains a series of fully connected layers that connect every neuron in one layer to every neuron in the other layer. The biggest advantage of fully connected networks is that no special assumptions need to be made about the input so they are broadly applicable. However this flexibility also tends to make fully connected neural networks weaker compared to more specific or special purpose networks such as CNNs [26].

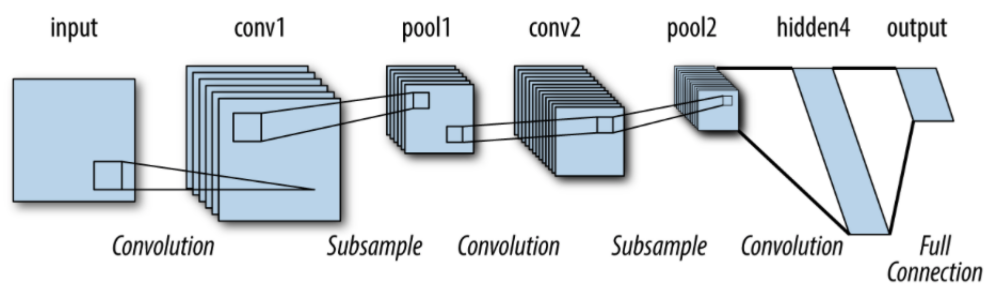

Figure 5.10: Convolutional Neural Network Diagram [31]

CNNs are type of neural network usually used for image data. As shown in Figure 5.6.2, they consist of three layers: a convolutional layer, a pooling layer, and a fully connected layer. The variation of layers allows CNNs to detect simple patterns at first, such as lines or shapes, and then detect more complex patterns, such as faces or objects. Although CNNs are typically used for image data, they can also be used for sentiment analysis or text classification so it was thought to be a good fit for this research. In order to format the data for ingestion into the CNN, the raw text data was convert into word embeddings using python's Word2Vec module. The testimonies were parsed, tokenized, and transformed into a word index. This word index was then transformed into an embedding index, using Stanford's GloVe 100d word embeddings corpus. By iterating through the word index and referencing the embedding index an embedding matrix was created. This embedding matrix was then used as weights for an embedding layer for the CNN. The embedding layer was added prior to the convolutional layers [31].

Although this practice would be effective in theory, the CNN with an embedding layer was not optimal for this data set. After several variations in implementation, it was decided that this model was not feasible given the time requirements of the project. A Convolutional Neural Network would be beneficial for predicting the outcome of a testimony but it would not provide significant information on successful components of a testimony.

Therefore only a fully connected neural network was implemented, using Tensorflow [27] and Keras [8]. The neural network consists of a sequential model with five dense layers and the following hyperparameters:

- Activation Functions: Rectified linear activation function (ReLU) (layer 1-4), Softmax (layer 5)
- Loss Function: Binary Crossentropy
- Optimizer: Stochastic Gradient Descent (SGD)
- Learning Rate: 0.01
- Epochs: 20
- Batch Size: 200

These hyperparameters were selected after testing various options as shown in Table below. Binary Crossentropy was selected as the loss function because it works best with binary classes like the outcome field in this research [8]. Similar to testing viable features, these hyperparameters were tested using the smaller dataset of 10,000 testimonies.

Fully Connnected Neural Network	Accuracy
activation= relu/softmax, optimizer = SGD, learning rate = 0.2, loss = categorical_crossentropy, batch = 100, epochs = 20	19.28
activation= sigmoid/softmax, optimizer = SGD, learning rate = 0.01, loss = binary_crossentropy, batch = 200, epochs = 20	72.327
activation= relu/softmax, optimizer = SGD, learning rate = 0.01, loss = binary_crossentropy, batch = 200, epochs = 20	72.75
activation= relu/softmax, optimizer = Adam, learning rate = 0.01, loss = binary_crossentropy, batch = 200, epochs = 20	71.6712
activation= relu/softmax, optimizer = SGD, learning rate = 0.001, loss = binary_crossentropy, batch = 120, epochs = 25	71.63
activation= relu/softmax, optimizer = SGD, learning rate = 0.01, loss = binary_crossentropy, batch = 120, epochs = 20	72.0185

Table 5.1: FCNN Hyperparameter Test

5.6 TF-IDF

In addition to manually extracting features and inputting them into a classifier, TF-IDF was also used to extract features based on the frequency of words used in testimonies [43]. TF-IDF stands for term frequency-inverse document frequency and it is used to quantify the importance or relevance of string representations in a document amongst a collection of documents (corpus). In this case a document would be an individual testimony and the collection of documents would be all the testimonies in the dataset. Figure 5.5.1 shows the formula for both term frequency and inverse document frequency.

$$TF(i,j) = \frac{\text{Term i frequency in document j}}{\text{Total words in document j}}$$

$$IDF(i) = \log_2\left(\frac{\text{Total documents}}{\text{documents with term i}}\right)$$

Figure 5.11: Formula for Term Frequency and Inverse-Document Frequency [43]

Term frequency is measured either as the number of occurrences of a term within a document (raw count) or term frequency adjusted for the length of the document (raw count/total number of terms in document). Inverse document frequency measures how common or uncommon a term is in all documents. It is calculated as log of total documents divided by the number of documents containing a specific term. Multiplying TF and IDF generates the TF-IDF.

Prior to ingestion in ML classifiers the data needs to be converted into a vector of numerical data. TF-IDF vectorization calculates the TF-IDF score for each term (word) in the corpus relative to the document (testimony) it is in and then formats the scores into a vector [34]. Therefore, each testimony in the dataset, has its own vector containing the TF-IDF score for each word in the collection of testimonies. These vectors are then used for ingestion into two different a Naïve Bayes models: Gaussian Naive Bayes and Multinomial Naive Bayes.

```
  (0, 22625)	0.10737263647226114
  (0, 1988)	0.33745873490713063
  (0, 2019)	0.580976146657532
  (0, 1904)	0.5432487742315509
  (0, 7250)	0.49187417647479503
A Jordan, CSJ's strong support.
```

Figure 5.12: TF-IDF vectors for sample testimony

The vectors follow the general form: (A,B) C where A is the document index, B is the specific word-vector index, and C is the Tf-IDF score for word B in document A. In this example A is 0 since this is the first testimony in the training dataset. This testimony produces five TF-IDF vectors, one for each term in the testimony.

5.7 Transformers

As mentioned in Chapter 2: Related Works, Transformer is an attention mechanism that learns contextual relations between words in a text. Transformer includes two separate mechanisms, an encoder that reads the text input and a decoder that produces a prediction for the task. In BERT, the Transformer encoder reads the entire sequence of words at once unlike directional models, which read the text input sequentially. This allows the model to learn the context of the words based on its surroundings. The input for the Transformer encoder is a sequence of tokens, which are first embedded into vectors and then processed in the neural network, and the output is a sequence of vectors, in which each vector corresponds to an input token with the same index [47].

5.8 Bidirectional Encoder Representations from Transformers (BERT)

As mentioned in Chapter 2: Related Works as well, Bidirectional Encoder Representations from Transformers, also known as BERT, is a transformer-based machine learning technique for natural language processing pre-training developed by Google. BERT's model architecture is a multi-layer bidirectional Transformer encoder with various model sizes, such as BERTBASE (L=12, H=768, A=12, Total Parameters=110M) and BERTLARGE (L=24, H=1024, A=16, Total Parameters=340M). For this research, various BERT models were tested to find the most efficient and based on the results two were selected to run on the full dataset. For this dataset, BERT was used to do basic text classification and unlike previous models, feature extraction was not done prior to ingestion into the model. Each BERT model has a corresponding BERT preprocessing model. Each testimony was preprocessed using a BERT preprocessing model,encoded using BERT's encoder, and the encoded data was then used to train the respective BERT model [19].

Chapter 6

EXPERIMENTS AND RESULTS

6.1 Feature Analysis

Prior to ingestion into the model, the features themselves were analyzed to infer which had the most impact on the output or if there was any correlation between features.

A correlation matrix was created, as shown in Figure 6.1.1 below. A correlation matrix is a table showing correlation coefficients between variables [5]. Each cell in the table shows the correlation between two variables. To include the categorical features in the correlation matrix, the categorical values were converted to numerical values. For person type General Public and Lobbyist were changed to 2 and -2 respectively. For alignment "For" was changed to 3, "Against" -3, and Neutral/Indeterminate 0. After the categorical features were transformed, python's correlation and heatmap modules were used to generate the correlation matrix.

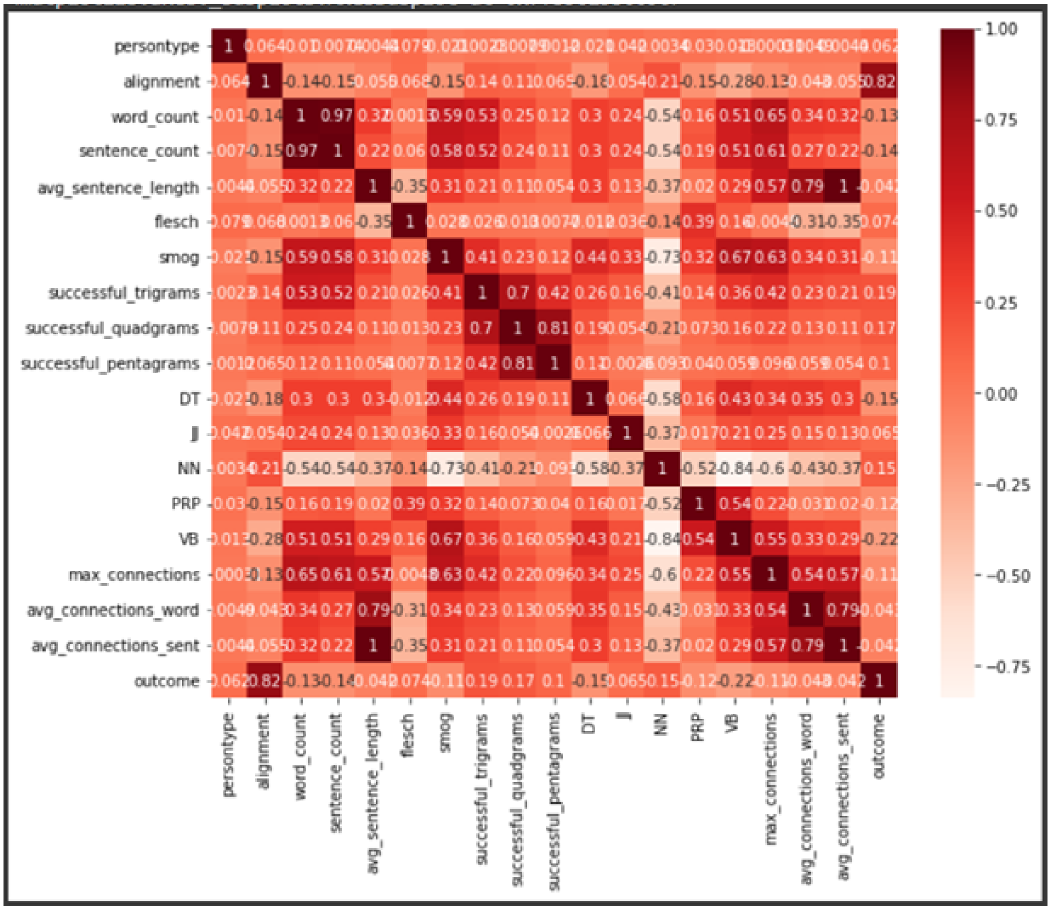

Figure 6.1: Correlation Matrix for Features

The darker squares indicate a higher correlation value; features with higher correlation typically have direct relationships such as sentence count and word count. If the sentence count is increases than the word count will most likely also increase. These direct relationships can also be seen with the successful phrases: successful trigrams, successful quadgrams, successful pentagrams have higher correlations most likely since they were extracted in the same manner and represent similar information but contain varying lengths of phrases.

However some correlations between features were less expected, such as average connections per sentence and average sentence length. Average connections per sentence is the summation of connections per word divided by the number of sentences, whereas average sentence length is the total number of words in the testimony divided by the number of sentences. The correlation between these two features being high indicates that the average number of connections per word is close 1.

Tables 6.1.1 and 6.1.2 below show a subset of the values as well as the averages for these features and provide an explanation for the higher correlation value.

avg_sentence_length	avg_connections_sent	avg_connections_word	max_connections
8	7	0.875	4
11.5	10.5	0.913043478	7
17	16	0.941176471	11
11	10	0.909090909	6
29.72727273	28.72727273	0.966360856	10
11	10	0.909090909	4
23.35	22.35	0.957173448	10
14	13	0.928571429	6
17.66	16.66	0.943374858	11
34	33	0.970588235	9

Table 6.1: Subset of data for average sentence length, average connections per sentence, average connections per word, and maximum connections of a word in each testimony

Features	avg_sentence_length	avg_connections_sent	avg_connections_word	max_connections
Averages	17.60805717	16.60805717	0.932713123	6.656194519

Table 6.2: Average values for average sentence length, average connections per sentence, average connections per word, and maximum connections

Aside from the first entry, the average connections per word ranges from 0.9-0.97 which is fairly close to 1 and explains why the average sentence length and average connections per sentence are very similar values. This is slightly unexpected since some words had 11 or 10 connections as shown in the maximum connections column but the average values for average sentence length and average connections per sentence are very similar in value, 17.6 and 16.8 respectively. This indicates that some words had several connections, whereas others had only 1 or none.

In addition to analyzing correlations between features, the values of each feature were analyzed based on when the testimony aligns with the bill outcome or does not align with the bill outcome. This was calculated by separating the feature data set into two data sets based on the outcome. The first data set only contained feature values when outcome equals 1, which means the testimony was successful and the second data set contained feature values when the outcome equals 0 which means the testimony alignment was not successful. The numerical features within each data set were then averaged and compared as shown in Table 6.1.3.

Feature	Successful Testimony	Unsuccessful Testimony	Absolute Difference	P-value T-Test	P-value F-Test
word_count	117.833047	190.881546	73.048499	7.67E-56	1
sentence_count	5.85722	9.209213	3.351993	2.16E-56	1
flesch	74.516194	71.222013	3.294181	3.25E-17	0.999999953
smog	4.570162	5.986948	1.416786	1.94E-41	0.999491085
successful_trigrams	1.921143	0.591171	1.329972	9.52E-112	1.11E-16
outcome	1	0	1	0.00E+00	N/A
avg_connections_sent	16.402933	17.131582	0.728649	3.64E-08	0.787904079
avg_sentence_length	17.402933	18.131582	0.728649	3.64E-08	0.787904079
max_connections	6.487538	7.086647	0.599109	3.82E-39	1
successful_quadgrams	0.746562	0.211955	0.534607	6.66E-100	1.11E-16
successful_pentagrams	0.311775	0.089389	0.222386	1.58E-54	1.11E-16
NN	0.495376	0.439062	0.056314	5.84E-65	0.442864534
VB	0.080675	0.117679	0.037004	1.87E-140	7.51E-05
DT	0.073353	0.088084	0.014731	2.03E-62	4.88E-11
PRP	0.046135	0.058055	0.01192	1.81E-40	0.002234407
JJ	0.045557	0.039488	0.006069	7.57E-14	1.11E-16
avg_connections_word	0.931904	0.934777	0.002873	7.72E-07	0.006017191

Table 6.3: Averages of each Feature for Successful and Unsuccessful testimonies, Absolute Difference, and P-Values

In addition to averages, the p-values for the two datasets were calculated using two tests: T-Test and F-Test. T-test is typically used to measure the difference between means of two sample sets, whereas F-tests are used to calculate the variance between two sample sets[39]. The p-values in the table above show that there is significant variance between sample tests for certain features such as successful trigrams, quadgrams, and pentagrams. The p-values from the T-tests were all very small which indicates that this test is not an appropriate measure of statistical significance.

The results showhat successful testimonies are shorter and easier to comprehend. This can be seen with the difference in average word count, sentence sount, flesch score, and smog score, as shown in the figures below.

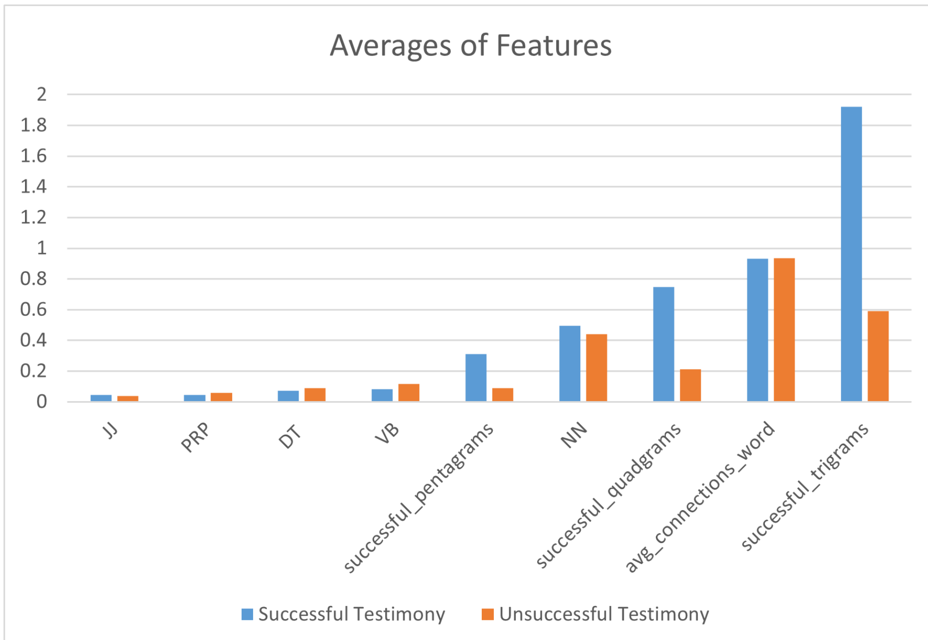

Figure 6.2: Comparison of average values for first set of features

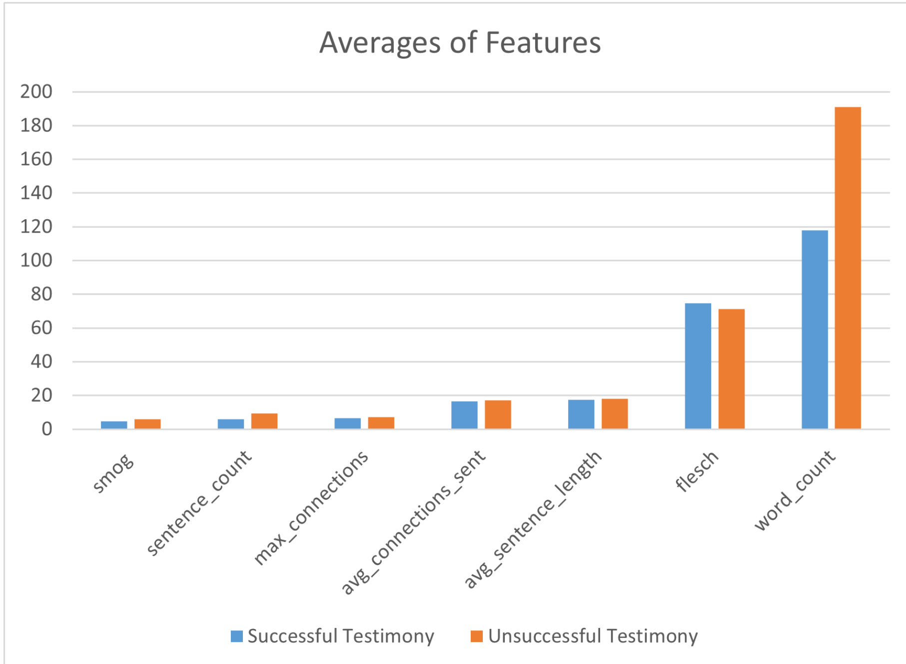

Figure 6.3: Comparison of average values for second set of features

A higher flesch score indicates the text is easier to read, where as a higher smog score indicates the text is harder to read. Successful testimonies tended to have higher flesch score and lower smog score. Successful testimonies also had a lower word count but similar sentence count as unsuccessful testimonies which indicates that unsuccessful testimonies have longer sentences and more run ons. Additionally, unsuccessful testimonies have slightly higher values for average number of connections and max connections which means the sentences within those testimonies contain more dependencies. The feature analysis results indicate that shorter and less complex (easier to comprehend) testimonies are more successful.

Successful testimonies also contain more successful phrases: trigrams, quadgrams, and pentagrams. The difference between the number of successful phrases for successful testimonies and unsuccessful testimonies indicate that these phrases have an impact on the bill outcome. Looking at these successful phrases more closely provides insight into what content makes a testimony more successful. For example, the positive 5 word phrase or pentagram, "don't have an official position" has a success rate of 100% which means each time this phrase appears in a testimony with an alignment of "For" or "Indeterminate" the bill passes. This indicates that mentioning an official position is a key aspect of those testimonies. Additionally, successful testimonies with an alignment of "For" often had phrases containing the word "oppose" as shown in Table 6.1.4 below.

4 word phrase (quadgram)	Success Rate
('org', 'in', 'opposition', 'to')	1
('were', 'opposed', 'to', 'the')	1
('but', 'at', 'this', 'time')	1
('we', 'share', 'the', 'concerns')	1
('org', 'we', 'oppose', 'this')	1
('dont', 'have', 'an', 'official')	1
('you', 'to', 'oppose', 'this')	1
('to', 'oppose', 'this', 'bill')	1
('org', 'in', 'respectful', 'opposition')	1
('opposition', 'to', 'this', 'measure')	1
('you', 'to', 'vote', 'no')	1
('in', 'opposition', 'person', 'on')	1
('opposition', 'person', 'on', 'behalf')	1
('with', 'org', 'also', 'opposed')	1
('were', 'going', 'to', 'continue')	1
('those', 'reasons', 'were', 'opposed')	1

Table 6.4: Examples of Positive Successful Quadgrams (4 word phrases)

The phrases in this table all have a success rate of 100% meaning any time these phrases appeared in a testimony with an alignment of "For," the bill passed. Based on these phrases it can be inferred that mentioning reasons why not to oppose to bill or addressing the concerns of those who might vote no are valuable to testimonies in favor of the bill. The full list of these phrases can be found in Appendix A.

Some features such as proportions of Parts of Speech (POS) did not have a large difference in values for successful and unsuccessful testimonies, as shown in Table 6.1.3, however these features were still valuable for training the classifier models. This means that changing the number of nouns, adjectives, or verbs within the testimony does not affect the bill outcome as much as the structure of the testimony.

6.2 Results of Classification Models

Two types of classification models were implemented in this research: the first type had with manual feature extraction done prior to ingestion into the model (Naive Bayes, SVM, and FCNN) and the second utilized the model's preprocessing capabilities to do feature extraction (TF-IDF, BERT).

Additionally, there are two sets of results for this research: pre-data balancing and post-data balancing. It was found that the original dataset contained 9308 successful testimonies and 3646 unsuccessful testimonies which led to skewed results and over fitted classifier models. In order to combat this issue the dataset was balanced using an undersampling technique in which 5660 successful testimonies were randomly removed prior to feature analysis or ingestion into the classifier models.

The results for all classifier models pre-data balancing and post-data balancing are shown below.

6.2.1 Results of Classification Models with Feature Extraction

For implementing the first type of classification models, various feature sets were tested using the smaller dataset as discussed in Section 4.6.9. Once the list of features was finalized, the full dataset was tested. Each of the models was executed three times, as depicted by the Trial Numbers on Figure 6.2.1 below. For each trial the dataset was randomly split using Sklearn's train-test-split module [34] with 80% of the data being in the training set and 20% in testing data. This random division of the dataset meant the model was given a new testing set each time. The average on the trials was taken to get the final accuracy, precision, and recall metrics of the model.

6.2.1.1 Pre-Data Balancing

As shown in Fig 6.2.1 below, Support Vector Machine received the highest accuracy followed by Gaussian Naive Bayes, Multinomial Naive Bayes, and Fully Connected Neural Network respectively. This is most likely due to the fact the feature dataset was all numerical variables and the label class was binary which is better suited for SVM models compared to Naive Bayes models.

Model	Trial Number	Accuracy	Precision	Recall	F1
Gaussian NB	1	0.90274025	0.952381	0.9105517	0.930997
	2	0.90274025	0.9476584	0.9163559	0.931744
	3	0.91470475	0.9533333	0.9260658	0.939502
	Average	**0.90672842**	**0.9511242**	**0.9176578**	**0.934081**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
Multinomial NB	1	0.89695098	0.9	0.9641136	0.930954
	2	0.88267078	0.8895493	0.956846	0.921971
	3	0.89116171	0.8856161	0.9735564	0.927506
	Average	**0.89026116**	**0.8917218**	**0.9648387**	**0.926811**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
SVM	1	0.91933616	0.9329107	0.9556277	0.944133
	2	0.93516017	0.9466737	0.9639785	0.955248
	3	0.91933616	0.937799	0.950431	0.944073
	Average	**0.92461083**	**0.9391278**	**0.9566791**	**0.947818**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
FCN	1	0.721729	0.721729	1	0.838377
	2	0.715554	0.715554	1	0.834196
	3	0.72787	0.72787	1	0.842506
	Average	**0.72171767**	**0.7217177**	**1**	**0.838359**

Table 6.5: Accuracy, Precision, and Recall metrics for Guassian Naive Bayes, Multinomial Naive Bayes, Support Vector Machine, and Fully Connected Neural Network Pre-Data Balancing

6.2.1.2 Post-Data Balancing

After balancing the dataset, the accuracy values for the classifier models stayed mostly the same with the exception of the Fully Connected Neural Network model as shown in Table 6.2.2 below. The reason the Fully Connected Neural Network (FCNN) may have seen a drastic decrease in accuracy is due to the smaller dataset and lack of over fitting.

Model	Trial Number	Accuracy	Precision	Recall	F1
Gaussian NB	1	0.89924606	0.8979021	0.896648	0.897275
	2	0.89239205	0.8790637	0.9135135	0.895958
	3	0.89102125	0.8935574	0.8848821	0.889199
	Average	**0.89421979**	**0.8901744**	**0.8983479**	**0.894144**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
Multinomial NB	1	0.91226868	0.879845	0.9511173	0.914094
	2	0.91021247	0.8681983	0.9702703	0.916401
	3	0.91432488	0.8753149	0.963939	0.917492
	Average	**0.91226868**	**0.8744527**	**0.9617755**	**0.915995**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
SVM	1	0.90541467	0.8696203	0.9515235	0.90873
	2	0.93351611	0.9103896	0.9615912	0.93529
	3	0.91501028	0.879081	0.9680426	0.92142
	Average	**0.91798035**	**0.8863636**	**0.9603858**	**0.921813**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
FCN	1	0.488005	0.488005	1	0.655918
	2	0.517478	0.517478	1	0.682024
	3	0.502399	0.502399	1	0.668796
	Average	**0.50262733**	**0.5026273**	**1**	**0.668913**

Table 6.6: Accuracy, Precision, and Recall metrics for Guassian Naive Bayes, Multinomial Naive Bayes, Support Vector Machine, and Fully Connected Neural Network Post-Data Balancing

6.2.2 Results of Classification Model without Feature Extraction

In addition to manual feature extraction, models with built in preprocessing capabilities such as TF-IDF and BERT were also developed. TF-IDF was implemented on both Gaussian and Multinomial Naive Bayes.

6.2.2.1 Pre-Data Balancing

As shown in Fig 6.3.1 below, Similar to the first set of results, the Gaussian Naive Bayes model achieved a higher accuracy than Multinomial, with 77.57% and 59.03% respectively. However, neither TF-IDF Naive Bayes models produced the same level of accuracy as the manually feature extraction models which indicates that the features extracted do play a significant role in the bill outcome.

Model	Trial Number	Accuracy	Precision	Recall	F1
TF-IDF Guassian NB	1	0.58510228	0.8188235	0.5529661	0.660133
	2	0.59590892	0.8179036	0.5692226	0.671272
	3	0.59011964	0.8296068	0.5612937	0.669571
	Average	**0.59037695**	**0.8221113**	**0.5611608**	**0.666992**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
TF-IDF Multinomial NB	1	0.76804323	0.7572895	0.9946063	0.859874
	2	0.77267464	0.7623393	0.99137	0.861899
	3	0.78656889	0.7760634	0.9914758	0.870643
	Average	**0.77576225**	**0.7652307**	**0.992484**	**0.864139**

Table 6.7: Accuracy, Precision, and Recall metrics for TF-IDF implementation of Naive Bayes Pre-data balancing

As for BERT, each training model has a corresponding preprocessing model through which the raw text data and bill outcome were inputted. After preprocessing the data was then encoded and then run through the BERT model. In order to find the most optimal BERT model, various options were ran on a smaller dataset of 10,000 testimonies and the results were compared.

BERT Model	epochs/batch size	Accuracy
small_bert/bert_en_uncased_L-2_H-768_A-12	epochs=10, batch_size = 32	80.06
small_bert/bert_en_uncased_L-4_H-768_A-12	epochs=10, batch_size = 32	**81.67**
small_bert/bert_en_uncased_L-6_H-768_A-12	epochs=10, batch_size = 32	81.03
small_bert/bert_en_uncased_L-8_H-768_A-12	epochs=10, batch_size = 32	81.35
small_bert/bert_en_uncased_L-10_H-768_A-12	epochs=10, batch_size = 32	80.71
small_bert/bert_en_uncased_L-12_H-768_A-12	epochs=10, batch_size = 32	**81.99**
small_bert/bert_en_uncased_L-2_H-128_A-2	epochs=10, batch_size = 32	81.03
small_bert/bert_en_uncased_L-4_H-128_A-2	epochs=10, batch_size = 32	80.71

Table 6.8: Accuracy metrics for various BERT Models, tested on smaller data set

As shown in Fig 6.3.2 above, the accuracies achieved were all around 80%, however the two models with the highest accuracies, small_bert/bert_en_uncased_L-4_H-768_A-12 and small_bert/bert_en_uncased_L-12_H-768_A-12, were utilized for the full dataset. Here L is the number of layers also known as Transformer blocks, H is the hidden size, and A represents the number of self-attention heads. The results of these two models can be seen in Fig 6.3.3 below. The L-4 model achieved a slightly higher accuracy compared to the L-12 model, 82.44% and 81.89% respectively.

BERT Model	Trial Number	Accuracy	Precision	Recall	F1
small_bert/bert_en_uncased_L-4_H-768_A-12	1	0.818432	0.857201	0.896977	0.876638
	2	0.831598	0.874636	0.893997	0.884211
	3	0.823025	0.878257	0.875266	0.876759
	Average	**0.82435167**	**0.87003133**	**0.8887467**	**0.879203**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
small_bert/bert_en_uncased_L-12_H-768_A-12	1	0.813227	0.833781	0.924649	0.876867
	2	0.822719	0.855707	0.906343	0.880297
	3	0.820882	0.85837	0.899106	0.878266
	Average	**0.81894267**	**0.849286**	**0.9100327**	**0.878477**

Table 6.9: Accuracy metrics for selected BERT Models Pre-data balancing

6.2.2.2 Post-Data Balancing

The results for the Tf-IDF models, after balancing the data set are shown in Table 6.2.6 below. Both TF-IDF models have a higher accuracy post-data balancing which is expected since this model is dependent on the relative frequency of each word in each testimony compared to all the testimonies in the dataset. With an equal number of successful and unsuccessful testimonies in the dataset, the models are more accurate.

Model	Trial Number	Accuracy	Precision	Recall	F1
TF-IDF Guassian NB	1	0.65867032	0.62740385	0.7352113	0.677043
	2	0.66552433	0.63233533	0.7447109	0.683938
	3	0.67169294	0.65925059	0.7496671	0.701558
	Average	**0.66529586**	**0.63966325**	**0.7431964**	**0.687513**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
TF-IDF Multinomial NB	1	0.81768334	0.86633663	0.7394366	0.797872
	2	0.82316655	0.85964912	0.7602257	0.806886
	3	0.81288554	0.8734375	0.7443409	0.803738
	Average	**0.81791181**	**0.86647442**	**0.7480011**	**0.802832**

Table 6.10: Accuracy, Precision, and Recall metrics for TF-IDF implementation of Naive Bayes Post-data balancing

The selected BERT models were also rereun after balancing the dataset and the results are shown in Table 6.2.7 below. Similar to Naive Bayes and Support Vector Machine models, the accuracy for BERT models stayed the same.

BERT Model	Trial Number	Accuracy	Precision	Recall	F1
small_bert/bert_en_uncased_L-4_H-768_A-12	1	0.810373	0.830125	0.784797	0.806825
	2	0.794165	0.831138	0.743041	0.784624
	3	0.799028	0.786735	0.825482	0.805643
	Average	**0.80118867**	**0.81599933**	**0.78444**	**0.799031**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**	**F1**
small_bert/bert_en_uncased_L-12_H-768_A-12	1	0.806051	0.804233	0.813704	0.808941
	2	0.786602	0.779855	0.804069	0.791777
	3	0.799028	0.810154	0.785867	0.797826
	Average	**0.797227**	**0.79808067**	**0.8012133**	**0.799514**

Table 6.11: Accuracy metrics for selected BERT Models Post-data balancing

6.2.3 Comparison of Classification Models

6.2.3.1 Pre-Data Balancing

Figure 6.4.1 below shows a comparison of all the classification models implemented.

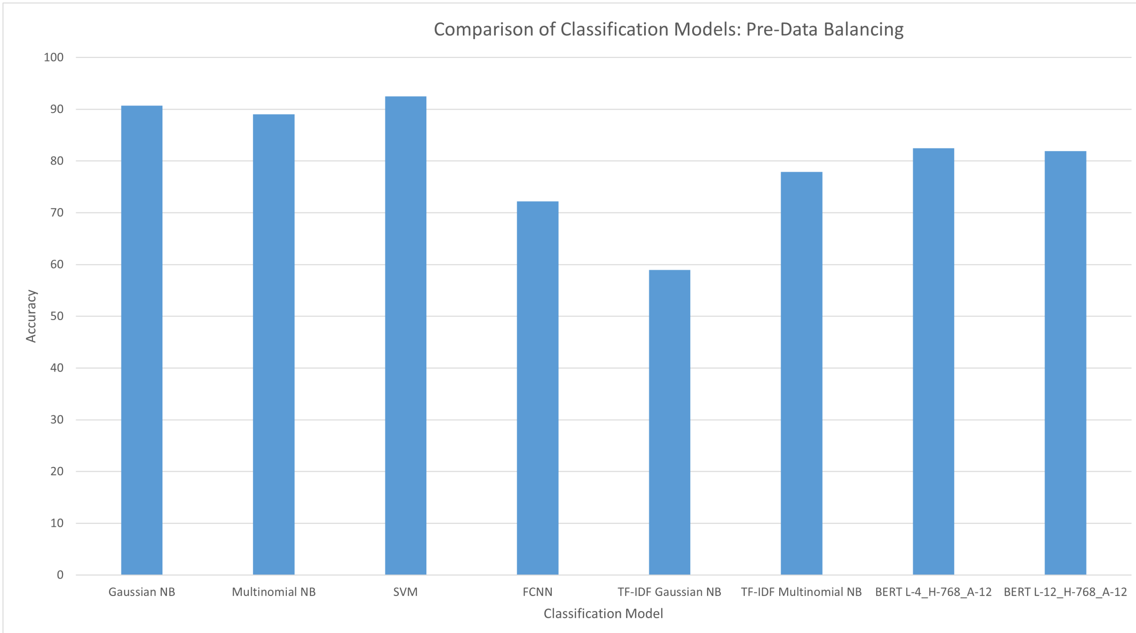

Figure 6.4: Comparison of Classification Models Pre-Data Balancing

The support vector machine model produced the highest accuracy with the Guassian Naive Bayes model being a close second. This comparison also indicates that manual feature extraction is a necessary step prior to classification. The models that did not have manual feature extraction, TF-IDF Guassian NB, TF-IDF Multinomial NB, and both BERT models, had lower accuracy values compared to the SVM and Naive Bayes models. The exception to this was the Fully Connected Neural Network (FCNN), which did not perform as well as expected. This could be due to the structure of the neural network. A simple FCNN with 5 layers was utilized for this data set, but with further testing and more specific layers, the FCNN could produce a higher accuracy.

6.2.3.2 Post-Data Balancing

After balancing the dataset the accuracy values for each model differed. The support Vector Machine model still has the highest accuracy of 91%, however the next accurate model was Multinomial Naive Bayes instead of the Gaussian Naive Bayes. The model with the lowest accuracy was initially the TF-IDF Gaussian Naive Bayes model, but post-data balancing the model with the lowest accuracy was the Fully Connected Neural Network.

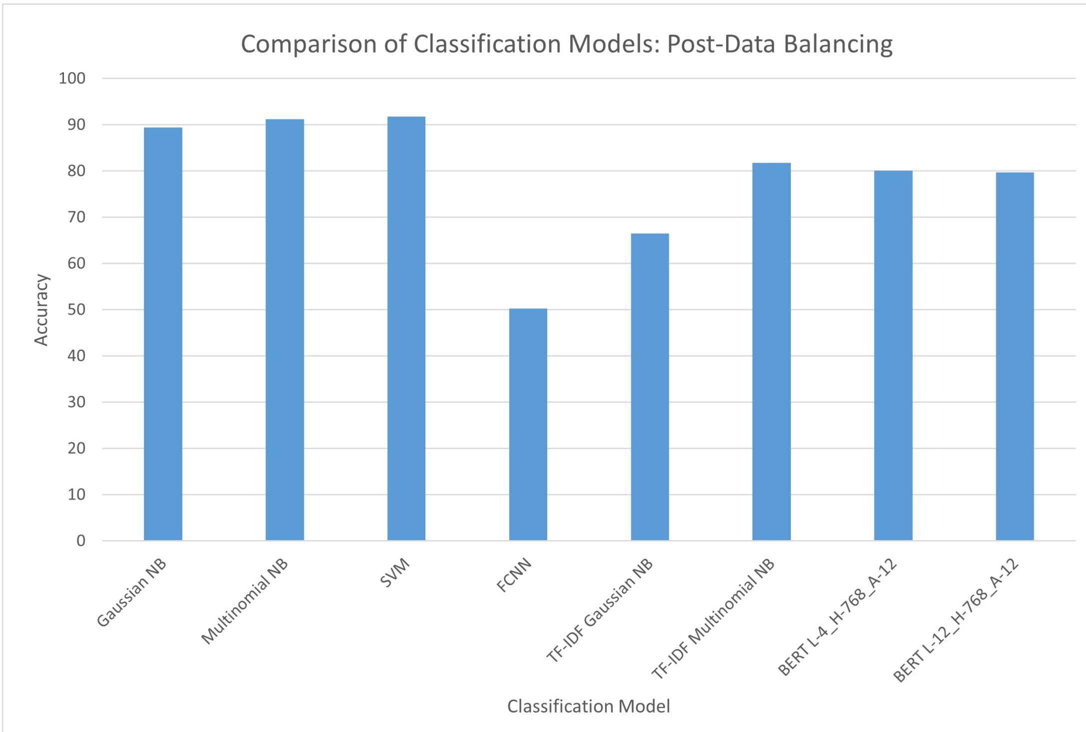

Figure 6.5: Comparison of Classification Models Post-Data Balancing

Figure 6.2.3 below, shows a comparison of all the classifier models before and after balancing the dataset. As mentioned above the accuracies from most models stayed consistent with the exception of the FCNN and the TF-IDF Gaussian Naive Bayes.

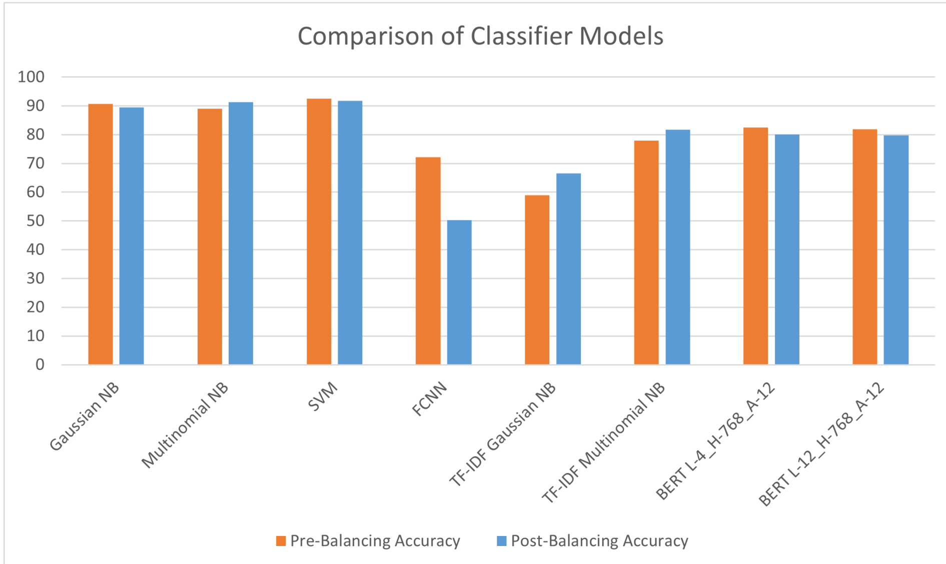

Figure 6.6: Comparison of Classification Models Before and After Data Balancing

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

This research used Natural Language Processing and Machine Learning techniques to analyze testimonies from California Legislative Meetings from 2015-2016 to determine what aspects of a testimony make it successful. In this research a successful testimony is defined as when the testimony's alignment matched the bill outcome (the alignment is "For" and the bill pass or the alignment is "Against" and the bill fails). This was done through feature extraction, implementation of classification models, and feature analysis. Filtering the data and identifying significant features to use was a challenge. However, the results show that certain aspects within a testimony significantly impact the bill outcome.

Based on the feature analysis and classifier results, it is clear that shorter and less complex (easier to comprehend) testimonies are more successful. This can be determined based on the difference in word count, sentence count, flesch, and smog score between successful and unsuccessful testimonies. Successful testimonies tended to have higher flesch score and lower smog score. Unsuccessful testimonies also had a higher word count but similar sentence count as successful testimonies which indicates that unsuccessful testimonies have longer sentences and more run ons.

Additionally, using certain phrases, which can be found in Appendix A, can greatly aid a testimony's success. For example, successful testimonies with an alignment of "For" often contain phrases which explain possible concerns related to the bill or justify why the legislative members should not oppose the bill.

Other features such as proportions of Parts of Speech (POS) are valuable for training the classifier models, however the average values of these features for successful and unsuccessful testimonies are very similar as shown in Table 6.1.3. This indicates that changing the number of nouns, adjectives, or verbs within the testimony does not affect the bill outcome as much as the structure of the testimony. The best performing classifiers, SVM and Gaussian Naive Bayes, were able to predict testimonies' success with an accuracy of about 90%, which means, if given a random testimony and its alignment, this system can extract the required features and determine if the bill will pass or fail with 90% accuracy.

Currently this research provides good insight on what makes a testimony successful, however if expanded upon and further tested, it can become a very powerful tool for lobbyists, legislative members, and general public involved in state legislature.

7.2 Future Work

There are three aspects in which this research can be expanded: feature extraction, classification, and application

Although several features were extracted, there are more which would further provide insight into the data. These include doing detailed sentiment analysis and analyzing sentence structure by creating parse trees for each testimony. A parse tree is an ordered, rooted tree that represents the syntactic structure of a string according to a context-free grammar. The levels of these parse trees can can be compared to indicate the complexity of the testimony [23]. However, generating these parse trees would require a good grammar for all of English which currently is unavailable which makes this a difficult feature to implement. Aside from the given data, it would be interesting to alter the scope of this research and analyze if external factors outside from language, such as time stamps of the speech, gender, position, donations, or organizations play a significant role in the outcome of the bill.

In this research two types of classification models were developed: models that used manually extracted features and models that had their own preprocessing model. In future work, these types of models could be combined to create a more well-rounded model. For example, in addition the the extracted features, TF-IDF vectors could be ingested into the model. The TF-IDF vectors can be weighed lower than the extracted features in order to prevent a bias model.

Given more data, a possible application of this research would be to use the models to predict the outcome for possible testimonies. If a testimony is inputted the model would be able to output if the bill will pass or fail with a certain value of confidence. Additionally, this research focused primarily on California State Legislative data from 2015-2016 meetings. The machine learning techniques implemented can be scaled to more current meetings or even applicable to federal legislative meetings.

BIBLIOGRAPHY

[1] M. G. Ariani, F. Sajedi, and M. Sajedi. Forensic linguistics: A brief overview of the key elements. *Procedia - Social and Behavioral Sciences*, 158:222–225, 2014. 14th Language, Literature and Stylistics Symposium.

[2] Ballotpedia. California open meeting act. 2021.

[3] P. Bhandari. The standard normal distribution. 11 2020.

[4] S. Bird, E. Klein, and E. Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[5] T. Bock. What is a correlation matrix? *Displayr*, 12 2020.

[6] A. Budhwar. Predicting the vote using legislative speech. *California Polytechnic State University San Luis Obispo, ACM Publication, 2018, pp. 1–62.*, pages 1–62, 2018.

[7] G. Chauhan. All about naive bayes. 10 2018.

[8] F. Chollet et al. Keras. 2015.

[9] F. A. COALITION. Brown act primer: Access to meetings. 05 2021.

[10] F. P. P. Commission. Lobbyist rules. 2022.

[11] P. Documentation. collections — container datatypes. 04 2022.

[12] A. F. Gad. Evaluating deep learning models: The confusion matrix, accuracy, precision, and recall. 2020.

[13] R. Gandhi. Support vector machine — introduction to machine learning algorithms. 06 2018.

[14] H. Heraa. Language shaping thought: Can this impact eye-witness testimonies? 04 2019.

[15] M. Honnibal and I. Montani. Spacy: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. 2017.

[16] J. Hunter. Visualization with python. *Matplotlib*, 2021.

[17] IATPP. Digital democracy. 2020.

[18] IATPP. Institute for advanced technology and public policy. 2020.

[19] K. L. K. T. Jacob Devlin, Ming-Wei Chang. Bidirectional encoder representations from transformers (bert). *Transformers for Machine Learning*, pages 43–70, 2022.

[20] S. Jaiswal. Natural language processing dependency parsing. *Medium*, 08 2021.

[21] A. Jones. Google colab: An online jupyter notebook that you should definitely try. 01 2022.

[22] D. A. Kauffman. Learning alignments from legislative discourse. *California Polytechnic State University San Luis Obispo, ACM Publication, 2017, pp. 1–73.*, pages 1–73, 2017.

[23] R. Kline. Parse trees. *Robert Kline Westchester University*, 2018.

[24] E. Loper and S. Bird. Nltk: The natural language toolkit. *CoRR*, 2002.

[25] J. Lucy. Sapir-whorf hypothesis. *International Encyclopedia of the Social Behavioral Sciences*, page 903–906, 03 2015.

[26] P. Mahajan. Fully connected vs convolutional neural networks. 10 2020.

[27] P. B. E. B. Z. C. C. C. Martín Abadi, Ashish Agarwal. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org.

[28] D. G. Maryam Davoodi, Eric Waltenburg. Understanding the language of political agreement and disagreement in legislative texts. *Purdue University , Association for Computational Linguistics, 2020, pp. 1–11.*, pages 1–11, 2020.

[29] W. McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[30] S. McLeod. Loftus and palmer. 2021.

[31] M. Mishra. Convolutional neural networks, explained. 08 2020.

[32] MonkeyLearn. Text classification: What it is and why it matters. 2022.

[33] PacktEditorialStaff. Implementing 3 naive bayes classifiers in scikit-learn. 05 2018.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[35] Readable. Flesch reading ease and the flesch kincaid grade level. 07 2020.

[36] Readable. The gunning fog index. 07 2021.

[37] Readable. The new dale chall readability formula. 07 2021.

[38] Readable. The smog index. 07 2021.

[39] S. S. Difference between t-test and f-test. 05 2022.

[40] A. Saini. Naive bayes algorithm: A complete guide for data science enthusiasts. 09 2021.

[41] B. Scott. How do i decide which readability formula or formulas to use on my document? 2018.

[42] R. W. Shuy. Language in the american courtroom. *Language and Linguistics Compass, vol. 1, no. 1-2*, pages 100–114, 03 2007.

[43] A. Simha. Understanding tf-idf for machine learning. 2021.

[44] M. Thomas, B. Pang, and L. Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327–335, Sydney, Australia, July 2006. Association for Computational Linguistics.

[45] TIBCO. What is a neural network? 2022.

[46] P. P. Tutorials. Part of speech tagging with nltk. 2022.

[47] W. E. Uday Kamath, Kenneth L. Graham. Transformers for machine learning. *Transformers for Machine Learning: A Deep Dive (1st ed.). Chapman and Hall/CRC.*, pages 280–285, 05 2021.

[48] J. Uszkoreit. Transformer: A novel neural network architecture for language understanding. *Google AI Blog*, 08 2017.

[49] S. N. Weber. Cal channel. *Cal Channel: California Secretary of State*, 2022.

APPENDICES

Appendix A

A.1 Feature Set Testing Results

Features: speech, person type, alignment, flesch score, dale chall score, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.80428954	0.8264463	0.96774194
	2	0.20107239	0.8571429	0.01980198
	3	0.83378016	0.8337802	1
	4	0.78284182	0.7956403	0.97986577
Model	**Average**	**0.65549598**	**0.8282524**	**0.74185242**
SVM	**Trial Number**	**Accuracy**	**Precision**	**Recall**
	1	0.80428954	0.8042895	1
	2	0.8310992	0.8310992	1
	3	0.83914209	0.8391421	1
	4	0.75871314	0.7587131	1
	Average	**0.80831099**	**0.808311**	**1**

Table A.1: Feature Testing Results 3

Features: speech, person type, alignment, flesch score, smog score, number of successful phrases, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.79116466	0.8392857	0.92156863
	2	0.7751004	0.7966805	0.96482412
	3	0.77911647	0.8604651	0.88095238
	4	0.75502008	0.8317757	0.87684729
	5	0.76706827	0.8141593	0.92
	Average	**0.77349398**	**0.8284733**	**0.91283848**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
SVM	1	0.80722892	0.8072289	1
	2	0.84337349	0.8433735	1
	3	0.81927711	0.8192771	1
	4	0.81526104	0.815261	1
	5	0.83935743	0.8393574	1
	Average	**0.8248996**	**0.8248996**	**1**

Table A.2: Feature Testing Results 4

Features = text, persontype, alignment, word_count, sentence_count, avg_sentence_length, flesch, smog, successful_phrases, DT, JJ, NN,PRP, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.70281124	0.8424658	0.70689655
	2	0.73895582	0.8896552	0.7247191
	3	0.69076305	0.8055556	0.7030303
	Average	**0.71084337**	**0.8458922**	**0.71154865**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
SVM	1	0.79919679	0.8448276	0.86470588
	2	0.77911647	0.8121547	0.875
	3	0.7751004	0.7853107	0.88535032
	Average	**0.78447122**	**0.8140977**	**0.87501873**
Model	**Trial Number**	**Accuracy**		
FCNN	1	68.67		
	2	68.67		
	3	68.67		
	Average	**68.67**		

Table A.3: Feature Testing Results 5

Features = text, persontype, alignment, word_count, sentence_count, avg_sentence_length, flesch, smog, successful_phrases, DT, JJ, NN, PRP, VB, outcome				
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
Gaussian Naïve Bayes	1	0.69879518	0.8368794	0.69411765
	2	0.68674699	0.8682171	0.64739884
	3	0.69477912	0.8529412	0.6744186
	Average	**0.69344043**	**0.8526792**	**0.67197837**
Model	**Trial Number**	**Accuracy**	**Precision**	**Recall**
SVM	1	0.84738956	0.8914286	0.89142857
	2	0.79116466	0.8112245	0.9137931
	3	0.80722892	0.8255814	0.8875
	Average	**0.81526104**	**0.8427448**	**0.89757389**
Model	**Trial Number**	**Accuracy**		
FCNN	1	68.67		
	2	68.67		
	3	68.67		
	Average	**68.67**		

Table A.4: Feature Testing Results 6

A.2 Successful Phrases

Successful 3 word phrases for testimonies with "For" or "Indeterminate" alignment:

3 word phrase (trigram)	Success Rate
('reasons', 'were', 'opposed')	1
('opposition', 'mr', 'person')	1
('in', 'its', 'current')	1
('we', 'remain', 'opposed')	1
('proponents', 'of', 'the')	1
('opposition', 'person', 'on')	1
('dont', 'think', 'this')	1
('are', 'strongly', 'opposed')	1
('date', 'in', 'opposition')	1
('associated', 'with', 'it')	1
('you', 'to', 'oppose')	1
('an', 'oppose', 'position')	1
('we', 'strongly', 'oppose')	1
('our', 'opposition', 'thank')	1
('were', 'in', 'opposition')	1
('that', 'we', 'oppose')	1
('amendments', 'that', 'are')	1
('the', 'proponents', 'of')	1
('unless', 'amended', 'to')	1
('were', 'not', 'sure')	1
('association', 'also', 'opposed')	1
('the', 'way', 'this')	1
('we', 'oppose', 'this')	1
('to', 'continuing', 'those')	1
('lot', 'of', 'concerns')	1
('cooling', 'contractors', 'of')	1
('orientation', 'change', 'efforts')	1
('so', 'were', 'concerned')	1
('share', 'the', 'concerns')	1
('reasons', 'we', 'oppose')	1
('reasons', 'we', 'remain')	1
('the', 'public', 'entities')	1
('to', 'vote', 'no')	1
('position', 'at', 'this')	1
('oppose', 'this', 'measure')	1
('there', 'hasnt', 'been')	1
('strongly', 'opposed', 'to')	1
('its', 'present', 'form')	1
('the', 'same', 'concerns')	1
('in', 'opposition', 'i')	1
('technology', 'association', 'also')	1
('regional', 'apartment', 'associations')	1
('org', 'in', 'respectful')	1
('represent', 'cardinal', 'regional')	1
('opposed', 'for', 'the')	1

('its', 'current', 'form')	1
('are', 'still', 'opposed')	1
('cardinal', 'regional', 'apartment')	1
('in', 'its', 'present')	1
('our', 'concerns', 'are')	1
('sexual', 'orientation', 'change')	1
('org', 'we', 'oppose')	1
('bill', 'is', 'currently')	1
('we', 'are', 'neutral')	1
('some', 'concerns', 'with')	0.928571429
('our', 'concerns', 'and')	0.928571429
('have', 'a', 'position')	0.923076923
('neutral', 'on', 'the')	0.909090909
('a', 'position', 'on')	0.9
('the', 'victims', 'and')	0.888888889
('senator', 'person', 'person')	0.888888889
('comments', 'about', 'the')	0.888888889
('position', 'on', 'the')	0.875
('forward', 'to', 'continuing')	0.875
('look', 'at', 'it')	0.857142857
('like', 'to', 'work')	0.857142857
('concerns', 'about', 'the')	0.857142857
('policies', 'and', 'procedures')	0.857142857
('of', 'which', 'is')	0.833333333
('with', 'us', 'we')	0.833333333
('the', 'property', 'owner')	0.833333333
('going', 'to', 'go')	0.833333333
('consistent', 'with', 'the')	0.833333333
('opposition', 'thank', 'you')	0.833333333
('on', 'the', 'bill')	0.833333333
('be', 'made', 'in')	0.8
('the', 'rental', 'housing')	0.8
('with', 'the', 'ordinal')	0.8
('official', 'position', 'on')	0.8
('take', 'into', 'account')	0.8
('org', 'we', 'appreciate')	0.8
('not', 'have', 'a')	0.8
('it', 'with', 'the')	0.8
('to', 'continue', 'working')	0.8
('also', 'need', 'to')	0.8
('and', 'the', 'staff')	0.8
('want', 'to', 'speak')	0.8
('org', 'we', 'have')	0.8
('and', 'the', 'sponsors')	0.8
('org', 'we', 'dont')	0.8
('a', 'letter', 'of')	0.8
('forward', 'to', 'continue')	0.8

('of', 'that', 'is')	0.8
('with', 'org', 'with')	0.8
('all', 'the', 'work')	0.8
('the', 'authors', 'staff')	0.8
('have', 'an', 'official')	0.8
('and', 'many', 'others')	0.8
('the', 'same', 'rules')	0.75
('gpe', 'manufacturers', 'and')	0.75
('office', 'and', 'the')	0.75
('in', 'opposition', 'person')	0.75
('that', 'is', 'a')	0.75
('this', 'bill', 'will')	0.75
('we', 'share', 'the')	0.75
('want', 'to', 'work')	0.75
('to', 'find', 'a')	0.75
('are', 'a', 'lot')	0.75
('i', 'speak', 'to')	0.75
('with', 'the', 'sponsor')	0.75
('is', 'the', 'ordinal')	0.75
('in', 'the', 'way')	0.75
('dont', 'have', 'an')	0.75
('with', 'mr', 'persons')	0.75
('a', 'public', 'nuisance')	0.75
('and', 'technology', 'association')	0.75
('manufacturers', 'and', 'technology')	0.75
('the', 'bill', 'at')	0.75
('position', 'but', 'we')	0.75
('a', 'neutral', 'position')	0.75
('so', 'we', 'will')	0.75
('authors', 'office', 'to')	0.75
('her', 'staff', 'and')	0.75
('as', 'i', 'know')	0.75
('however', 'we', 'do')	0.75
('person', 'person', 'on')	0.75
('get', 'it', 'right')	0.75
('concerns', 'and', 'we')	0.75
('we', 'are', 'talking')	0.75
('taken', 'a', 'position')	0.75
('so', 'im', 'going')	0.75
('that', 'is', 'in')	0.75
('this', 'bill', 'this')	0.75
('youre', 'not', 'going')	0.75
('here', 'date', 'and')	0.75
('time', 'but', 'we')	0.75
('the', 'amendments', 'but')	0.75
('with', 'this', 'org')	0.75
('to', 'oppose', 'this')	0.75

('and', 'staff', 'for')	0.75
('if', 'there', 'are')	0.75
('committed', 'to', 'working')	0.75
('we', 'appreciate', 'all')	0.75
('authors', 'staff', 'and')	0.75
('the', 'bill', 'the')	0.75
('hope', 'that', 'the')	0.75
('and', 'in', 'the')	0.75
('they', 'were', 'not')	0.666666667
('im', 'person', 'here')	0.666666667
('i', 'did', 'want')	0.666666667
('affected', 'by', 'this')	0.666666667
('the', 'very', 'least')	0.666666667
('the', 'author', 'the')	0.666666667
('around', 'for', 'date')	0.666666667
('the', 'state', 'can')	0.666666667
('the', 'chief', 'probation')	0.666666667
('the', 'staff', 'and')	0.666666667
('of', 'the', 'amendments')	0.666666667
('associated', 'with', 'that')	0.666666667
('issues', 'with', 'the')	0.666666667
('the', 'bill', 'but')	0.666666667
('to', 'the', 'authors')	0.666666667
('to', 'ask', 'you')	0.666666667
('go', 'back', 'to')	0.666666667
('from', 'org', 'to')	0.666666667
('to', 'continuing', 'to')	0.666666667
('and', 'i', 'believe')	0.666666667
('that', 'there', 'is')	0.666666667
('for', 'date', 'in')	0.666666667
('the', 'concept', 'of')	0.666666667
('our', 'opposition', 'and')	0.666666667
('do', 'have', 'some')	0.666666667
('org', 'weve', 'been')	0.666666667
('there', 'may', 'be')	0.666666667
('because', 'there', 'is')	0.666666667
('this', 'issue', 'and')	0.666666667
('position', 'on', 'this')	0.666666667
('we', 'will', 'be')	0.666666667
('the', 'gpe', 'manufacturers')	0.666666667
('is', 'that', 'if')	0.666666667
('have', 'a', 'formal')	0.666666667
('my', 'colleague', 'from')	0.666666667
('to', 'understand', 'the')	0.666666667
('that', 'org', 'and')	0.666666667
('to', 'ask', 'the')	0.666666667
('given', 'the', 'fact')	0.666666667

('and', 'then', 'theres')	0.666666667
('of', 'org', 'ill')	0.666666667
('go', 'out', 'and')	0.666666667
('be', 'used', 'for')	0.666666667
('an', 'opposed', 'position')	0.666666667
('bill', 'this', 'is')	0.666666667
('it', 'in', 'a')	0.666666667
('cap', 'and', 'trade')	0.666666667
('how', 'this', 'bill')	0.666666667
('components', 'of', 'the')	0.666666667
('this', 'time', 'but')	0.666666667
('it', 'would', 'be')	0.666666667
('that', 'this', 'bill')	0.666666667
('it', 'was', 'date')	0.666666667
('probation', 'officers', 'of')	0.666666667
('forward', 'with', 'the')	0.666666667
('the', 'authors', 'intent')	0.666666667
('and', 'so', 'again')	0.666666667
('do', 'have', 'concerns')	0.666666667
('we', 'do', 'have')	0.666666667
('about', 'how', 'the')	0.666666667
('but', 'we', 'also')	0.666666667
('bill', 'does', 'not')	0.666666667
('we', 'are', 'trying')	0.666666667
('or', 'not', 'the')	0.666666667
('would', 'still', 'be')	0.666666667
('in', 'which', 'to')	0.666666667
('i', 'believe', 'that')	0.666666667
('something', 'that', 'would')	0.666666667
('in', 'the', 'previous')	0.666666667
('org', 'will', 'be')	0.666666667
('continuing', 'to', 'work')	0.666666667
('of', 'the', 'chief')	0.666666667
('these', 'kinds', 'of')	0.666666667
('is', 'required', 'to')	0.666666667
('opposed', 'to', 'this')	0.666666667
('to', 'this', 'bill')	0.666666667
('end', 'up', 'in')	0.666666667
('have', 'an', 'oppose')	0.666666667
('appreciate', 'all', 'the')	0.666666667
('to', 'do', 'more')	0.666666667
('there', 'is', 'the')	0.666666667
('things', 'that', 'the')	0.666666667
('want', 'to', 'start')	0.666666667
('on', 'thank', 'you')	0.666666667
('bill', 'will', 'be')	0.666666667
('would', 'agree', 'with')	0.666666667

('the', 'author', 'has')	0.666666667
('required', 'by', 'law')	0.666666667
('able', 'to', 'support')	0.666666667
('many', 'of', 'which')	0.666666667
('to', 'put', 'the')	0.666666667
('but', 'when', 'you')	0.666666667
('on', 'to', 'the')	0.666666667
('out', 'of', 'org')	0.666666667
('as', 'has', 'been')	0.666666667
('time', 'and', 'the')	0.666666667
('but', 'at', 'this')	0.666666667
('the', 'resources', 'that')	0.666666667
('that', 'weve', 'had')	0.666666667
('bill', 'at', 'this')	0.666666667
('want', 'to', 'protect')	0.666666667
('from', 'gpe', 'to')	0.666666667
('can', 'get', 'a')	0.666666667
('moving', 'forward', 'and')	0.666666667
('down', 'the', 'road')	0.666666667
('have', 'an', 'issue')	0.666666667
('to', 'us', 'to')	0.666666667
('back', 'to', 'the')	0.666666667
('even', 'if', 'the')	0.666666667
('due', 'to', 'the')	0.666666667
('to', 'apply', 'to')	0.666666667
('the', 'sort', 'of')	0.666666667
('none', 'of', 'that')	0.666666667
('chief', 'probation', 'officers')	0.666666667
('be', 'clear', 'about')	0.666666667
('they', 'can', 'be')	0.666666667
('be', 'removing', 'our')	0.666666667
('the', 'bill', 'weve')	0.666666667
('definition', 'of', 'what')	0.666666667
('the', 'sponsor', 'and')	0.666666667
('gpe', 'to', 'gpe')	0.666666667
('cardinal', 'of', 'which')	0.666666667
('them', 'in', 'the')	0.666666667
('hard', 'work', 'and')	0.666666667
('reality', 'is', 'that')	0.666666667
('we', 'can', 'get')	0.666666667
('that', 'with', 'the')	0.666666667
('we', 'do', 'see')	0.666666667
('been', 'in', 'the')	0.666666667
('in', 'the', 'committee')	0.666666667
('the', 'committee', 'analysis')	0.666666667
('there', 'is', 'nothing')	0.666666667
('of', 'org', 'just')	0.666666667

('and', 'we', 'appreciate')	0.666666667
('that', 'i', 'know')	0.666666667
('number', 'of', 'issues')	0.666666667
('opposition', 'to', 'gpe')	0.666666667
('i', 'understand', 'the')	0.666666667
('the', 'org', 'staff')	0.666666667
('what', 'the', 'author')	0.666666667
('with', 'org', 'this')	0.666666667
('safety', 'and', 'reliability')	0.666666667
('there', 'is', 'a')	0.666666667
('and', 'the', 'authors')	0.666666667
('have', 'a', 'problem')	0.666666667
('to', 'take', 'out')	0.666666667
('by', 'this', 'bill')	0.666666667
('the', 'analysis', 'that')	0.666666667
('for', 'listening', 'to')	0.666666667
('and', 'trade', 'program')	0.666666667
('work', 'with', 'us')	0.666666667
('officers', 'of', 'gpe')	0.666666667
('at', 'a', 'minimum')	0.666666667
('agree', 'with', 'mr')	0.666666667
('that', 'were', 'not')	0.666666667
('on', 'this', 'issue')	0.666666667
('the', 'only', 'thing')	0.666666667
('are', 'trying', 'to')	0.666666667
('an', 'issue', 'with')	0.666666667
('thats', 'what', 'were')	0.666666667
('this', 'opportunity', 'to')	0.666666667
('org', 'i', 'just')	0.666666667
('where', 'they', 'can')	0.666666667
('work', 'with', 'them')	0.666666667
('in', 'opposition', 'we')	0.666666667
('at', 'all', 'of')	0.666666667
('the', 'mental', 'health')	0.666666667
('out', 'that', 'there')	0.666666667
('speak', 'to', 'that')	0.666666667
('concerns', 'with', 'the')	0.666666667
('the', 'sponsors', 'and')	0.666666667
('think', 'that', 'there')	0.666666667
('greatly', 'appreciate', 'the')	0.666666667
('so', 'wed', 'like')	0.6
('and', 'i', 'agree')	0.571428571
('a', 'cardinal', 'vote')	0.5
('what', 'i', 'was')	0.5
('with', 'us', 'and')	0.5
('the', 'opportunity', 'for')	0.5
('put', 'in', 'place')	0.5

('this', 'and', 'i')	0.5
('of', 'all', 'of')	0.5
('working', 'with', 'our')	0.5
('in', 'the', 'commission')	0.5
('but', 'its', 'not')	0.5
('with', 'our', 'members')	0.5
('with', 'norp', 'for')	0.5
('org', 'i', 'appreciate')	0.5
('we', 'dont', 'think')	0.5
('county', 'and', 'municipal')	0.5
('all', 'of', 'those')	0.5
('concern', 'is', 'the')	0.5
('in', 'the', 'position')	0.5
('when', 'we', 'have')	0.5
('i', 'would', 'suggest')	0.5
('is', 'the', 'time')	0.5
('come', 'up', 'with')	0.5
('for', 'the', 'public')	0.5
('in', 'our', 'communities')	0.5
('we', 'have', 'some')	0.5
('issues', 'that', 'were')	0.5
('this', 'bill', 'to')	0.5
('passed', 'in', 'date')	0.5
('and', 'that', 'it')	0.5
('as', 'a', 'cosponsor')	0.5
('in', 'date', 'is')	0.5
('as', 'it', 'was')	0.5
('no', 'way', 'for')	0.5
('org', 'that', 'is')	0.5
('sort', 'of', 'the')	0.5
('are', 'not', 'being')	0.5
('to', 'come', 'up')	0.5
('the', 'public', 'interest')	0.5
('part', 'of', 'their')	0.5
('most', 'of', 'these')	0.5
('and', 'municipal', 'employees')	0.5
('be', 'part', 'of')	0.5
('amendments', 'that', 'the')	0.5
('is', 'a', 'concern')	0.5
('to', 'go', 'down')	0.5
('staff', 'and', 'we')	0.5
('for', 'those', 'who')	0.5
('you', 'would', 'be')	0.5
('to', 'say', 'the')	0.5
('really', 'appreciate', 'the')	0.5
('and', 'we', 'oppose')	0.5
('talk', 'about', 'this')	0.5

('gpe', 'org', 'i')	0.5
('date', 'before', 'the')	0.5
('have', 'the', 'ability')	0.5
('a', 'few', 'things')	0.5
('date', 'so', 'if')	0.5
('do', 'support', 'the')	0.5
('to', 'make', 'these')	0.5
('a', 'chilling', 'effect')	0.5
('in', 'the', 'statute')	0.5
('we', 'would', 'encourage')	0.5
('bill', 'moves', 'forward')	0.5
('because', 'that', 'is')	0.5
('at', 'all', 'and')	0.5
('here', 'at', 'the')	0.5
('but', 'we', 'dont')	0.5
('clear', 'and', 'convincing')	0.5
('and', 'we', 'hope')	0.5
('working', 'with', 'org')	0.5
('with', 'some', 'of')	0.5
('to', 'the', 'city')	0.5
('the', 'org', 'did')	0.5
('that', 'you', 'may')	0.5
('then', 'the', 'other')	0.5
('that', 'the', 'author')	0.5
('that', 'need', 'to')	0.5
('as', 'weve', 'seen')	0.5
('law', 'enforcement', 'agency')	0.5
('to', 'meet', 'our')	0.5
('well', 'i', 'think')	0.5
('appreciate', 'senator', 'persons')	0.5
('that', 'these', 'are')	0.5
('and', 'then', 'the')	0.5
('from', 'the', 'org')	0.5
('not', 'be', 'the')	0.5
('we', 'just', 'think')	0.5
('the', 'amendments', 'date')	0.5
('to', 'talk', 'to')	0.5
('the', 'language', 'that')	0.5
('as', 'to', 'how')	0.5
('like', 'to', 'start')	0.5
('can', 'do', 'it')	0.5
('the', 'state', 'the')	0.5
('can', 'get', 'to')	0.5
('be', 'a', 'lot')	0.5
('programs', 'that', 'are')	0.5
('do', 'not', 'know')	0.5
('are', 'based', 'on')	0.5

('to', 'be', 'working')	0.5
('we', 'do', 'want')	0.5
('date', 'we', 'are')	0.5
('again', 'we', 'have')	0.5
('done', 'by', 'org')	0.5
('local', 'elected', 'officials')	0.5
('say', 'a', 'few')	0.5
('skilled', 'and', 'trained')	0.5
('of', 'those', 'cardinal')	0.5
('because', 'it', 'makes')	0.5
('date', 'from', 'the')	0.5
('a', 'firearm', 'in')	0.5
('as', 'to', 'the')	0.5
('and', 'our', 'members')	0.5
('staff', 'for', 'their')	0.5
('thank', 'you', 'the')	0.5
('and', 'this', 'was')	0.5
('this', 'bill', 'should')	0.5
('and', 'org', 'has')	0.5
('we', 'are', 'doing')	0.5
('requirement', 'that', 'the')	0.5
('we', 'continue', 'to')	0.5
('depending', 'on', 'how')	0.5
('well', 'be', 'able')	0.5
('for', 'public', 'safety')	0.5
('it', 'should', 'be')	0.5
('you', 'want', 'to')	0.5
('if', 'youre', 'going')	0.5
('if', 'you', 'are')	0.5
('most', 'of', 'the')	0.5
('to', 'start', 'by')	0.5
('who', 'has', 'a')	0.5
('we', 'have', 'had')	0.5
('i', 'tried', 'to')	0.5
('hard', 'on', 'this')	0.5
('advantage', 'of', 'the')	0.5
('know', 'that', 'in')	0.5
('to', 'provide', 'these')	0.5
('that', 'were', 'just')	0.5
('it', 'is', 'a')	0.5
('person', 'here', 'on')	0.5
('willing', 'to', 'work')	0.5
('that', 'we', 'really')	0.5
('for', 'their', 'work')	0.5
('for', 'our', 'members')	0.5
('take', 'advantage', 'of')	0.5
('to', 'date', 'and')	0.5

('is', 'the', 'problem')	0.5
('removed', 'from', 'the')	0.5
('mr', 'person', 'and')	0.5
('in', 'the', 'room')	0.5
('goal', 'of', 'the')	0.5
('you', 'mr', 'person')	0.5
('that', 'to', 'be')	0.5
('but', 'the', 'bill')	0.5
('on', 'the', 'record')	0.5
('from', 'org', 'we')	0.5
('the', 'removal', 'of')	0.5
('process', 'thank', 'you')	0.5
('we', 'have', 'concerns')	0.5
('the', 'fact', 'that')	0.5
('person', 'mentioned', 'the')	0.5
('date', 'gpe', 'has')	0.5
('thank', 'org', 'member')	0.5
('for', 'individuals', 'who')	0.5
('the', 'data', 'and')	0.5
('do', 'that', 'i')	0.5
('person', 'person', 'org')	0.5
('the', 'information', 'that')	0.5
('the', 'same', 'reasons')	0.5
('of', 'org', 'that')	0.5
('if', 'you', 'want')	0.5
('concerns', 'raised', 'by')	0.5
('we', 'just', 'need')	0.5
('language', 'of', 'the')	0.5
('respectfully', 'oppose', 'this')	0.5
('and', 'so', 'those')	0.5
('dont', 'have', 'a')	0.5
('the', 'issue', 'that')	0.5
('org', 'also', 'opposed')	0.5
('the', 'hands', 'of')	0.5
('so', 'in', 'the')	0.5
('from', 'the', 'bill')	0.5
('so', 'all', 'of')	0.5
('of', 'gpes', 'org')	0.5
('on', 'both', 'sides')	0.5
('date', 'and', 'this')	0.5
('the', 'area', 'of')	0.5
('to', 'take', 'care')	0.5
('the', 'amendments', 'we')	0.5
('subject', 'to', 'a')	0.5
('does', 'not', 'have')	0.5
('of', 'gpe', 'gpe')	0.5
('with', 'the', 'org')	0.5

('gpe', 'we', 'do')	0.5
('org', 'with', 'the')	0.5
('of', 'the', 'norp')	0.5
('if', 'we', 'had')	0.5
('goal', 'is', 'to')	0.5
('appreciative', 'of', 'the')	0.5
('by', 'my', 'colleagues')	0.5
('i', 'think', 'there')	0.5
('that', 'senator', 'person')	0.5
('we', 'do', 'this')	0.5
('with', 'org', 'for')	0.5
('that', 'if', 'there')	0.5
('under', 'the', 'org')	0.5
('to', 'do', 'for')	0.5
('moves', 'forward', 'thank')	0.5
('the', 'answer', 'is')	0.5
('the', 'next', 'generation')	0.5
('org', 'we', 'also')	0.5
('were', 'concerned', 'that')	0.5
('of', 'the', 'orgs')	0.5
('and', 'we', 'believe')	0.5
('this', 'bill', 'creates')	0.5
('the', 'age', 'of')	0.5
('the', 'ordinal', 'time')	0.5
('lot', 'of', 'the')	0.5
('to', 'remove', 'our')	0.5
('and', 'there', 'is')	0.5
('a', 'staff', 'attorney')	0.5
('part', 'of', 'our')	0.5
('we', 'are', 'looking')	0.5
('set', 'of', 'amendments')	0.5
('the', 'purpose', 'of')	0.5
('on', 'the', 'board')	0.5
('we', 'think', 'the')	0.5
('to', 'take', 'advantage')	0.5
('and', 'we', 'dont')	0.5
('the', 'terms', 'of')	0.5
('the', 'commission', 'of')	0.5
('of', 'issues', 'that')	0.5
('is', 'likely', 'to')	0.5
('a', 'date', 'notice')	0.5
('we', 'can', 'come')	0.5
('have', 'concerns', 'with')	0.5
('improve', 'the', 'health')	0.5
('this', 'measure', 'we')	0.5
('on', 'the', 'org')	0.5
('any', 'type', 'of')	0.5

('that', 'were', 'trying')	0.5
('that', 'theres', 'no')	0.5
('other', 'forms', 'of')	0.5
('way', 'that', 'the')	0.5
('and', 'working', 'with')	0.5
('of', 'the', 'urban')	0.5
('no', 'place', 'to')	0.5
('it', 'needs', 'to')	0.5
('committee', 'person', 'on')	0.5
('so', 'when', 'you')	0.5
('so', 'if', 'the')	0.5
('that', 'org', 'will')	0.5
('in', 'opposition', 'to')	0.5
('not', 'only', 'the')	0.5
('of', 'gpe', 'also')	0.5
('they', 'have', 'the')	0.5
('it', 'is', 'also')	0.5
('one', 'of', 'them')	0.5
('to', 'come', 'out')	0.5
('by', 'the', 'proponents')	0.5
('why', 'we', 'have')	0.5
('in', 'the', 'language')	0.5
('going', 'forward', 'we')	0.5
('we', 'will', 'not')	0.5
('with', 'the', 'authors')	0.5
('the', 'idea', 'of')	0.5
('dont', 'have', 'to')	0.5
('and', 'as', 'a')	0.5
('not', 'the', 'way')	0.5
('and', 'then', 'if')	0.5
('been', 'around', 'for')	0.5
('gpe', 'chapters', 'of')	0.5
('proud', 'to', 'sponsor')	0.5
('with', 'my', 'colleague')	0.5
('the', 'other', 'issue')	0.5
('industry', 'and', 'we')	0.5
('thank', 'you', 'mr')	0.5
('but', 'i', 'dont')	0.5
('do', 'not', 'have')	0.5
('its', 'just', 'a')	0.5
('to', 'take', 'into')	0.5
('sponsor', 'and', 'the')	0.5
('associated', 'with', 'the')	0.5
('that', 'we', 'cant')	0.5
('have', 'to', 'go')	0.5
('that', 'we', 'want')	0.5
('so', 'there', 'are')	0.5

('org', 'staff', 'for')	0.5
('about', 'the', 'fact')	0.5
('think', 'there', 'are')	0.5
('working', 'closely', 'with')	0.5
('just', 'say', 'that')	0.5
('and', 'we', 'understand')	0.5
('if', 'we', 'were')	0.5
('with', 'it', 'and')	0.5
('we', 'also', 'are')	0.5
('i', 'think', 'as')	0.5
('the', 'effects', 'of')	0.5
('but', 'it', 'is')	0.5
('to', 'thank', 'org')	0.5
('and', 'id', 'like')	0.5
('to', 'see', 'if')	0.5
('in', 'other', 'words')	0.5
('in', 'the', 'hands')	0.5
('process', 'in', 'place')	0.5
('remind', 'you', 'that')	0.5
('we', 'have', 'with')	0.5
('that', 'as', 'the')	0.5
('want', 'to', 'just')	0.5
('person', 'who', 'has')	0.5
('regardless', 'of', 'their')	0.5
('need', 'to', 'get')	0.5
('up', 'on', 'the')	0.5
('bill', 'so', 'that')	0.5
('able', 'to', 'come')	0.5
('with', 'org', 'in')	0.5
('out', 'of', 'compliance')	0.5
('going', 'to', 'come')	0.5
('address', 'our', 'concerns')	0.5
('of', 'the', 'public')	0.5
('we', 'want', 'to')	0.5
('whats', 'going', 'on')	0.5
('bill', 'because', 'of')	0.5
('im', 'sure', 'that')	0.5
('i', 'agree', 'with')	0.5
('in', 'gpe', 'since')	0.5
('nothing', 'in', 'this')	0.5
('theyre', 'not', 'going')	0.5
('so', 'while', 'we')	0.5
('opposition', 'to', 'org')	0.5
('an', 'amendment', 'that')	0.5
('we', 'greatly', 'appreciate')	0.5
('for', 'a', 'lot')	0.5
('org', 'and', 'it')	0.5

('information', 'about', 'the')	0.5
('a', 'lot', 'of')	0.5
('also', 'on', 'the')	0.5
('time', 'org', 'members')	0.5
('appreciate', 'the', 'work')	0.5
('date', 'in', 'order')	0.5
('morning', 'mr', 'chair')	0.5
('in', 'support', 'but')	0.5
('thank', 'you', 'i')	0.5
('to', 'meet', 'that')	0.5
('know', 'how', 'much')	0.5
('of', 'the', 'data')	0.5
('members', 'i', 'am')	0.5
('as', 'one', 'of')	0.5
('org', 'to', 'the')	0.5
('or', 'not', 'we')	0.5
('here', 'we', 'are')	0.5
('with', 'a', 'cardinal')	0.5
('we', 'are', 'proud')	0.5
('in', 'support', 'also')	0.5
('but', 'for', 'the')	0.5
('staff', 'attorney', 'at')	0.5
('his', 'leadership', 'on')	0.5
('bill', 'is', 'an')	0.5
('in', 'cardinal', 'of')	0.5
('cardinal', 'of', 'all')	0.5
('version', 'of', 'the')	0.5
('they', 'would', 'not')	0.5
('with', 'that', 'as')	0.5
('the', 'time', 'and')	0.5
('person', 'if', 'i')	0.5
('to', 'the', 'measure')	0.5
('dont', 'think', 'it')	0.5
('and', 'can', 'be')	0.5
('from', 'being', 'a')	0.5
('is', 'a', 'process')	0.5
('apply', 'to', 'all')	0.5
('that', 'can', 'be')	0.5
('according', 'to', 'the')	0.5
('strongly', 'believe', 'that')	0.5
('is', 'a', 'lot')	0.5
('the', 'cap', 'and')	0.5
('reduce', 'greenhouse', 'gas')	0.5
('member', 'person', 'for')	0.5
('to', 'try', 'to')	0.5
('thats', 'the', 'problem')	0.5
('looking', 'for', 'a')	0.5

('what', 'i', 'want')	0.5
('amend', 'the', 'bill')	0.5
('get', 'to', 'a')	0.5
('of', 'care', 'and')	0.5
('org', 'just', 'want')	0.5
('cardinal', 'or', 'cardinal')	0.5
('to', 'sponsor', 'this')	0.5
('of', 'that', 'information')	0.5
('if', 'the', 'bill')	0.5
('as', 'well', 'i')	0.5
('in', 'the', 'same')	0.5
('bill', 'but', 'we')	0.5
('the', 'money', 'to')	0.5
('the', 'collective', 'bargaining')	0.5
('working', 'on', 'this')	0.5
('to', 'have', 'that')	0.5
('heard', 'a', 'lot')	0.5
('may', 'be', 'a')	0.5
('this', 'kind', 'of')	0.5
('the', 'cause', 'of')	0.5
('bill', 'would', 'require')	0.5
('we', 'dont', 'need')	0.5
('operate', 'in', 'gpe')	0.5
('on', 'a', 'regular')	0.5
('the', 'grid', 'and')	0.5
('bill', 'this', 'bill')	0.5
('think', 'the', 'bill')	0.5
('to', 'serve', 'the')	0.5
('think', 'that', 'is')	0.5
('like', 'to', 'have')	0.5
('for', 'org', 'were')	0.5
('of', 'the', 'cardinal')	0.5
('of', 'their', 'income')	0.5
('with', 'the', 'amendments')	0.5
('and', 'that', 'would')	0.5
('of', 'it', 'and')	0.5
('the', 'folks', 'who')	0.5
('to', 'see', 'the')	0.5
('align', 'myself', 'with')	0.5
('of', 'people', 'that')	0.5
('in', 'gpe', 'because')	0.5
('colleague', 'from', 'org')	0.5
('other', 'parts', 'of')	0.5
('the', 'reason', 'why')	0.5
('for', 'date', 'im')	0.5
('trying', 'to', 'figure')	0.5
('a', 'chance', 'to')	0.5

('were', 'also', 'concerned')	0.5
('to', 'the', 'local')	0.5
('people', 'who', 'dont')	0.5
('those', 'are', 'our')	0.5
('in', 'date', 'a')	0.5
('bill', 'that', 'we')	0.5
('hopeful', 'that', 'we')	0.5
('i', 'wanted', 'to')	0.5
('in', 'gpe', 'that')	0.5
('dont', 'have', 'that')	0.5
('that', 'it', 'will')	0.5
('to', 'org', 'the')	0.5
('the', 'definition', 'of')	0.5
('we', 'too', 'are')	0.5
('youre', 'going', 'to')	0.5
('its', 'a', 'very')	0.5
('but', 'there', 'are')	0.5
('here', 'is', 'that')	0.5
('we', 'think', 'there')	0.5
('in', 'the', 'area')	0.5
('actually', 'have', 'a')	0.5
('to', 'put', 'a')	0.5
('the', 'way', 'the')	0.5
('will', 'be', 'a')	0.5
('a', 'condition', 'of')	0.5
('there', 'thank', 'you')	0.5
('as', 'i', 'understand')	0.5
('that', 'we', 'are')	0.5
('because', 'there', 'are')	0.5
('leadership', 'on', 'this')	0.5
('sure', 'that', 'they')	0.5
('with', 'the', 'local')	0.5
('of', 'my', 'colleagues')	0.5
('if', 'we', 'can')	0.5
('made', 'by', 'org')	0.5
('conversations', 'with', 'the')	0.5
('in', 'place', 'that')	0.5
('they', 'did', 'not')	0.5
('something', 'that', 'i')	0.5
('to', 'acknowledge', 'the')	0.5
('this', 'area', 'and')	0.5
('oppose', 'this', 'bill')	0.5
('are', 'things', 'that')	0.5
('and', 'i', 'work')	0.5
('that', 'if', 'we')	0.5
('on', 'this', 'we')	0.5
('of', 'them', 'and')	0.5

('is', 'very', 'very')	0.5
('asking', 'for', 'a')	0.5
('answer', 'to', 'that')	0.5
('also', 'would', 'like')	0.5
('heard', 'from', 'the')	0.5
('that', 'it', 'was')	0.5
('we', 'all', 'agree')	0.5
('across', 'the', 'nation')	0.5
('that', 'would', 'require')	0.5
('have', 'a', 'concern')	0.5
('under', 'the', 'federal')	0.5
('is', 'the', 'org')	0.5
('who', 'has', 'been')	0.5
('removing', 'our', 'opposition')	0.5
('moving', 'forward', 'with')	0.5
('executive', 'director', 'for')	0.5
('this', 'issue', 'for')	0.5
('and', 'in', 'this')	0.5
('are', 'eligible', 'for')	0.5
('thank', 'you', 'chairman')	0.5
('with', 'the', 'sponsors')	0.5
('should', 'be', 'given')	0.5
('of', 'these', 'things')	0.5
('how', 'many', 'people')	0.5
('org', 'to', 'do')	0.5
('to', 'this', 'measure')	0.5
('you', 'to', 'vote')	0.5
('law', 'enforcement', 'officers')	0.5
('it', 'has', 'not')	0.5
('the', 'most', 'effective')	0.5
('but', 'for', 'now')	0.5
('not', 'have', 'an')	0.5
('work', 'and', 'we')	0.5
('also', 'wanted', 'to')	0.5
('of', 'the', 'reason')	0.5
('be', 'very', 'clear')	0.5
('when', 'youre', 'talking')	0.5
('trying', 'to', 'address')	0.5
('the', 'kind', 'of')	0.5
('are', 'in', 'this')	0.5
('the', 'reasons', 'previously')	0.5
('are', 'committed', 'to')	0.5
('there', 'are', 'a')	0.5
('gpe', 'and', 'cardinal')	0.5
('and', 'were', 'going')	0.5
('would', 'be', 'happy')	0.5
('right', 'now', 'and')	0.5

('even', 'though', 'the')	0.5
('to', 'go', 'back')	0.5
('of', 'our', 'students')	0.5
('it', 'to', 'be')	0.5
('so', 'that', 'the')	0.5
('those', 'reasons', 'were')	0.5
('this', 'in', 'a')	0.5
('under', 'this', 'bill')	0.5
('believe', 'that', 'we')	0.5
('any', 'of', 'those')	0.5
('that', 'if', 'you')	0.5
('on', 'it', 'and')	0.5
('not', 'talking', 'about')	0.5
('had', 'a', 'chance')	0.5
('and', 'they', 'dont')	0.5
('of', 'a', 'problem')	0.5
('gpe', 'that', 'have')	0.5
('most', 'of', 'our')	0.5
('was', 'that', 'the')	0.5
('and', 'we', 'still')	0.5
('and', 'i', 'appreciate')	0.5
('is', 'subject', 'to')	0.5
('that', 'org', 'has')	0.5
('weve', 'been', 'working')	0.5
('of', 'org', 'so')	0.5
('number', 'of', 'reasons')	0.5
('looking', 'at', 'this')	0.5
('language', 'thank', 'you')	0.5
('come', 'back', 'to')	0.5
('safety', 'of', 'our')	0.5
('person', 'for', 'his')	0.5
('we', 'do', 'not')	0.5
('support', 'of', 'the')	0.5
('i', 'hope', 'that')	0.5
('that', 'its', 'not')	0.5
('on', 'the', 'table')	0.5
('as', 'a', 'state')	0.5
('that', 'the', 'state')	0.5
('that', 'is', 'that')	0.5
('that', 'the', 'gpe')	0.5
('us', 'to', 'address')	0.5
('with', 'my', 'colleagues')	0.5
('distributed', 'energy', 'resources')	0.5
('go', 'through', 'the')	0.5
('the', 'ordinal', 'thing')	0.5
('date', 'in', 'date')	0.5
('to', 'the', 'issue')	0.5

('it', 'provides', 'a')	0.5
('that', 'should', 'be')	0.5
('attorney', 'at', 'org')	0.5
('are', 'not', 'a')	0.5
('the', 'bill', 'it')	0.5
('a', 'regular', 'basis')	0.5
('not', 'opposed', 'to')	0.5
('if', 'youre', 'not')	0.5
('with', 'many', 'of')	0.5
('the', 'bill', 'which')	0.5
('component', 'of', 'the')	0.5
('that', 'have', 'already')	0.5
('authority', 'in', 'support')	0.5
('so', 'much', 'for')	0.5
('over', 'the', 'course')	0.5
('org', 'i', 'think')	0.5
('no', 'vote', 'on')	0.5
('ask', 'you', 'to')	0.5
('talking', 'about', 'a')	0.5
('the', 'intent', 'of')	0.5
('and', 'we', 'cant')	0.5
('of', 'law', 'and')	0.5
('author', 'and', 'sponsors')	0.5
('as', 'i', 'said')	0.5
('to', 'suggest', 'that')	0.5
('gpe', 'we', 'also')	0.5
('appreciate', 'the', 'opportunity')	0.5
('issue', 'and', 'we')	0.5
('in', 'opposition', 'thank')	0.5
('that', 'i', 'had')	0.5
('were', 'not', 'talking')	0.5
('of', 'org', 'id')	0.5
('to', 'come', 'back')	0.5
('for', 'some', 'of')	0.5
('ensure', 'that', 'we')	0.5
('its', 'not', 'a')	0.5
('you', 'can', 'do')	0.5
('the', 'hard', 'work')	0.5
('does', 'that', 'mean')	0.5
('so', 'that', 'would')	0.5
('in', 'gpe', 'the')	0.5
('the', 'committee', 'for')	0.5
('see', 'in', 'the')	0.5
('so', 'i', 'will')	0.5
('org', 'org', 'as')	0.5
('you', 'person', 'for')	0.5
('we', 'are', 'now')	0.5

('up', 'to', 'the')	0.5
('say', 'that', 'we')	0.5
('bill', 'we', 'do')	0.5
('to', 'be', 'paid')	0.5
('pointed', 'out', 'in')	0.5
('in', 'respectful', 'opposition')	0.5
('on', 'trying', 'to')	0.5
('time', 'my', 'names')	0.5
('opposition', 'of', 'the')	0.5
('to', 'gpe', 'cardinal')	0.5
('gpe', 'since', 'date')	0.5
('reasons', 'previously', 'stated')	0.5
('that', 'needs', 'to')	0.5
('the', 'bill', 'thank')	0.5
('from', 'the', 'public')	0.5
('theyre', 'supposed', 'to')	0.5
('does', 'not', 'take')	0.5
('through', 'the', 'org')	0.5
('bill', 'would', 'not')	0.5
('by', 'mr', 'person')	0.5
('the', 'law', 'enforcement')	0.5
('be', 'the', 'ordinal')	0.5
('dont', 'know', 'that')	0.5
('yes', 'thank', 'you')	0.5
('im', 'a', 'staff')	0.5
('id', 'be', 'happy')	0.5
('bill', 'we', 'are')	0.5
('author', 'and', 'staff')	0.5
('in', 'that', 'particular')	0.5
('the', 'safety', 'of')	0.5
('if', 'we', 'do')	0.5
('and', 'then', 'to')	0.5
('things', 'that', 'were')	0.5
('id', 'also', 'like')	0.5
('of', 'things', 'and')	0.5
('of', 'the', 'employee')	0.5
('and', 'that', 'we')	0.5
('thats', 'something', 'that')	0.5
('date', 'we', 'have')	0.5
('cardinal', 'other', 'states')	0.5
('analysis', 'that', 'was')	0.5
('provisions', 'of', 'this')	0.5
('by', 'cardinal', 'of')	0.5
('or', 'the', 'other')	0.5
('the', 'recent', 'amendments')	0.5
('org', 'id', 'like')	0.5
('could', 'lead', 'to')	0.5

('that', 'people', 'can')	0.5
('we', 'have', 'an')	0.5
('need', 'for', 'the')	0.5
('sponsor', 'this', 'bill')	0.5
('this', 'is', 'all')	0.5
('want', 'to', 'go')	0.5
('the', 'funding', 'for')	0.5
('need', 'for', 'a')	0.5
('thing', 'to', 'do')	0.5
('about', 'how', 'this')	0.5
('with', 'org', 'thank')	0.5
('so', 'were', 'not')	0.5
('us', 'and', 'we')	0.5
('continue', 'working', 'with')	0.5
('of', 'the', 'health')	0.5
('the', 'org', 'in')	0.5
('and', 'have', 'to')	0.5
('believe', 'that', 'that')	0.5
('the', 'org', 'that')	0.5
('so', 'that', 'people')	0.5
('get', 'to', 'the')	0.5
('person', 'with', 'org')	0.5
('were', 'in', 'a')	0.5
('would', 'love', 'to')	0.5
('madam', 'person', 'members')	0.5
('over', 'date', 'to')	0.5
('of', 'this', 'and')	0.5
('worked', 'with', 'the')	0.5
('trying', 'to', 'make')	0.5
('as', 'a', 'condition')	0.5
('we', 'do', 'support')	0.5
('were', 'proud', 'to')	0.5
('on', 'the', 'road')	0.5
('for', 'his', 'leadership')	0.5
('the', 'author', 'we')	0.5
('cities', 'we', 'are')	0.5
('come', 'back', 'and')	0.5
('of', 'the', 'victims')	0.5
('the', 'need', 'for')	0.5
('want', 'to', 'get')	0.5
('been', 'on', 'the')	0.5
('they', 'are', 'a')	0.5
('author', 'and', 'his')	0.5
('in', 'opposition', 'of')	0.5
('we', 'havent', 'seen')	0.5
('that', 'information', 'is')	0.5
('on', 'gpe', 'date')	0.5

('understand', 'the', 'need')	0.5
('this', 'time', 'we')	0.5
('understand', 'that', 'the')	0.5
('bill', 'we', 'appreciate')	0.5
('harassment', 'or', 'discrimination')	0.5
('we', 'think', 'it')	0.5
('this', 'bill', 'date')	0.5
('this', 'legislation', 'thank')	0.5
('for', 'a', 'no')	0.5
('mr', 'person', 'the')	0.5
('of', 'them', 'is')	0.5
('what', 'may', 'be')	0.5
('this', 'bill', 'passes')	0.5
('raised', 'by', 'the')	0.5
('and', 'we', 'need')	0.5
('does', 'not', 'affect')	0.5
('if', 'they', 'do')	0.5
('this', 'bill', 'so')	0.5
('to', 'what', 'we')	0.5
('take', 'very', 'seriously')	0.5
('person', 'org', 'we')	0.5
('you', 'would', 'have')	0.5
('all', 'the', 'way')	0.5
('say', 'that', 'were')	0.5
('this', 'bill', 'does')	0.5
('the', 'people', 'in')	0.5
('not', 'in', 'opposition')	0.5
('are', 'proud', 'to')	0.5
('org', 'were', 'proud')	0.5
('org', 'representing', 'org')	0.5
('that', 'in', 'gpe')	0.5
('a', 'process', 'in')	0.5
('gpe', 'is', 'the')	0.5
('org', 'for', 'date')	0.5
('by', 'my', 'colleague')	0.5
('has', 'to', 'be')	0.5
('is', 'an', 'issue')	0.5
('i', 'echo', 'the')	0.5

Successful 3 word phrases for testimonies with "Against" or "Indeterminate" alignment

3 word phrase (trigram)	Success Rate
('to', 'read', 'thank')	1
('move', 'america', 'and')	1
('oppose', 'along', 'with')	1
('gpe', 'hi', 'im')	1
('la', 'org', 'org')	1
('org', 'esperanza', 'community')	1
('corporation', 'person', 'org')	1
('health', 'councils', 'org')	1
('take', 'up', 'more')	1
('esperanza', 'community', 'housing')	1
('advocates', 'here', 'to')	1
('you', 'org', 'esperanza')	1
('i', 'wont', 'bother')	1
('in', 'place', 'jobs')	1
('justice', 'org', 'coalition')	1
('up', 'more', 'time')	1
('person', 'org', 'investing')	1
('org', 'fair', 'rents')	1
('wont', 'bother', 'to')	1
('america', 'and', 'cardinal')	1
('to', 'move', 'america')	1
('counsel', 'act', 'la')	1
('for', 'justice', 'org')	1
('public', 'counsel', 'act')	1
('fair', 'rents', 'for')	1
('org', 'org', 'beyond')	1
('org', 'investing', 'in')	1
('community', 'technologies', 'community')	1
('community', 'health', 'councils')	1
('to', 'oppose', 'along')	1
('community', 'housing', 'corporation')	1
('read', 'thank', 'you')	1
('act', 'la', 'org')	1
('economic', 'survival', 'community')	1
('beyond', 'the', 'arc')	1
('councils', 'org', 'fair')	1
('for', 'economic', 'survival')	1
('place', 'jobs', 'to')	1
('technologies', 'community', 'health')	1
('org', 'beyond', 'the')	1
('rents', 'for', 'gpe')	1
('bother', 'to', 'take')	1
('housing', 'corporation', 'person')	1
('coalition', 'for', 'economic')	1
('jobs', 'to', 'move')	1

('survival', 'community', 'technologies')	1
('gpe', 'public', 'counsel')	1
('investing', 'in', 'place')	1
('for', 'gpe', 'public')	1
('that', 'i', 'wont')	0.958333333
('org', 'coalition', 'for')	0.956521739
('other', 'organizations', 'that')	0.956521739
('to', 'take', 'up')	0.956521739
('cardinal', 'other', 'organizations')	0.956521739
('time', 'to', 'read')	0.956521739
('of', 'gpe', 'hi')	0.954545455
('organizations', 'that', 'i')	0.88
('and', 'person', 'on')	0.833333333
('here', 'to', 'oppose')	0.806451613
('down', 'the', 'state')	0.8
('high', 'speed', 'rail')	0.8
('you', 'can', 'get')	0.8
('up', 'and', 'down')	0.8
('and', 'down', 'the')	0.8
('for', 'org', 'for')	0.75
('if', 'this', 'is')	0.75
('what', 'do', 'we')	0.75
('we', 'would', 'argue')	0.75
('of', 'the', 'cities')	0.714285714
('of', 'domestic', 'violence')	0.666666667
('representing', 'org', 'also')	0.666666667
('be', 'paid', 'for')	0.666666667
('still', 'have', 'some')	0.666666667
('as', 'a', 'result')	0.666666667
('gpe', 'right', 'now')	0.666666667
('nothing', 'in', 'the')	0.666666667
('opportunity', 'to', 'work')	0.666666667
('date', 'i', 'have')	0.666666667
('is', 'a', 'small')	0.666666667
('thats', 'not', 'going')	0.666666667
('are', 'supportive', 'of')	0.666666667
('to', 'stand', 'up')	0.666666667
('to', 'invest', 'in')	0.666666667
('in', 'opposition', 'date')	0.666666667
('org', 'from', 'gpe')	0.666666667
('not', 'allowed', 'to')	0.666666667
('to', 'acknowledge', 'that')	0.666666667
('opposition', 'to', 'this')	0.666666667
('who', 'work', 'in')	0.666666667
('senators', 'person', 'on')	0.666666667
('believe', 'it', 'is')	0.666666667
('we', 'were', 'the')	0.666666667

('and', 'also', 'a')	0.666666667
('problem', 'and', 'we')	0.666666667
('that', 'reason', 'we')	0.666666667
('for', 'affordable', 'housing')	0.666666667
('money', 'an', 'hour')	0.666666667
('for', 'the', 'purpose')	0.666666667
('the', 'cities', 'of')	0.625
('person', 'with', 'all')	0.6
('this', 'is', 'something')	0.6
('and', 'org', 'also')	0.571428571
('to', 'address', 'this')	0.5
('be', 'on', 'the')	0.5
('in', 'your', 'analysis')	0.5
('for', 'in', 'the')	0.5
('sponsor', 'of', 'org')	0.5
('mr', 'persons', 'comments')	0.5
('look', 'at', 'how')	0.5
('senator', 'person', 'on')	0.5
('are', 'on', 'the')	0.5
('a', 'date', 'basis')	0.5
('to', 'work', 'for')	0.5
('his', 'or', 'her')	0.5
('in', 'a', 'different')	0.5
('of', 'affordable', 'housing')	0.5
('kinds', 'of', 'things')	0.5
('that', 'the', 'court')	0.5
('but', 'it', 'does')	0.5
('to', 'see', 'it')	0.5
('date', 'in', 'the')	0.5
('the', 'hook', 'for')	0.5
('the', 'bill', 'with')	0.5
('that', 'they', 'will')	0.5
('something', 'that', 'they')	0.5
('gpe', 'would', 'be')	0.5
('this', 'is', 'very')	0.5
('that', 'is', 'the')	0.5
('as', 'well', 'we')	0.5
('cardinal', 'of', 'them')	0.5
('will', 'create', 'a')	0.5
('that', 'theyre', 'going')	0.5
('a', 'problem', 'and')	0.5
('because', 'we', 'do')	0.5
('we', 'are', 'also')	0.5
('to', 'kind', 'of')	0.5
('than', 'happy', 'to')	0.5
('assure', 'you', 'that')	0.5
('who', 'wants', 'to')	0.5

('org', 'county', 'and')	0.5
('this', 'bill', 'its')	0.5
('with', 'org', 'county')	0.5
('to', 'this', 'issue')	0.5
('there', 'is', 'not')	0.5
('have', 'in', 'the')	0.5
('bill', 'does', 'and')	0.5
('again', 'we', 'are')	0.5
('of', 'the', 'california')	0.5
('but', 'we', 'would')	0.5
('is', 'responsible', 'for')	0.5
('mr', 'person', 'senators')	0.5
('when', 'it', 'was')	0.5
('time', 'person', 'with')	0.5
('that', 'is', 'to')	0.5
('that', 'you', 'know')	0.5
('were', 'one', 'of')	0.5
('purpose', 'of', 'the')	0.5
('org', 'to', 'make')	0.5
('thats', 'going', 'to')	0.5
('is', 'a', 'statewide')	0.5
('and', 'in', 'addition')	0.5
('what', 'has', 'been')	0.5
('a', 'result', 'of')	0.5
('went', 'on', 'to')	0.5
('for', 'date', 'of')	0.5
('the', 'cosponsors', 'of')	0.5
('on', 'a', 'date')	0.5
('going', 'to', 'put')	0.5
('that', 'were', 'concerned')	0.5
('for', 'people', 'who')	0.5
('impact', 'of', 'the')	0.5
('of', 'them', 'are')	0.5
('to', 'come', 'in')	0.5
('behalf', 'of', 'norp')	0.5
('is', 'a', 'date')	0.5
('an', 'attempt', 'to')	0.5
('mr', 'person', 'im')	0.5
('someone', 'who', 'has')	0.5
('this', 'reason', 'we')	0.5
('the', 'requirements', 'of')	0.5
('reach', 'out', 'to')	0.5
('of', 'a', 'bill')	0.5
('to', 'live', 'in')	0.5
('in', 'date', 'which')	0.5
('for', 'the', 'state')	0.5
('point', 'of', 'view')	0.5

('it', 'creates', 'a')	0.5
('that', 'we', 'need')	0.5
('date', 'or', 'cardinal')	0.5
('we', 'are', 'supportive')	0.5
('thats', 'exactly', 'what')	0.5
('date', 'is', 'not')	0.5
('from', 'org', 'in')	0.5
('bill', 'seeks', 'to')	0.5
('compared', 'to', 'the')	0.5
('this', 'bill', 'seeks')	0.5
('in', 'charge', 'of')	0.5
('so', 'this', 'is')	0.5
('gpe', 'and', 'also')	0.5
('allow', 'them', 'to')	0.5
('are', 'also', 'in')	0.5
('want', 'to', 'ensure')	0.5
('for', 'percent', 'of')	0.5
('to', 'communicate', 'with')	0.5
('municipal', 'employees', 'in')	0.5
('to', 'us', 'and')	0.5
('people', 'are', 'going')	0.5
('this', 'is', 'really')	0.5
('do', 'not', 'want')	0.5
('we', 'really', 'do')	0.5
('allow', 'for', 'the')	0.5
('for', 'this', 'reason')	0.5
('of', 'the', 'cosponsors')	0.5
('the', 'law', 'to')	0.5
('of', 'the', 'biggest')	0.5
('money', 'to', 'pay')	0.5
('going', 'to', 'pay')	0.5
('to', 'get', 'that')	0.5
('reason', 'for', 'that')	0.5
('members', 'cesar', 'diaz')	0.5
('youre', 'talking', 'about')	0.5
('to', 'understand', 'that')	0.5
('in', 'gpe', 'i')	0.5
('bill', 'does', 'it')	0.5
('for', 'a', 'couple')	0.5
('it', 'easier', 'for')	0.5
('to', 'close', 'the')	0.5
('i', 'understand', 'that')	0.5
('we', 'try', 'to')	0.5
('the', 'problems', 'that')	0.5
('we', 'do', 'think')	0.5
('in', 'date', 'so')	0.5

Successful 4 word phrases for testimonies with "For" or "Indeterminate" alignment:

4 word phrase (quadgram)	Success Rate
('org', 'in', 'opposition', 'to')	1
('were', 'opposed', 'to', 'the')	1
('but', 'at', 'this', 'time')	1
('we', 'share', 'the', 'concerns')	1
('org', 'we', 'oppose', 'this')	1
('dont', 'have', 'an', 'official')	1
('you', 'to', 'oppose', 'this')	1
('to', 'oppose', 'this', 'bill')	1
('org', 'in', 'respectful', 'opposition')	1
('opposition', 'to', 'this', 'measure')	1
('you', 'to', 'vote', 'no')	1
('in', 'opposition', 'person', 'on')	1
('opposition', 'person', 'on', 'behalf')	1
('with', 'org', 'also', 'opposed')	1
('were', 'going', 'to', 'continue')	1
('those', 'reasons', 'were', 'opposed')	1
('we', 'strongly', 'oppose', 'this')	1
('the', 'proponents', 'of', 'the')	1
('i', 'would', 'note', 'that')	1
('we', 'do', 'have', 'concerns')	1
('no', 'vote', 'on', 'this')	1
('the', 'amendments', 'that', 'the')	1
('in', 'its', 'current', 'form')	1
('concerns', 'that', 'we', 'have')	1
('sexual', 'orientation', 'change', 'efforts')	1
('we', 'are', 'strongly', 'opposed')	1
('are', 'strongly', 'opposed', 'to')	1
('strongly', 'opposed', 'to', 'this')	1
('to', 'this', 'bill', 'thank')	1
('in', 'its', 'present', 'form')	1
('oppose', 'this', 'measure', 'thank')	1
('in', 'opposition', 'to', 'org')	1
('your', 'no', 'vote', 'on')	1
('to', 'the', 'bill', 'as')	1
('org', 'in', 'opposition', 'person')	1
('those', 'reasons', 'we', 'remain')	1
('a', 'lot', 'of', 'concerns')	1
('in', 'opposition', 'to', 'gpe')	1
('opposed', 'to', 'the', 'measure')	1
('the', 'bill', 'is', 'currently')	1
('these', 'reasons', 'we', 'oppose')	1
('our', 'opposition', 'thank', 'you')	1
('that', 'this', 'is', 'going')	1
('and', 'technology', 'association', 'also')	1
('technology', 'association', 'also', 'opposed')	1

('here', 'date', 'in', 'opposition')	1
('the', 'bill', 'in', 'its')	1
('bill', 'in', 'its', 'current')	1
('person', 'i', 'represent', 'cardinal')	1
('i', 'represent', 'cardinal', 'regional')	1
('represent', 'cardinal', 'regional', 'apartment')	1
('cardinal', 'regional', 'apartment', 'associations')	1
('not', 'have', 'a', 'position')	0.941176471
('we', 'dont', 'have', 'an')	0.933333333
('org', 'we', 'dont', 'have')	0.916666667
('have', 'a', 'position', 'on')	0.909090909
('a', 'position', 'on', 'the')	0.909090909
('neutral', 'on', 'the', 'bill')	0.909090909
('some', 'concerns', 'with', 'the')	0.909090909
('have', 'some', 'concerns', 'with')	0.9
('do', 'have', 'some', 'concerns')	0.888888889
('look', 'forward', 'to', 'continuing')	0.875
('position', 'on', 'the', 'bill')	0.875
('dont', 'have', 'a', 'position')	0.875
('on', 'the', 'bill', 'but')	0.857142857
('like', 'to', 'work', 'with')	0.857142857
('to', 'continuing', 'to', 'work')	0.833333333
('of', 'org', 'of', 'the')	0.833333333
('and', 'the', 'org', 'staff')	0.833333333
('want', 'to', 'speak', 'to')	0.833333333
('on', 'the', 'bill', 'we')	0.8
('an', 'official', 'position', 'on')	0.8
('need', 'to', 'be', 'made')	0.8
('look', 'forward', 'to', 'continue')	0.8
('author', 'and', 'the', 'sponsors')	0.8
('im', 'going', 'to', 'be')	0.8
('all', 'the', 'work', 'that')	0.8
('do', 'not', 'have', 'a')	0.75
('so', 'im', 'going', 'to')	0.75
('person', 'person', 'on', 'behalf')	0.75
('gpe', 'manufacturers', 'and', 'technology')	0.75
('manufacturers', 'and', 'technology', 'association')	0.75
('continuing', 'to', 'work', 'with')	0.75
('org', 'we', 'do', 'not')	0.75
('have', 'an', 'official', position')	0.75
('with', 'the', 'author', 'and')	0.75
('and', 'org', 'in', 'opposition')	0.75
('the', 'bill', 'as', 'it')	0.75
('of', 'org', 'we', 'have')	0.75
('so', 'much', 'for', 'your')	0.75
('because', 'they', 'dont', 'have')	0.75
('for', 'the', 'state', 'and')	0.75

('there', 'are', 'a', 'lot')	0.75
('are', 'a', 'lot', 'of')	0.75
('org', 'and', 'org', 'have')	0.75
('and', 'i', 'dont', 'know')	0.75
('to', 'continue', 'working', 'with')	0.75
('the', 'authors', 'staff', 'and')	0.75
('trying', 'to', 'find', 'a')	0.75
('we', 'are', 'talking', 'about')	0.75
('youre', 'not', 'going', 'to')	0.75
('the', 'authors', 'office', 'to')	0.75
('we', 'are', 'trying', 'to')	0.666666667
('a', 'number', 'of', 'issues')	0.666666667
('with', 'the', 'author', 'to')	0.666666667
('the', 'gpe', 'manufacturers', 'and')	0.666666667
('forward', 'to', 'continuing', 'to')	0.666666667
('we', 'do', 'not', 'have')	0.666666667
('in', 'the', 'bill', 'we')	0.666666667
('like', 'to', 'thank', 'the')	0.666666667
('we', 'do', 'have', 'some')	0.666666667
('cap', 'and', 'trade', 'program')	0.666666667
('the', 'bill', 'we', 'think')	0.666666667
('greenhouse', 'gas', 'emissions', 'and')	0.666666667
('opposed', 'to', 'this', 'bill')	0.666666667
('at', 'this', 'time', 'we')	0.666666667
('that', 'there', 'is', 'a')	0.666666667
('to', 'the', 'authors', 'office')	0.666666667
('have', 'an', 'issue', 'with')	0.666666667
('i', 'will', 'be', 'brief')	0.666666667
('at', 'the', 'very', 'least')	0.666666667
('this', 'bill', 'does', 'not')	0.666666667
('were', 'concerned', 'about', 'the')	0.666666667
('the', 'reality', 'is', 'that')	0.666666667
('the', 'majority', 'of', 'the')	0.666666667
('on', 'the', 'bill', 'and')	0.666666667
('of', 'org', 'we', 'too')	0.666666667
('some', 'of', 'the', 'concerns')	0.666666667
('mr', 'person', 'person', 'on')	0.666666667
('have', 'a', 'formal', 'position')	0.666666667
('to', 'work', 'with', 'them')	0.666666667
('that', 'we', 'will', 'be')	0.666666667
('the', 'provisions', 'of', 'this')	0.666666667
('whether', 'or', 'not', 'the')	0.666666667
('local', 'law', 'enforcement', 'agencies')	0.666666667
('and', 'i', 'appreciate', 'the')	0.666666667
('believe', 'that', 'this', 'bill')	0.666666667
('and', 'we', 'appreciate', 'the')	0.666666667
('position', 'on', 'this', 'bill')	0.666666667

('they', 'do', 'not', 'have')	0.666666667
('on', 'some', 'of', 'the')	0.666666667
('at', 'this', 'time', 'but')	0.666666667
('the', 'things', 'that', 'we')	0.666666667
('be', 'able', 'to', 'support')	0.666666667
('org', 'we', 'have', 'a')	0.666666667
('and', 'thats', 'why', 'the')	0.666666667
('you', 'dont', 'have', 'to')	0.666666667
('this', 'bill', 'will', 'be')	0.666666667
('we', 'want', 'to', 'work')	0.666666667
('want', 'to', 'work', 'with')	0.666666667
('with', 'org', 'we', 'dont')	0.666666667
('need', 'for', 'the', 'bill')	0.666666667
('we', 'continue', 'to', 'work')	0.666666667
('to', 'tell', 'you', 'that')	0.666666667
('are', 'opposed', 'to', 'this')	0.666666667
('and', 'i', 'just', 'wanted')	0.666666667
('this', 'bill', 'this', 'is')	0.666666667
('feel', 'that', 'this', 'bill')	0.666666667
('i', 'did', 'want', 'to')	0.666666667
('the', 'chief', 'probation', 'officers')	0.666666667
('our', 'opposition', 'to', 'this')	0.666666667
('going', 'to', 'go', 'to')	0.666666667
('from', 'gpe', 'to', 'gpe')	0.666666667
('we', 'have', 'to', 'be')	0.666666667
('thank', 'the', 'author', 'the')	0.666666667
('behalf', 'of', 'the', 'chief')	0.666666667
('of', 'the', 'chief', 'probation')	0.666666667
('chief', 'probation', 'officers', 'of')	0.666666667
('probation', 'officers', 'of', 'gpe')	0.666666667
('we', 'appreciate', 'the', 'opportunity')	0.666666667
('in', 'the', 'committee', 'analysis')	0.666666667
('like', 'to', 'continue', 'to')	0.666666667
('for', 'all', 'the', 'work')	0.666666667
('provisions', 'of', 'the', 'bill')	0.6
('so', 'wed', 'like', 'to')	0.6
('thank', 'you', 'mr', 'person')	0.5
('you', 'mr', 'person', 'and')	0.5
('mr', 'person', 'and', 'members')	0.5
('in', 'opposition', 'thank', 'you')	0.5
('person', 'with', 'org', 'in')	0.5
('you', 'to', 'vote', 'yes')	0.5
('to', 'vote', 'yes', 'on')	0.5
('person', 'for', 'org', 'in')	0.5
('the', 'language', 'in', 'the')	0.5
('with', 'org', 'in', 'support')	0.5
('work', 'with', 'org', 'and')	0.5

('i', 'dont', 'know', 'that')	0.5
('we', 'need', 'to', 'take')	0.5
('in', 'support', 'of', 'the')	0.5
('support', 'of', 'the', 'bill')	0.5
('and', 'id', 'like', 'to')	0.5
('of', 'org', 'i', 'just')	0.5
('opposed', 'to', 'the', 'bill')	0.5
('that', 'needs', 'to', 'be')	0.5
('also', 'in', 'support', 'of')	0.5
('issues', 'with', 'the', 'bill')	0.5
('with', 'some', 'of', 'the')	0.5
('of', 'org', 'in', 'opposition')	0.5
('of', 'org', 'we', 'also')	0.5
('if', 'youre', 'going', 'to')	0.5
('behalf', 'of', 'org', 'just')	0.5
('over', 'the', 'course', 'of')	0.5
('want', 'to', 'continue', 'to')	0.5
('of', 'all', 'of', 'the')	0.5
('in', 'opposition', 'of', 'the')	0.5
('time', 'my', 'names', 'person')	0.5
('ask', 'you', 'to', 'vote')	0.5
('this', 'bill', 'we', 'have')	0.5
('we', 'have', 'concerns', 'with')	0.5
('have', 'concerns', 'with', 'the')	0.5
('we', 'really', 'appreciate', 'the')	0.5
('person', 'here', 'on', 'behalf')	0.5
('names', 'person', 'im', 'the')	0.5
('executive', 'director', 'for', 'org')	0.5
('person', 'for', 'his', 'leadership')	0.5
('we', 'are', 'proud', 'to')	0.5
('org', 'in', 'support', 'of')	0.5
('for', 'a', 'lot', 'of')	0.5
('do', 'not', 'have', 'an')	0.5
('names', 'person', 'im', 'with')	0.5
('thank', 'the', 'author', 'and')	0.5
('the', 'author', 'and', 'his')	0.5
('time', 'person', 'of', 'org')	0.5
('and', 'we', 'do', 'have')	0.5
('id', 'be', 'happy', 'to')	0.5
('to', 'take', 'care', 'of')	0.5
('the', 'cap', 'and', 'trade')	0.5
('and', 'then', 'the', 'other')	0.5
('intent', 'of', 'the', 'bill')	0.5
('to', 'ensure', 'that', 'we')	0.5
('and', 'so', 'this', 'bill')	0.5
('in', 'this', 'area', 'and')	0.5
('and', 'i', 'have', 'to')	0.5

('when', 'youre', 'talking', 'about')	0.5
('with', 'org', 'and', 'we')	0.5
('org', 'id', 'like', 'to')	0.5
('org', 'member', 'person', 'for')	0.5
('org', 'in', 'opposition', 'thank')	0.5
('would', 'like', 'to', 'have')	0.5
('the', 'author', 'and', 'sponsors')	0.5
('for', 'his', 'leadership', 'on')	0.5
('his', 'leadership', 'on', 'this')	0.5
('leadership', 'on', 'this', 'issue')	0.5
('this', 'issue', 'and', 'we')	0.5
('to', 'the', 'author', 'and')	0.5
('the', 'bill', 'thank', 'you')	0.5
('with', 'org', 'thank', 'you')	0.5
('you', 'so', 'much', 'for')	0.5
('im', 'a', 'staff', 'attorney')	0.5
('behalf', 'of', 'org', 'ill')	0.5
('want', 'to', 'thank', 'org')	0.5
('the', 'need', 'for', 'the')	0.5
('vote', 'on', 'this', 'bill')	0.5
('to', 'this', 'bill', 'we')	0.5
('as', 'the', 'org', 'member')	0.5
('in', 'gpe', 'and', 'also')	0.5
('needs', 'to', 'be', 'a')	0.5
('the', 'bill', 'that', 'we')	0.5
('theyre', 'not', 'going', 'to')	0.5
('of', 'the', 'bill', 'thank')	0.5
('is', 'a', 'lot', 'of')	0.5
('in', 'the', 'bill', 'but')	0.5
('and', 'we', 'understand', 'that')	0.5
('we', 'understand', 'that', 'the')	0.5
('so', 'we', 'think', 'that')	0.5
('in', 'the', 'hands', 'of')	0.5
('for', 'those', 'reasons', 'were')	0.5
('work', 'with', 'the', 'author')	0.5
('ask', 'for', 'a', 'no')	0.5
('for', 'a', 'no', 'vote')	0.5
('been', 'working', 'on', 'this')	0.5
('working', 'on', 'this', 'issue')	0.5
('org', 'we', 'too', 'are')	0.5
('for', 'the', 'reasons', 'previously')	0.5
('to', 'come', 'up', 'with')	0.5
('org', 'i', 'just', 'want')	0.5
('we', 'do', 'support', 'the')	0.5
('on', 'a', 'regular', 'basis')	0.5
('need', 'to', 'be', 'addressed')	0.5
('that', 'need', 'to', 'be')	0.5

('to', 'work', 'with', 'us')	0.5
('myself', 'with', 'the', 'comments')	0.5
('this', 'bill', 'and', 'i')	0.5
('we', 'do', 'want', 'to')	0.5
('willing', 'to', 'work', 'with')	0.5
('provisions', 'of', 'this', 'bill')	0.5
('in', 'the', 'bill', 'and')	0.5
('needs', 'to', 'be', 'addressed')	0.5
('working', 'with', 'us', 'and')	0.5
('weve', 'been', 'working', 'with')	0.5
('been', 'working', 'with', 'the')	0.5
('we', 'dont', 'have', 'to')	0.5
('this', 'bill', 'is', 'an')	0.5
('staff', 'attorney', 'at', 'org')	0.5
('had', 'a', 'chance', 'to')	0.5
('work', 'with', 'the', 'authors')	0.5
('and', 'as', 'a', 'result')	0.5
('that', 'were', 'trying', 'to')	0.5
('that', 'are', 'in', 'place')	0.5
('on', 'this', 'bill', 'but')	0.5
('behalf', 'of', 'org', 'id')	0.5
('of', 'org', 'id', 'like')	0.5
('much', 'for', 'the', 'opportunity')	0.5
('a', 'lot', 'of', 'the')	0.5
('nothing', 'in', 'this', 'bill')	0.5
('in', 'this', 'bill', 'that')	0.5
('and', 'i', 'dont', 'think')	0.5
('to', 'the', 'bill', 'thank')	0.5
('that', 'we', 'want', 'to')	0.5
('we', 'think', 'that', 'there')	0.5
('well', 'be', 'able', 'to')	0.5
('version', 'of', 'the', 'bill')	0.5
('my', 'colleague', 'from', 'org')	0.5
('you', 'mr', 'person', 'my')	0.5
('is', 'an', 'issue', 'that')	0.5
('and', 'we', 'need', 'to')	0.5
('for', 'the', 'same', 'reasons')	0.5
('the', 'bill', 'has', 'been')	0.5
('and', 'that', 'would', 'be')	0.5
('have', 'to', 'go', 'through')	0.5
('just', 'want', 'to', 'say')	0.5
('if', 'you', 'want', 'to')	0.5
('members', 'person', 'with', 'the')	0.5
('of', 'the', 'board', 'of')	0.5
('this', 'bill', 'we', 'are')	0.5
('to', 'address', 'our', 'concerns')	0.5
('org', 'of', 'gpe', 'also')	0.5

('it', 'needs', 'to', 'be')	0.5
('work', 'on', 'this', 'bill')	0.5
('that', 'it', 'should', 'be')	0.5
('here', 'in', 'opposition', 'to')	0.5
('we', 'greatly', 'appreciate', 'the')	0.5
('with', 'us', 'to', 'address')	0.5
('to', 'take', 'advantage', 'of')	0.5
('time', 'thank', 'you', 'for')	0.5
('the', 'org', 'staff', 'for')	0.5
('org', 'in', 'gpe', 'we')	0.5
('we', 'think', 'the', 'bill')	0.5
('to', 'sponsor', 'this', 'bill')	0.5
('be', 'able', 'to', 'do')	0.5
('thank', 'you', 'person', 'for')	0.5
('to', 'look', 'at', 'the')	0.5
('and', 'we', 'hope', 'to')	0.5
('good', 'time', 'org', 'members')	0.5
('a', 'number', 'of', 'reasons')	0.5
('with', 'the', 'authors', 'staff')	0.5
('be', 'part', 'of', 'the')	0.5
('of', 'the', 'bill', 'but')	0.5
('we', 'need', 'to', 'get')	0.5
('i', 'want', 'to', 'start')	0.5
('want', 'to', 'start', 'by')	0.5
('i', 'think', 'there', 'are')	0.5
('the', 'sponsor', 'and', 'the')	0.5
('some', 'of', 'the', 'amendments')	0.5
('county', 'and', 'municipal', 'employees')	0.5
('have', 'the', 'ability', 'to')	0.5
('id', 'also', 'like', 'to')	0.5
('and', 'were', 'going', 'to')	0.5
('language', 'of', 'the', 'bill')	0.5
('were', 'not', 'talking', 'about')	0.5
('person', 'org', 'in', 'opposition')	0.5
('but', 'i', 'think', 'that')	0.5
('this', 'bill', 'this', 'bill')	0.5
('the', 'state', 'so', 'we')	0.5
('if', 'this', 'bill', 'passes')	0.5
('also', 'would', 'like', 'to')	0.5
('of', 'org', 'i', 'want')	0.5
('about', 'the', 'fact', 'that')	0.5
('as', 'a', 'condition', 'of')	0.5
('just', 'want', 'to', 'thank')	0.5
('the', 'reasons', 'previously', 'stated')	0.5
('the', 'author', 'and', 'staff')	0.5
('org', 'we', 'appreciate', 'the')	0.5
('take', 'advantage', 'of', 'the')	0.5

('would', 'not', 'apply', 'to')	0.5
('this', 'legislation', 'thank', 'you')	0.5
('noted', 'in', 'the', 'analysis')	0.5
('that', 'the', 'author', 'and')	0.5
('with', 'the', 'fact', 'that')	0.5
('not', 'going', 'to', 'go')	0.5
('with', 'the', 'bill', 'and')	0.5
('for', 'the', 'org', 'of')	0.5
('bill', 'this', 'bill', 'is')	0.5
('are', 'the', 'ones', 'that')	0.5
('to', 'go', 'to', 'a')	0.5
('not', 'going', 'to', 'have')	0.5
('opposition', 'to', 'gpe', 'date')	0.5
('may', 'not', 'be', 'the')	0.5
('that', 'we', 'have', 'with')	0.5
('we', 'have', 'with', 'the')	0.5
('an', 'increase', 'in', 'the')	0.5
('org', 'just', 'want', 'to')	0.5
('committee', 'person', 'on', 'behalf')	0.5
('were', 'proud', 'to', 'support')	0.5
('the', 'way', 'that', 'the')	0.5
('in', 'the', 'commission', 'of')	0.5
('need', 'to', 'take', 'a')	0.5
('we', 'are', 'looking', 'at')	0.5
('about', 'some', 'of', 'the')	0.5
('what', 'i', 'want', 'to')	0.5
('the', 'bill', 'we', 'do')	0.5
('many', 'of', 'which', 'are')	0.5
('say', 'a', 'few', 'words')	0.5
('bill', 'moves', 'forward', 'thank')	0.5
('moves', 'forward', 'thank', 'you')	0.5
('behalf', 'of', 'gpes', 'org')	0.5
('i', 'agree', 'with', 'the')	0.5
('the', 'bill', 'but', 'we')	0.5
('counties', 'of', 'gpe', 'also')	0.5
('of', 'the', 'gpe', 'org')	0.5
('with', 'the', 'sponsors', 'and')	0.5
('of', 'the', 'urban', 'counties')	0.5
('behalf', 'of', 'the', 'urban')	0.5

Successful 4 word phrases for testimonies with "Against" or "Indeterminate" alignment:

4 word phrase (quadgram)	Success Rate
('with', 'public', 'advocates', 'here')	1
('public', 'advocates', 'here', 'to')	1
('advocates', 'here', 'to', 'oppose')	1
('here', 'to', 'oppose', 'along')	1
('to', 'oppose', 'along', 'with')	1
('oppose', 'along', 'with', 'org')	1
('with', 'org', 'org', 'of')	1
('of', 'gpe', 'hi', 'im')	1
('gpe', 'hi', 'im', 'person')	1
('for', 'gpe', 'public', 'counsel')	1
('gpe', 'public', 'counsel', 'act')	1
('public', 'counsel', 'act', 'la')	1
('counsel', 'act', 'la', 'org')	1
('act', 'la', 'org', 'org')	1
('la', 'org', 'org', 'beyond')	1
('org', 'org', 'beyond', 'the')	1
('org', 'beyond', 'the', 'arc')	1
('beyond', 'the', 'arc', 'a')	1
('partnership', 'for', 'justice', 'org')	1
('for', 'justice', 'org', 'coalition')	1
('justice', 'org', 'coalition', 'for')	1
('org', 'coalition', 'for', 'economic')	1
('coalition', 'for', 'economic', 'survival')	1
('for', 'economic', 'survival', 'community')	1
('economic', 'survival', 'community', 'technologies')	1
('survival', 'community', 'technologies', 'community')	1
('community', 'technologies', 'community', 'health')	1
('technologies', 'community', 'health', 'councils')	1
('community', 'health', 'councils', 'org')	1
('health', 'councils', 'org', 'fair')	1
('councils', 'org', 'fair', 'rents')	1
('org', 'fair', 'rents', 'for')	1
('fair', 'rents', 'for', 'gpe')	1
('rents', 'for', 'gpe', 'public')	1
('org', 'esperanza', 'community', 'housing')	1
('esperanza', 'community', 'housing', 'corporation')	1
('community', 'housing', 'corporation', 'person')	1
('housing', 'corporation', 'person', 'org')	1
('corporation', 'person', 'org', 'investing')	1
('person', 'org', 'investing', 'in')	1
('org', 'investing', 'in', 'place')	1
('investing', 'in', 'place', 'jobs')	1
('in', 'place', 'jobs', 'to')	1
('place', 'jobs', 'to', 'move')	1
('jobs', 'to', 'move', 'america')	1

('to', 'move', 'america', 'and')	1
('move', 'america', 'and', 'cardinal')	1
('america', 'and', 'cardinal', 'other')	1
('and', 'cardinal', 'other', 'organizations')	1
('other', 'organizations', 'that', 'i')	1
('organizations', 'that', 'i', 'wont')	1
('that', 'i', 'wont', 'bother')	1
('i', 'wont', 'bother', 'to')	1
('wont', 'bother', 'to', 'take')	1
('bother', 'to', 'take', 'up')	1
('to', 'take', 'up', 'more')	1
('take', 'up', 'more', 'time')	1
('up', 'more', 'time', 'to')	1
('more', 'time', 'to', 'read')	1
('time', 'to', 'read', 'thank')	1
('to', 'read', 'thank', 'you')	1
('read', 'thank', 'you', 'org')	1
('thank', 'you', 'org', 'esperanza')	1
('you', 'org', 'esperanza', 'community')	1
('gpe', 'org', 'of', 'gpe')	0.956521739
('cardinal', 'other', 'organizations', 'that')	0.956521739
('along', 'with', 'org', 'org')	0.956521739
('org', 'of', 'gpe', 'hi')	0.954545455
('up', 'and', 'down', 'the')	0.8
('and', 'down', 'the', 'state')	0.8
('behalf', 'of', 'the', 'cities')	0.714285714
('of', 'the', 'cities', 'of')	0.714285714
('for', 'that', 'reason', 'we')	0.666666667
('thats', 'going', 'to', 'be')	0.666666667
('senators', 'person', 'on', 'behalf')	0.666666667
('i', 'have', 'been', 'a')	0.666666667
('person', 'representing', 'org', 'also')	0.666666667
('of', 'org', 'i', 'would')	0.666666667
('this', 'is', 'something', 'that')	0.6
('we', 'would', 'argue', 'that')	0.6
('chairman', 'person', 'and', 'members')	0.571428571
('was', 'one', 'of', 'the')	0.5
('in', 'opposition', 'to', 'this')	0.5
('person', 'from', 'org', 'we')	0.5
('names', 'person', 'im', 'a')	0.5
('have', 'some', 'concerns', 'about')	0.5
('and', 'for', 'that', 'reason')	0.5
('youre', 'going', 'to', 'have')	0.5
('person', 'from', 'org', 'in')	0.5
('on', 'a', 'date', 'basis')	0.5
('in', 'gpe', 'which', 'is')	0.5
('way', 'to', 'do', 'that')	0.5

('from', 'org', 'and', 'we')	0.5
('good', 'time', 'person', 'with')	0.5
('time', 'person', 'with', 'org')	0.5
('org', 'we', 'are', 'also')	0.5
('more', 'than', 'happy', 'to')	0.5
('i', 'would', 'just', 'add')	0.5
('as', 'a', 'result', 'of')	0.5
('that', 'theyre', 'going', 'to')	0.5
('you', 'would', 'have', 'to')	0.5
('on', 'behalf', 'of', 'norp')	0.5
('behalf', 'of', 'org', 'california')	0.5
('we', 'are', 'also', 'in')	0.5
('throughout', 'the', 'state', 'of')	0.5
('that', 'we', 'have', 'and')	0.5
('were', 'one', 'of', 'the')	0.5
('one', 'of', 'the', 'cosponsors')	0.5
('to', 'come', 'in', 'and')	0.5
('of', 'the', 'cosponsors', 'of')	0.5
('that', 'would', 'be', 'a')	0.5
('do', 'not', 'want', 'to')	0.5
('for', 'this', 'reason', 'we')	0.5
('this', 'bill', 'seeks', 'to')	0.5
('org', 'and', 'org', 'also')	0.5
('that', 'we', 'have', 'to')	0.5
('we', 'are', 'supportive', 'of')	0.5
('im', 'from', 'gpe', 'gpe')	0.5
('org', 'county', 'and', 'municipal')	0.5
('person', 'with', 'org', 'county')	0.5
('with', 'org', 'county', 'and')	0.5
('to', 'do', 'that', 'is')	0.5
('and', 'members', 'cesar', 'diaz')	0.5
('should', 'not', 'have', 'to')	0.5
('for', 'a', 'couple', 'of')	0.5
('and', 'municipal', 'employees', 'in')	0.5

Successful 5 word phrases for testimonies with "For" or "Indeterminate" alignment:

5 word Phrase (Pentagram)	Success Rate
('that', 'this', 'is', 'going', 'to')	1
('with', 'org', 'in', 'opposition', 'thank')	1
('we', 'ask', 'for', 'your', 'no')	1
('for', 'your', 'no', 'vote', 'on')	1
('you', 'to', 'oppose', 'this', 'bill')	1
('do', 'not', 'have', 'a', 'position')	1
('were', 'opposed', 'to', 'the', 'bill')	1
('opposed', 'to', 'the', 'bill', 'thank')	1
('opposition', 'person', 'on', 'behalf', 'of')	1
('for', 'those', 'reasons', 'were', 'opposed')	1
('represent', 'cardinal', 'regional', 'apartment', 'associations')	1
('and', 'technology', 'association', 'also', 'opposed')	1
('ask', 'you', 'to', 'vote', 'no')	1
('dont', 'have', 'an', 'official', 'position')	1
('we', 'are', 'strongly', 'opposed', 'to')	1
('strongly', 'opposed', 'to', 'this', 'bill')	1
('for', 'those', 'reasons', 'we', 'remain')	1
('we', 'dont', 'have', 'an', 'official')	1
('for', 'these', 'reasons', 'we', 'oppose')	1
('are', 'opposed', 'to', 'this', 'measure')	1
('and', 'org', 'in', 'opposition', 'person')	1
('to', 'this', 'bill', 'thank', 'you')	1
('no', 'vote', 'on', 'this', 'bill')	1
('in', 'opposition', 'person', 'on', 'behalf')	1
('are', 'opposed', 'to', 'this', 'bill')	1
('were', 'going', 'to', 'continue', 'to')	1
('with', 'the', 'author', 'and', 'we')	1
('with', 'org', 'in', 'opposition', 'to')	1
('your', 'no', 'vote', 'on', 'this')	1
('person', 'i', 'represent', 'cardinal', 'regional')	1
('oppose', 'this', 'measure', 'thank', 'you')	1
('manufacturers', 'and', 'technology', 'association', 'also')	1
('we', 'do', 'not', 'believe', 'that')	1
('i', 'represent', 'cardinal', 'regional', 'apartment')	1
('not', 'have', 'a', 'position', 'on')	0.916666667
('a', 'position', 'on', 'the', 'bill')	0.909090909
('we', 'dont', 'have', 'a', 'position')	0.888888889
('have', 'a', 'position', 'on', 'the')	0.888888889
('with', 'org', 'we', 'dont', 'have')	0.888888889
('org', 'we', 'do', 'not', 'have')	0.888888889
('we', 'do', 'have', 'some', 'concerns')	0.875
('have', 'some', 'concerns', 'with', 'the')	0.875
('we', 'do', 'not', 'have', 'a')	0.875
('live', 'in', 'gpe', 'and', 'i')	0.857142857
('forward', 'to', 'continuing', 'to', 'work')	0.833333333

('to', 'continuing', 'to', 'work', 'with')	0.833333333
('org', 'we', 'dont', 'have', 'a')	0.833333333
('have', 'an', 'official', 'position', 'on')	0.8
('the', 'author', 'and', 'the', 'sponsors')	0.8
('like', 'to', 'work', 'with', 'the')	0.8
('we', 'look', 'forward', 'to', 'continuing')	0.8
('dont', 'have', 'a', 'position', 'on')	0.8
('and', 'look', 'forward', 'to', 'continuing')	0.75
('you', 'so', 'much', 'for', 'your')	0.75
('gpe', 'manufacturers', 'and', 'technology', 'association')	0.75
('position', 'on', 'the', 'bill', 'and')	0.75
('there', 'are', 'a', 'lot', 'of')	0.75
('person', 'person', 'on', 'behalf', 'of')	0.75
('to', 'make', 'sure', 'that', 'this')	0.75
('continuing', 'to', 'work', 'with', 'the')	0.666666667
('id', 'like', 'to', 'thank', 'the')	0.666666667
('behalf', 'of', 'the', 'chief', 'probation')	0.666666667
('of', 'org', 'we', 'have', 'a')	0.666666667
('and', 'i', 'just', 'wanted', 'to')	0.666666667
('the', 'provisions', 'of', 'this', 'bill')	0.666666667
('on', 'behalf', 'of', 'the', 'chief')	0.666666667
('person', 'with', 'org', 'we', 'dont')	0.666666667
('working', 'with', 'the', 'author', 'to')	0.666666667
('behalf', 'of', 'org', 'we', 'have')	0.666666667
('we', 'continue', 'to', 'work', 'with')	0.666666667
('the', 'chief', 'probation', 'officers', 'of')	0.666666667
('work', 'with', 'the', 'author', 'and')	0.666666667
('not', 'going', 'to', 'be', 'able')	0.666666667
('look', 'forward', 'to', 'continuing', 'to')	0.666666667
('behalf', 'of', 'org', 'we', 'too')	0.666666667
('mr', 'person', 'person', 'on', 'behalf')	0.666666667
('with', 'the', 'authors', 'office', 'to')	0.666666667
('chief', 'probation', 'officers', 'of', 'gpe')	0.666666667
('of', 'the', 'chief', 'probation', 'officers')	0.666666667
('some', 'of', 'the', 'concerns', 'that')	0.666666667
('cardinal', 'of', 'the', 'things', 'that')	0.666666667
('with', 'the', 'org', 'and', 'the')	0.666666667
('one', 'of', 'the', 'things', 'that')	0.666666667
('of', 'the', 'things', 'that', 'we')	0.666666667
('org', 'i', 'want', 'to', 'thank')	0.666666667
('the', 'gpe', 'manufacturers', 'and', 'technology')	0.666666667
('madame', 'chair', 'and', 'members', 'of')	0.5
('of', 'the', 'org', 'for', 'the')	0.5
('aye', 'vote', 'on', 'this', 'bill')	0.5
('person', 'for', 'his', 'leadership', 'on')	0.5
('the', 'intent', 'of', 'the', 'bill')	0.5
('on', 'behalf', 'of', 'the', 'urban')	0.5

('do', 'want', 'to', 'thank', 'the')	0.5
('org', 'in', 'opposition', 'thank', 'you')	0.5
('for', 'the', 'reasons', 'previously', 'stated')	0.5
('time', 'person', 'with', 'org', 'were')	0.5
('like', 'to', 'thank', 'the', 'org')	0.5
('support', 'of', 'the', 'bill', 'thank')	0.5
('like', 'to', 'thank', 'the', 'author')	0.5
('to', 'echo', 'the', 'comments', 'made')	0.5
('on', 'behalf', 'of', 'org', 'ill')	0.5
('his', 'leadership', 'on', 'this', 'issue')	0.5
('for', 'your', 'aye', 'vote', 'thank')	0.5
('to', 'work', 'with', 'the', 'author')	0.5
('of', 'org', 'id', 'like', 'to')	0.5
('so', 'for', 'those', 'reasons', 'were')	0.5
('org', 'in', 'support', 'of', 'the')	0.5
('just', 'want', 'to', 'thank', 'the')	0.5
('and', 'members', 'person', 'with', 'the')	0.5
('person', 'im', 'with', 'org', 'and')	0.5
('much', 'for', 'the', 'opportunity', 'to')	0.5
('org', 'i', 'just', 'want', 'to')	0.5
('bill', 'moves', 'forward', 'thank', 'you')	0.5
('names', 'person', 'im', 'with', 'org')	0.5
('with', 'the', 'author', 'and', 'the')	0.5
('my', 'names', 'person', 'im', 'with')	0.5
('behalf', 'of', 'the', 'urban', 'counties')	0.5
('thank', 'the', 'author', 'for', 'his')	0.5
('to', 'come', 'up', 'with', 'a')	0.5
('the', 'cap', 'and', 'trade', 'program')	0.5
('we', 'are', 'opposed', 'to', 'this')	0.5
('want', 'to', 'say', 'that', 'we')	0.5
('is', 'person', 'representing', 'org', 'and')	0.5
('of', 'org', 'i', 'want', 'to')	0.5
('so', 'we', 'urge', 'your', 'support')	0.5
('thank', 'you', 'mr', 'person', 'and')	0.5
('have', 'a', 'formal', 'position', 'on')	0.5
('person', 'with', 'org', 'in', 'support')	0.5
('gpe', 'mayor', 'person', 'in', 'support')	0.5
('just', 'want', 'to', 'say', 'that')	0.5
('there', 'needs', 'to', 'be', 'a')	0.5
('of', 'org', 'in', 'opposition', 'thank')	0.5
('continue', 'to', 'work', 'with', 'the')	0.5
('thank', 'you', 'mr', 'person', 'my')	0.5
('on', 'behalf', 'of', 'org', 'just')	0.5
('to', 'be', 'able', 'to', 'do')	0.5
('of', 'the', 'bill', 'thank', 'you')	0.5
('with', 'org', 'in', 'support', 'of')	0.5
('for', 'the', 'org', 'of', 'gpe')	0.5

('of', 'the', 'urban', 'counties', 'of')	0.5
('for', 'his', 'leadership', 'on', 'this')	0.5
('on', 'behalf', 'of', 'org', 'id')	0.5
('thank', 'you', 'so', 'much', 'for')	0.5
('mr', 'chair', 'and', 'members', 'person')	0.5
('you', 'mr', 'person', 'and', 'members')	0.5
('you', 'to', 'vote', 'yes', 'on')	0.5
('person', 'for', 'org', 'in', 'support')	0.5
('ask', 'for', 'a', 'no', 'vote')	0.5
('behalf', 'of', 'org', 'we', 'also')	0.5
('person', 'here', 'on', 'behalf', 'of')	0.5
('committee', 'person', 'on', 'behalf', 'of')	0.5
('good', 'time', 'person', 'of', 'org')	0.5
('on', 'behalf', 'of', 'gpes', 'org')	0.5
('to', 'thank', 'the', 'author', 'and')	0.5
('person', 'with', 'org', 'we', 'are')	0.5
('thank', 'the', 'author', 'and', 'his')	0.5
('behalf', 'of', 'org', 'id', 'like')	0.5
('to', 'the', 'bill', 'thank', 'you')	0.5
('urban', 'counties', 'of', 'gpe', 'also')	0.5
('behalf', 'of', 'org', 'i', 'want')	0.5
('you', 'mr', 'person', 'my', 'name')	0.5
('good', 'time', 'thank', 'you', 'for')	0.5
('madam', 'person', 'members', 'person', 'with')	0.5
('of', 'the', 'consumer', 'attorneys', 'of')	0.5
('working', 'with', 'the', 'author', 'and')	0.5
('org', 'also', 'in', 'support', 'of')	0.5
('org', 'id', 'like', 'to', 'thank')	0.5
('in', 'support', 'of', 'the', 'bill')	0.5
('my', 'names', 'person', 'im', 'the')	0.5

Successful 5 word phrases for testimonies with "Against" or "Indeterminate" alignment:

5 word phrase (pentagram)	Success Rate
('to', 'move', 'america', 'and', 'cardinal')	1
('with', 'org', 'org', 'of', 'gpe')	1
('more', 'time', 'to', 'read', 'thank')	1
('to', 'oppose', 'along', 'with', 'org')	1
('councils', 'org', 'fair', 'rents', 'for')	1
('gpe', 'org', 'of', 'gpe', 'hi')	1
('org', 'org', 'beyond', 'the', 'arc')	1
('partnership', 'for', 'justice', 'org', 'coalition')	1
('public', 'advocates', 'here', 'to', 'oppose')	1
('community', 'technologies', 'community', 'health', 'councils')	1
('to', 'take', 'up', 'more', 'time')	1
('thank', 'you', 'org', 'esperanza', 'community')	1
('person', 'with', 'public', 'advocates', 'here')	1
('oppose', 'along', 'with', 'org', 'org')	1
('for', 'gpe', 'public', 'counsel', 'act')	1
('time', 'to', 'read', 'thank', 'you')	1
('here', 'to', 'oppose', 'along', 'with')	1
('of', 'gpe', 'org', 'of', 'gpe')	1
('place', 'jobs', 'to', 'move', 'america')	1
('community', 'health', 'councils', 'org', 'fair')	1
('health', 'councils', 'org', 'fair', 'rents')	1
('gpe', 'public', 'counsel', 'act', 'la')	1
('read', 'thank', 'you', 'org', 'esperanza')	1
('org', 'beyond', 'the', 'arc', 'a')	1
('bother', 'to', 'take', 'up', 'more')	1
('justice', 'org', 'coalition', 'for', 'economic')	1
('i', 'wont', 'bother', 'to', 'take')	1
('jobs', 'to', 'move', 'america', 'and')	1
('esperanza', 'community', 'housing', 'corporation', 'person')	1
('along', 'with', 'org', 'org', 'of')	1
('housing', 'corporation', 'person', 'org', 'investing')	1
('for', 'economic', 'survival', 'community', 'technologies')	1
('in', 'place', 'jobs', 'to', 'move')	1
('up', 'more', 'time', 'to', 'read')	1
('move', 'america', 'and', 'cardinal', 'other')	1
('counsel', 'act', 'la', 'org', 'org')	1
('other', 'organizations', 'that', 'i', 'wont')	1
('for', 'justice', 'org', 'coalition', 'for')	1
('beyond', 'the', 'arc', 'a', 'norp')	1
('that', 'i', 'wont', 'bother', 'to')	1
('org', 'of', 'gpe', 'org', 'of')	1
('you', 'org', 'esperanza', 'community', 'housing')	1
('la', 'org', 'org', 'beyond', 'the')	1
('rents', 'for', 'gpe', 'public', 'counsel')	1
('survival', 'community', 'technologies', 'community', 'health')	1

('technologies', 'community', 'health', 'councils', 'org')	1
('org', 'org', 'of', 'gpe', 'org')	1
('advocates', 'here', 'to', 'oppose', 'along')	1
('take', 'up', 'more', 'time', 'to')	1
('org', 'of', 'gpe', 'hi', 'im')	1
('wont', 'bother', 'to', 'take', 'up')	1
('with', 'public', 'advocates', 'here', 'to')	1
('community', 'housing', 'corporation', 'person', 'org')	1
('act', 'la', 'org', 'org', 'beyond')	1
('economic', 'survival', 'community', 'technologies', 'community')	1
('america', 'and', 'cardinal', 'other', 'organizations')	1
('gpe', 'hi', 'im', 'person', 'with')	1
('org', 'fair', 'rents', 'for', 'gpe')	1
('and', 'cardinal', 'other', 'organizations', 'that')	1
('person', 'org', 'investing', 'in', 'place')	1
('to', 'read', 'thank', 'you', 'org')	1
('org', 'investing', 'in', 'place', 'jobs')	1
('org', 'esperanza', 'community', 'housing', 'corporation')	1
('investing', 'in', 'place', 'jobs', 'to')	1
('cardinal', 'other', 'organizations', 'that', 'i')	1
('public', 'counsel', 'act', 'la', 'org')	1
('corporation', 'person', 'org', 'investing', 'in')	1
('coalition', 'for', 'economic', 'survival', 'community')	1
('org', 'coalition', 'for', 'economic', 'survival')	1
('organizations', 'that', 'i', 'wont', 'bother')	1
('norp', 'partnership', 'for', 'justice', 'org')	1
('fair', 'rents', 'for', 'gpe', 'public')	1
('of', 'gpe', 'hi', 'im', 'person')	1
('up', 'and', 'down', 'the', 'state')	0.8
('on', 'behalf', 'of', 'the', 'cities')	0.714285714
('behalf', 'of', 'the', 'cities', 'of')	0.714285714
('senators', 'person', 'on', 'behalf', 'of')	0.666666667
('the', 'comments', 'of', 'my', 'colleagues')	0.5
('on', 'behalf', 'of', 'org', 'california')	0.5
('with', 'org', 'county', 'and', 'municipal')	0.5
('with', 'org', 'here', 'date', 'on')	0.5
('org', 'county', 'and', 'municipal', 'employees')	0.5
('with', 'org', 'also', 'in', 'opposition')	0.5
('good', 'time', 'person', 'with', 'org')	0.5
('throughout', 'the', 'state', 'of', 'gpe')	0.5
('org', 'we', 'are', 'also', 'in')	0.5
('with', 'org', 'also', 'in', 'strong')	0.5
('mr', 'person', 'and', 'members', 'of')	0.5
('time', 'mr', 'person', 'person', 'on')	0.5
('one', 'of', 'the', 'cosponsors', 'of')	0.5
('time', 'person', 'with', 'org', 'also')	0.5
('my', 'names', 'person', 'im', 'a')	0.5